

REGULAR ARTICLE

Open Access



# Using wearable proximity sensors to characterize social contact patterns in a village of rural Malawi

Laura Ozella<sup>1\*</sup> , Daniela Paolotti<sup>1</sup>, Guilherme Lichand<sup>2</sup>, Jorge P. Rodríguez<sup>1</sup>, Simon Haenni<sup>2</sup>, John Phuka<sup>3</sup>, Onicio B. Leal-Neto<sup>2</sup> and Ciro Cattuto<sup>1,4</sup>

\*Correspondence: [laura.ozella@isi.it](mailto:laura.ozella@isi.it)

<sup>1</sup>ISI Foundation, Turin, Italy

Full list of author information is available at the end of the article

## Abstract

Measuring close proximity interactions between individuals can provide key information on social contacts in human communities and related behaviours. This is even more essential in rural settings in low- and middle-income countries where there is a need to understand contact patterns for the implementation of strategies for social protection interventions. We report the quantitative assessment of contact patterns in a village in rural Malawi, based on proximity sensors technology that allows for high-resolution measurements of social contacts. Our results revealed that the community structure of the village was highly correlated with the household membership of the individuals, thus confirming the importance of the family ties within the village. Social contacts within households occurred mainly between adults and children, and adults and adolescents and most of the inter-household social relationships occurred among adults and among adolescents. At the individual level, age and gender social assortment were observed in the inter-household network, and age disassortativity was instead observed in intra-household networks. Moreover, we obtained a clear trend of the daily contact activity of the village. Family members congregated in the early morning, during lunch time and dinner time. In contrast, inter-household contact activity displayed a growth from the morning, reaching a maximum in the afternoon.

The proximity sensors technology used in this study provided high resolution temporal data characterized by timescales comparable with those intrinsic to social dynamics and it thus allowed to have access to the level of information needed to understand the social context of the village.

**Keywords:** Social contact patterns; Contact network; Wearable proximity sensors; Rural settings; Developing countries; Households

## 1 Introduction

Describing close proximity interactions allows to create contact networks representing frequency of social contacts in human communities. Contact network analysis can be used to better understand social interaction patterns and related behaviours (Borgatti et al. [3]; Chami et al. [8]) and the transmission of diseases (Funk, et al. [17]; Danon et al. [12]). Col-

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

lecting high resolution data on the contact rates between individuals is a major challenge in most settings, particularly in rural low- and middle-income areas. Furthermore, many infectious diseases have emerged or re-emerged in rural low- and middle-income settings in the last century (Fenollar et al. [15]). Including observed contact data, of these harder to reach populations, in stochastic models of transmissible diseases could help better predict epidemics

and has utility in the design of preventive and control measures such as vaccination and social distancing (Mossong et al. [35]; Salathé et al. [41]).

### 1.1 Related works

Contact diaries have usually been used to record contact information in low- and middle-income countries. Few studies have investigated social contact patterns in the Asian continent. Social networks in household-structured communities have been estimated in Vietnam (Horby et al. [19]) and in rural India (Johny et al. [23]) through the use of paper-based questionnaires. Banerjee et al. [1] studied the social network of rural villages of Southern India using questionnaire and they developed a model of information diffusion. In Africa, a robust literature documents the use of contact diaries to understand and quantify households' connectivity (Hosegood and Timaeus [20]; Cassidy and Barnes [5]) and social contacts relevant for diseases transmission (Johnstone-Robertson et al. [22]; Crampin et al. [11]; Chami et al. [8]; Kiti et al. [26]). Specifically, in Malawi, Rock et al. [41] have studied the social networks and social participation of youth living in extreme poverty in rural sites, and Helleringer and Kohler [18] have investigated the relationship between the sexual networks of young adult population of several villages and the position of HIV-positive individuals within these networks. However, such survey-based approach permits obtaining only an approximate estimate of the number of contacts and of their durations, especially when repeated over many days or weeks. Moreover, young or illiterate participants may face difficulty in comprehending the process of completion of the questionnaire and need the assistance of trained workers (Johnstone-Robertson et al. [22]).

Recently, the use of proximity sensing technology to collect social contact information has emerged as an alternative. Mobile phones have been used to continuously collect proximity information within rural communities by Bluetooth devices within the scanning range of the user (typically 5–10 m) (Yoneki and Crowcroft [48]). On the other hand, wireless proximity sensors can gather proximity interactions less than 1–2 meters separation distance every few seconds (Cattuto et al. [6]) in an objective way. These proximity events are relevant to detect social interactions, such as conversations, or spread of infections through physical touch or via aerosol. Proximity sensors were successfully deployed in rural Kenya to characterize contact patterns that may shape the transmission of respiratory infections in schools and households (Kiti et al. [28]; Kiti et al. [27]).

### 1.2 Objectives of the study

In this study, we report on the use of wearable proximity sensors to measure face-to-face proximity and pattern of contacts between individuals in a village in rural Malawi. The present study is part of a wider project on Child and Youth Development Study funded by UNICEF Malawi (Leal Neto et al. [30]) aimed to track children's health and to better understand child development in poor country settings. The primary aim of this study is to estimate social contact patterns to understand the social context within a village in rural

Malawi. Specifically, we investigated the social interactions within household and across households, to better understand the social relationships by age-categories. We analyzed separately the intra- and the inter-households' social relationships to assess the differences in contact patterns among individuals belonging to the same family group and the different family groups, in particular the relationships of children with the other age-categories. We measured group-level patterns by testing for community structure in the whole social network of the village, and we evaluated the performance of a community detection algorithm by their ability to find so-called ground truth communities (i.e., age, gender and household memberships), and the goal is to find communities that align with age, gender or household membership. We meant to deeply understand the sociality within the village by considering also the dyadic social interactions. Thus, we explored individual-level patterns by measuring the social assortment of individuals by age and by gender, to test the tendency of individuals to interact with individuals with whom they shared similar attributes. Finally, we assessed the social contact patterns at the temporal level, by studying the daily activity profile of contacts among individuals split in intra-household and inter-household contacts. This study is an extension of the Kiti et al. [28] study that took place in rural Kenya. Here, we used a larger sample and we analyzed inter-household contact patterns.

The remaining part of the paper is organized as follows. Material and methods section presents the proximity sensing technology used, and it introduces the setting in which the data collection took place, including the ethical aspect of the research. Moreover, this section illustrates how the data was cleaned and the data analysis was performed. The empirical findings on the social structure of the village are provided in the Results section. Discussion section then discusses the results and describes the related works more extensively. Conclusions and limitations of the study are drawn in the Conclusion session.

## 2 Materials and methods

### 2.1 Data collection

The present study is part of a wider project on Child and Youth Development Study funded by UNICEF Malawi (Leal Neto et al. [30]). The data collection was conducted between 16th December 2019 and 10th January 2020 in Mdoliro village in Dowa district in the Central Region of Malawi. Mdoliro is a small village with an estimated 2019 population of 147 distributed over 32 households (average household size is 4.5). The majority of the inhabitants were Christians, and the Chewas are the main ethnic group. Farming is the major source of income.

The data were obtained and processed using a proximity-sensing application previously used to measure individuals social contacts in a variety of real-world settings as hospital wards (Isella et al. [21]; Voirin et al. [46]), schools (Stehlé et al. [45]), social events (Smieszek, et al. [44]), households (Ozella et al. [38]; Kiti et al. [28]), and more recently it was used to detect proximity events between animals (e.g., Wilson-Aggarwal et al. [47]). This technology is based on wearable proximity sensors that exchange ultra-low power radio packets and the use of the sensors to detect face-to face interactions is described in detail in Cattuto *et al.* [6]. Sensors in close proximity exchange with one another a maximum of about 1 power packet per second, and the exchange of low-power radio-packets is used as a proxy for the spatial proximity of the individuals wearing the sensors (Cattuto *et al.* [6]). To estimate how close individuals are, the attenuation of the signals

with distance is computed as the difference between the received and transmitted power. Proximity between individuals is asserted when the median attenuation over a given time interval exceeds a specified attenuation threshold (in dBm). In this study, we set the attenuation threshold at  $-75$  dBm, this threshold was already used in previous deployments (e.g., Ozella et al. [37]) and allowed the detection of proximity events between devices situated in the range  $1\text{--}1.5$  m of one another. This distance between individuals allows detection of a close-contact situation during which social interactions might occur and a communicable disease infection might be directly transmitted. A ‘contact event’ between two individuals was identified when the devices exchanged at least one radio packet during a time interval of 20 sec. After a contact is established, it is considered ongoing as long as the devices continue to exchange at least one radio packet for every subsequent 20 s interval. Conversely, a contact was considered broken if a 20 s interval elapses with no exchange of radio packets (Stehlé et al. [45]; Kiti et al. [27]). Each device has a unique identification (ID) number that is used to link the information on the contacts established by the individual carrying the device. For the present study, the technology was operated in a distributed fashion: contact data were stored in the local memory of individual devices. After collecting the devices at the end of the study, data from individual devices were downloaded, and the (temporal) contact networks recorded by individual devices were combined together to build a time-resolved proximity graph. In addition to contact information, each device periodically logs its orientation in space as measured by a tri-axial accelerometer.

The participants wore a sensor enclosed in a pouch and pinned to the front of a blouse/shirt in order to detect close-range proximity. The low-power radio frequency in use cannot propagate through the human body, and the position of the sensor favours capturing face-to-face interactions. In addition, metadata on individuals were collected, i.e., gender, age and which household they belonged to, through the use of the app Survey CTO. A household was defined as the group of people living in the same house and eating from the same kitchen (Hosegood and Timaeus [20]). Participants were grouped into three age-categories: <10 years old (children), 11–18 years old (adolescents), and >18 years old (adults). Training sessions were conducted with Health Surveillance Assistants (HSAs) and volunteers in the use of sensors and how participants should have worn them over the study period, and HSAs visited the village in order to check if participants were wearing the sensors properly. The Ministry of Health (MOH) defines an HSA as a primary healthcare worker serving as a link between a health facility and the community (Chikaphupha et al. [9]). HSAs’ tasks include community health, family health, environmental health, prevention and control of communicable diseases, and community case management. Specifically, two HSAs were involved in the study: one of them lived in Mdoliro (i.e., one adult participant that had also the role of HSA), and one HSA was external from the village’s population, and visited the village at least once a week.

## 2.2 Ethical aspects

Only participants who gave their written consent (documented) were included in the research. In the case of children, consent was obtained from their guardians. In the case of adolescents, consent was obtained from both themselves and their guardians. The study was approved by Ethical Committee at the University of Zurich (OEC IRB #2018-046) and Ethical Committee at College of Medicine in Malawi (P.10/19/2825).

### 2.3 Data analysis

The proximity data were extracted from devices and cleaned by identifying anomalies in the recorded data that might point to sensors that were tampered with or suffered hardware/battery issues resulting in data loss or low-quality data. Participants were asked to remove the sensor overnight. Night contacts were disregarded from the analysis by using the tri-axial accelerometer data to identify the time periods during which the sensor did not move. This also allowed us to identify the time periods during which the sensors were not worn by the participants. This data was also disregarded from the analyses.

### 2.4 Network analysis and contact matrices

For each participant in the study, we computed the number of contact events and the duration of each contact. Time-aggregated, weighted contact networks were generated: nodes correspond to individuals, an edge between two individuals indicates that at least one contact event involving those individuals was recorded during the whole experimental period. The weight  $w_{ij}$  of an edge between nodes  $i$  and  $j$  is defined as the cumulative duration of the contact events recorded between those individuals. Network edges are undirected and the weights on the edges are regarded as symmetric ( $w_{ij} = w_{ji}$ ). The degree  $k_i$  of a node  $i$  in the above network corresponds to the number of distinct individuals with whom individual  $i$  has been in contact. Intra-household and inter-household contact matrices were generated based on the daily number and on the daily duration of contacts by age-category. We aimed to obtain the daily mean of the contact durations and the daily mean of contact events per capita for each pair of age-categories. To obtain the intra-household matrices, we divided the total contact durations and the total number of contacts by the days during which the family members wore sensors simultaneously (days of overlap), and by the number of persons belonging to the two age-categories ( $a$  and  $b$ ):

$$\text{daily mean contact duration}(a, b) = \frac{\left( \frac{\text{total contact durations } (a, b)}{\text{days of overlap}} \right)}{(a + b)/2},$$

$$\text{daily mean contact events } (a, b) = \frac{\left( \frac{\text{total number of contact events } (a, b)}{\text{days of overlap}} \right)}{(a + b)/2}.$$

To obtain the inter-household matrices, we divided the total contact durations and the total number of contacts by the days during which all the participants wore the sensors simultaneously. The observed values were compared to those obtained by a null model: we shuffled the nodes attributes of the intra- and inter-household edge lists and we computed for those realizations the daily number and the daily duration of contacts by age-category.

### 2.5 Community detection

Community detection seeks to describe the large-scale structure of a network by dividing its nodes into communities (or groups), based only on the pattern of links among those nodes (Contisciani et al. [10]). Nodes belonging to communities are more highly connected to each other than to the rest of the network and probably share common properties. We used the Louvain algorithm (Blondel et al. [2]) to identify community structure in the aggregated networks. The modularity is a measure of the structure of networks which measures the strength of division of a network into communities, and this method maximizes a modularity score for each community. The algorithm assesses how much more

closely connected the nodes within a group are, compared to how connected they would be in a random network (Borgatti et al. [3]; Lu et al. [32]). We used the Normalized Mutual Information (NMI) score to test the relationship between the community membership of participants and their attributes (i.e., gender, age-category and household membership). We aimed to evaluate if the communities detected by the algorithm corresponded to participants' attributes. NMI score ranges between 0 (no mutual information) and 1 (perfect correlation). NMI scores closer to 1 imply a greater overlap between community membership and attributes.

## 2.6 Assortativity

We studied the assortativity (i.e., the tendency of the individuals to associate with individuals of similar characteristics) in the gender and the age-category. Our aim was to understand if individuals with the same gender and the same age-category will be more likely to interact, and if there are differences between intra and inter-households contacts. We computed the number of contacts and the total time in contacts (weights) between individuals in aggregated observed networks. We compared the values obtained by the observed networks, with those obtained from a null model. We created an ensemble of realizations of a null model by shuffling nodes' attributes and computing for those realizations the assortativity. Then, we compared the empirical results with the distribution of values obtained from the null model using the z-test.

## 2.7 Daily activity profiles

We studied the daily activity profile of contacts among individuals, extracting the probability of observing a contact as a function of the time along the day. We computed these activity profiles for each household, split in intra-household and inter-household contacts. Additionally, we create two aggregated datasets that join all the timestamps of the observed contacts: the aggregated data for intra-household contacts and the aggregated data for inter-household contacts. We computed the Kolmogorov–Smirnov statistical distance  $d_{KS}$  between the aggregated data of all the households and the data observed for each household (for intra-household daily activity profile), and of all individuals and the data observed for each individual (for inter-households daily activity profile). The Kolmogorov–Smirnov distance between two probability distributions  $i$  and  $j$  depending on time  $\tau$  is the maximum, among all the times, of the absolute difference between the cumulative distribution functions  $C_i(\tau)$  and  $C_j(\tau)$ :

$$d_{KS}(i,j) = \max_{\tau} |C_i(\tau) - C_j(\tau)|.$$

This distance is bounded between 0, when the two compared distributions are identical, and 1, when the overlap between both distributions is null. In this particular case, we studied the distributions arising from the empirically observed contacts. The times of the contacts of each household (for intra-household data) and of each individual (for inter-households data)  $i$  are described by the cumulative distribution function  $C_i$  given by

$$C_i(\tau) = 1 - \frac{\sum_{t < \tau} N_i(t)}{\sum_t N_i(t)},$$

where  $N_i(t)$  is the number of observed contacts at time  $t$ , computed identically for the aggregated data. We considered a range of times of 24 h duration. We observe that the

periodicity of this range may influence the measurement of dKS. For example, if we consider the origin of times at midnight and there are contacts between 23:00 and 01:00 in the aggregated dataset, but there is no contact at one specific household along this range, the maximum difference between the cumulatives would be influenced by the origin of times, such that this difference would be lower if the origin of times is included along the observed range. To tackle this issue, we generated the origin of times as a uniform random number between 0 and 24 h. We generated, for each pair of distributions, 100 different origins of times, computed the Kolmogorov–Smirnov distance for each origin of times, and the considered  $d_{KS}$  was the minimum among all these samples.

### 3 Results

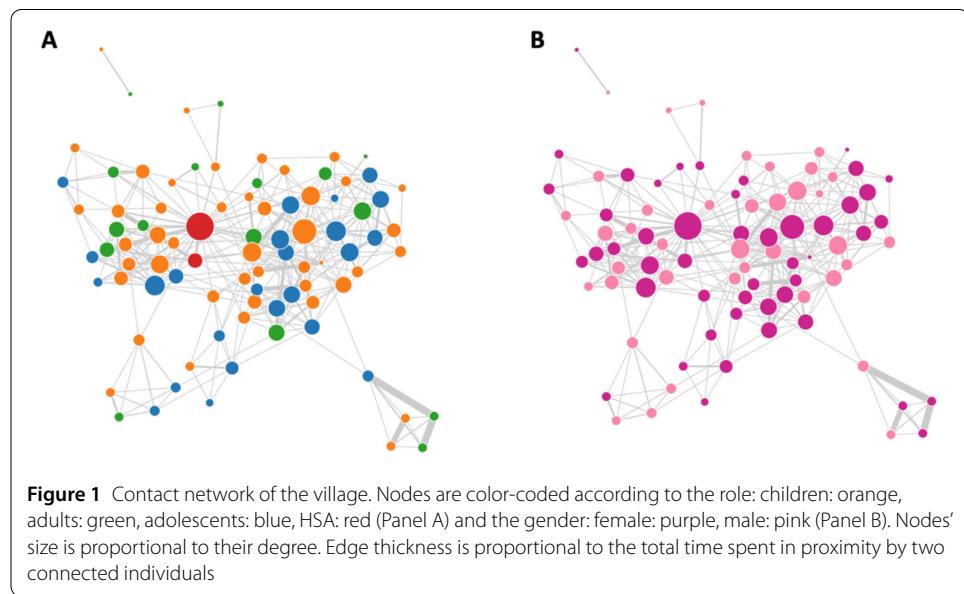
The whole study duration was 26 days (from 16th December 2019 to 10th January 2020), however, the proximity sensors were deployed and collected at different times within this time window. Therefore, each sensor had a different deployment period, and this varied from 16 days to 23 days (median 20 days).

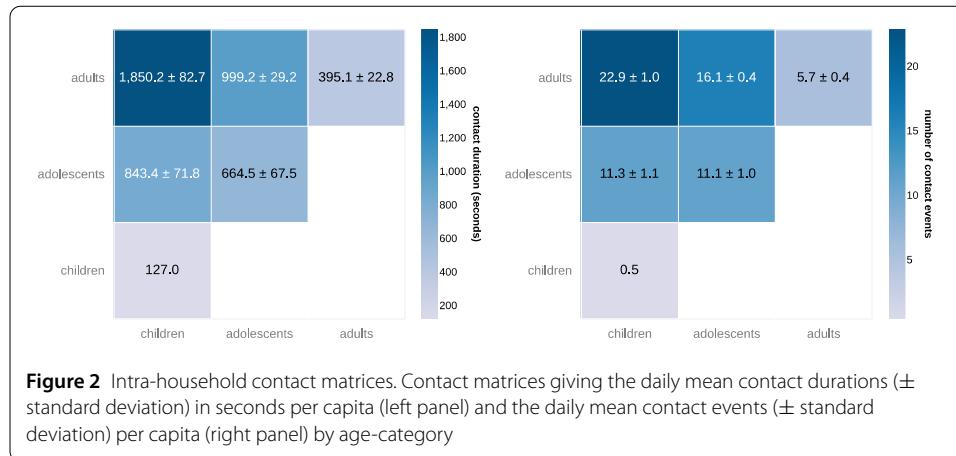
We included in the data analysis the contact data for all the participants when they wore the sensors simultaneously, and this overlapping deployment period was 13 days. On the other hand, we studied the intra-household contact data when the family members wore sensors simultaneously, and this overlapping deployment period was different according to the household (range 16–22 days). The initial number of sensors deployed was 99. However, a total of 86 sensors were included in the data analysis, since we excluded the sensors that had anomalies, the sensors that were not properly worn by the participants, and the sensors that did not register contacts for the entire duration of the study.

#### 3.1 Network analysis and contact matrices

##### 3.1.1 Contact network

Figure 1 displays a representation of the contact network of the village (86 nodes and 355 edges), where nodes are individuals and edges indicate the presence of at least one recorded contact during the study period. Nodes are color coded according to the role





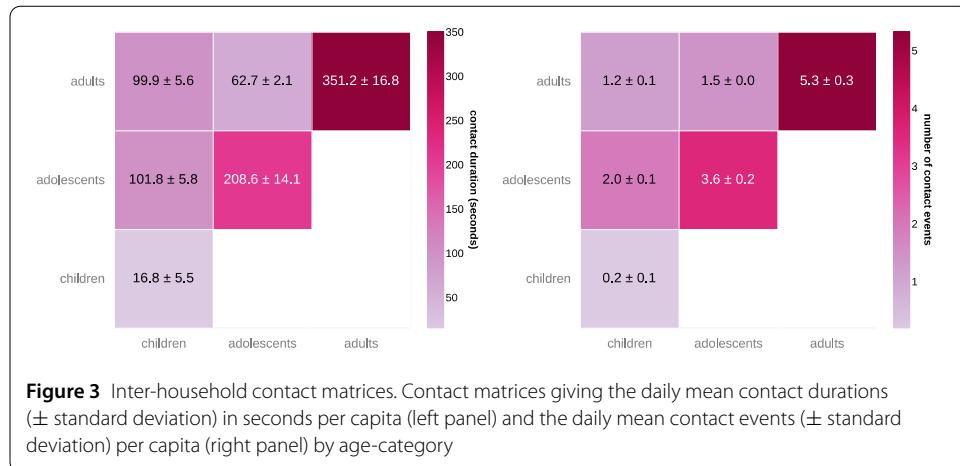
(child, adolescent, adult and HSA) (Panel A) and according to the gender (Panel B), and the edge thickness is proportional to the weight  $w_{ij}$ , that is the total amount of time spent in proximity by two individuals.

### 3.1.2 Intra-household contact data

Overall, 28 households were included in the data analysis, accounting for a total of 84 sensors, distributed as follows: 17 children, 24 adolescents, and 43 adults. The mean number of household members wearing a sensor was 3 (range 2–5). We generated contact matrices based on the duration of contacts (in seconds) and on the number of contacts (contact events) by age-category Fig. 2. The highest mean contact durations and mean number of contacts corresponded to the contacts between adults and children, and between adults and adolescents. However, we did not find significant differences between the empirical values observed between adolescents and adults and those obtained from the null model. On the other hand, the comparison between the empirical values and the null model showed that adults and children had significantly higher mean number of contacts and higher mean contact durations than we expected by chance. The mean contact durations and the mean number of contacts among adults living in the same household were significantly lower than we expected by chance. We did not find significant differences between the empirical values observed and those obtained from the null model for all the other age-categories (see Additional file 1).

### 3.1.3 Inter-household contact data

The aggregated contact network for the participants that had contacts with individuals not belonging to their household had 74 nodes and 264 edges. The nodes are distributed as follows: 10 children, 24 adolescents, 40 adults (one of them had also the role of Health Surveillance assistant (HSA)), and one HSA was external from the village's population. All the adolescents involved in the study had contacts with individuals not belonging to their household. Overall, the median degree (i.e., number of connections with other individuals) was 6 (range 1–33). The median degree of children was 7 (range 1–11) and of adults was 5 (range 1–33). The adolescents had the highest median degree: 8 (range 1–15). As we expected, the individual with the highest number of connections was an HSA (degree 33), however the HSA external from the village had links with 10 people. We also generated contact matrices based on daily mean of the contact durations and daily mean of contact



**Figure 3** Inter-household contact matrices. Contact matrices giving the daily mean contact durations ( $\pm$  standard deviation) in seconds per capita (left panel) and the daily mean contact events ( $\pm$  standard deviation) per capita (right panel) by age-category

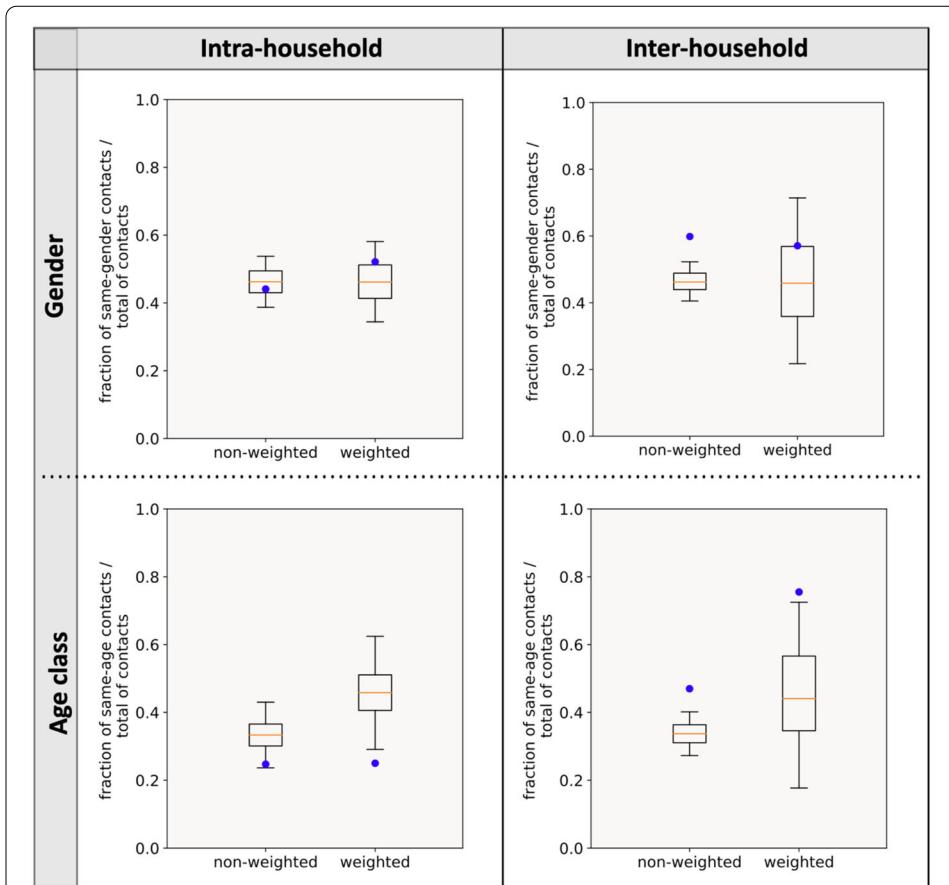
events *per capita* (Fig. 3). The highest mean contact durations and mean number of contacts corresponded to the contacts among adults and among adolescents. The comparison between the empirical values and the null model showed that the mean contact durations and the mean number of contacts among adults living in different households were higher than we expected by chance. We did not find significant differences between the empirical values observed and those obtained from the null model for all the other age-categories (see Additional file 1).

### 3.2 Community detection

We studied the community structure of the aggregated network by considering all the participants, both intra-household and inter-households contacts (86 nodes and 355 edges). Community analysis using Louvain's algorithm showed high modularity in the unweighted network (0.48) and the weighted network (0.75). The algorithm detected 8 communities for the unweighted network and 13 communities for the weighted network. Networks with high modularity have dense connections between the nodes within communities but sparse connections between nodes in different communities. The community membership was strongly correlated with household membership in both the unweighted network ( $NMI = 0.73$ ) and weighted network ( $NMI = 0.82$ ). Community membership had no significant relationship with either the individual's gender (unweighted  $NMI = 0.02$ ; weighted  $NMI = 0.04$ ) or age-category (unweighted  $NMI = 0.03$ ; weighted  $NMI = 0.11$ ).

### 3.3 Assortativity

We computed the fraction of contact events and the fraction of total time in contact involving individuals with same gender and the same age-category, intra and inter-households. We compared the fraction of contacts and the fraction of total time in contact obtained by observed networks, with those obtained from null models of graphs. To do so, we produced 1000 randomized equivalents of observed networks by shuffling individuals' characteristics. Then, we tested for the homophilous behaviour by computing the distribution of the fraction of contacts and fraction of total time in contact, linking nodes of same characteristics in the null model, and comparing it to the empirical value of this ratio. We considered the following hypothesis in relation to a null model: (H0) The observed fraction of contacts and fraction of time in contact involving individuals with same characteristic are included in the 95% confidence intervals of the null distribution.

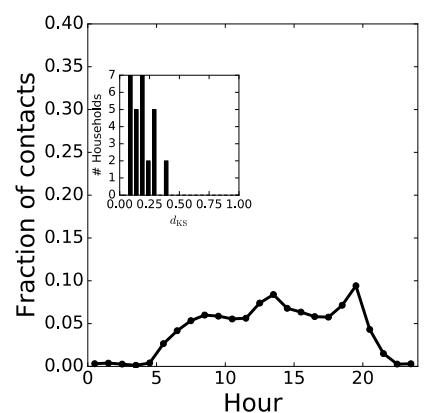


**Figure 4** Statistical test of gender and age assortativity for inter and intra household contacts (non-weighted = fraction of contact events; weighted = fraction of total time in contact). Blue dots indicate the empirical values involving individuals of the same gender and age. Box plots show the distribution of the fraction of contact events (non-weighted) and fraction of total time in contact (weighted) resulting from the null model. In each box, the orange line marks the median and the extremities of the box correspond to the 25 and 75 percentiles: the whiskers give the 5 and 95 percentiles of each distribution

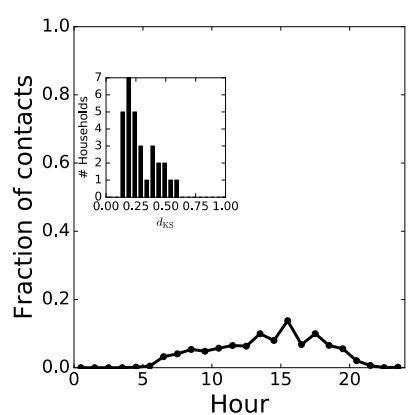
Regarding the gender, we do not reject the null hypothesis H0 for unweighted ( $p\text{-value} = 0.680$ ) and weighted contacts ( $p\text{-value} = 0.424$ ) among individuals living in the same household, and for weighted contacts ( $p\text{-value} = 0.426$ ) among individuals living in different households, the empirical values are included in the 95% confidence intervals of the null distribution. We can reject the H0 for unweighted inter-households contacts ( $p\text{-value} < 0.001$ ), the empirical value is above the 95% confidence intervals of the null distribution, by demonstrating the preference of individuals living in different household to have contacts with individuals of same gender (i.e., gender assortativity) (Fig. 4, upper panels).

Regarding the age-categories, we do not reject the null hypothesis H0 for unweighted contacts ( $p\text{-value} = 0.06$ ) among individuals living in the same households, the empirical values are included in the 95% confidence intervals of the null distribution. We can reject the H0 for weighted intra-household contacts ( $p\text{-value} = 0.006$ ), the empirical value is below the 95% confidence intervals of the null distribution, by demonstrating the preference of individuals living in the same household to spend time with individuals of different age (i.e., age dissimilarity). We can also reject the H0 for unweighted and weighted

**Figure 5** Intra-household daily activity profiles. Number of observed contacts happening at each hour of the day divided by total number of observed contacts. Each thin line corresponds to a different household and the thick line represents the aggregated dataset. Inset: number of households with a specific Kolmogorov–Smirnov distance  $d_{KS}$  to the aggregated dataset



**Figure 6** Inter-household daily activity profiles. Number of observed contacts happening at each hour of the day divided by total number of observed contacts. Each thin line corresponds to a different household and the thick line represents the aggregated dataset. Inset: number of households with a specific Kolmogorov–Smirnov distance  $d_{KS}$  to the aggregated dataset



inter-households contacts ( $p$ -values < 0.001 and 0.023, respectively), the empirical value is above the 95% confidence intervals, by demonstrating the preference of individuals living in different households to interact with individuals of the same age (i.e., age assortativity) (Fig. 4, bottom panels).

### 3.4 Daily activity profiles

We aggregated the observed contacts at intra-household and inter-household level, focusing on the time that they occur along a daily scale. Figures 5 and 6 show the number of contacts happening at each hour divided by the total number of contacts observed. Intra-household contacts showed an intense growth during the early morning (from 5 am) and two activity peaks around lunch time (1 pm) and dinner time (7 pm) (Fig. 5). Likewise, inter-household contacts increased from 5 am, but with a moderate intensity, and showed a peak of activity around 4 pm (Fig. 6).

## 4 Discussion

With the present study, we report the quantitative assessment of contact patterns in a village in rural Malawi, based on proximity sensors technology that allows for high-resolution measurements of social contacts. This technology provided information on community structure of the village, on social relationships and social assortment between individuals, and on daily contacts activity within the village, both intra and inter-

households. We described the social context of the village and the social interactions between individuals with respect to the age-category and to the gender.

#### 4.1 Community detection

Our findings revealed that the social network presented communities that were highly correlated with household membership, thus confirming the importance of family ties within the village. The household variable is the most important one in shaping the sub-groups of the network, while we did not find communities correlates with the age-category and the gender of the participants. The household is usually the fundamental social and economic unit in African villages, where individuals have more frequent and intense interactions (Hosegood and Timaeus [20]), and where a significant part of children's development occurs (Bradley and Putnick [4]). Families not only offer the access to basic necessities, such as food and shelter, but also safeguard a safe environment for young children and adolescents. The quality of care and parenting practices plays a key role in child growth (Engle et al. [14]) and a correct stimulation promotes optimal child development through responsive and appropriate interactions with caregivers (Landry et al. [29]). Moreover, households play a key role in the transmission of infectious diseases, and household composition may influence transmission risks (Fraser [16]; Horby et al. [19]).

Nevertheless, a household is not an independent entity, but it is embedded in the broader structure of the village, where there are kinships and other relationships between individuals. Understanding and quantifying interactions both within and across households can offer a general picture of social life. Children's early life experience is shaped not only by family contexts but also by the social ties formed within the village and social fabric characteristics play an important role in determining child development.

#### 4.2 Network analysis and contact matrices

We obtained contact matrices stratified by age-category on the number of contact events and the time spent in proximity. Our results showed a clear difference between intra and inter-household interaction patterns. Adults had a greater number of contacts and time spent in proximity with children and adolescents living in the same household. This result suggests that the role of adults within the family is the care for the youngest. In addition to age, gender plays an important role in shaping the burden of youth care. In our study, adult women had more interaction with children and adolescents than male (see Additional file 1). In developing countries, women are more likely than men to take on caregiving activities, where mainly elderly women have to cope with the care of grandchildren and children (Schatz [43]; Kalomo and Besthorn [24]). However, we found that the role of men is not negligible, in particular in relation to children. Men in families represent an important resource for children's well-being. In West African countries, children are reared in large extended families with a clan-based kinship centered around a polygynous headman with little parental involvement (Nsamenang [36]). Nonetheless, his role is crucial to guarantee a social position to the young since he provides social connections with the rest of the clan (Nsamenang [36]).

On the other hand, our results showed that most of the inter-household interactions occurred among adults. Kiti et al. [28] found similar results in rural Kenya, where most of the contacts and the total time spent in proximity across households are recorded between adults. We did not find relevant gender differences, and this shows that both adult

men and adult women have social contacts outside the household. Even adolescents had a high number of contact events and high time spent in proximity with individuals of the same age-category not belonging to their household. Moreover, adolescents had the highest median degree (i.e., number of distinct individuals with whom an individual has been in contact) in the inter-household network compared to the other age-categories, and this demonstrates the high sociality of youth individuals. Rock et al. [40] studied how the sociality of adolescents can positively influence their mental and physical health in a poor context in Malawi. However, a survey of sexual partnerships among young adults in several villages of Likoma, an island on Lake Malawi, showed that the high connectivity of social network leads to an increased risk of HIV infections (Helleringer and Kohler [18]).

#### 4.3 Assortativity

At an individual level, we studied the tendency of participants to interact with individuals with whom they shared similar attributes (i.e., assortativity), in particular we tested the influence of the age and the gender. In human societies assortativity is known to be a prime bond formation factor between a pair of individuals (McPherson et al. [34]). Social assortativity might have a large impact on several significant social phenomena like segregation (Leszczensky & Pink [31]), inequality (Karimi et al. [25]), and the transmission of information between groups of individuals (McPherson et al. [34]). The study of assortativity in rural low- and middle-income society is of paramount importance to understand the interactions people experience and which information they share. We observed age and gender assortativity in inter-household network, showing that individuals not belonging to the same family group prefer to interact with people with whom they share similar characteristics. Age disassortativity is instead observed in intra-household networks, this is easily explained by the parenthood and caregiving relationship between adults and youth within the family groups. Similarly, intra-household age assortativity was observed in rural Kenya by Kiti et al. [28], in particular between children aged 6–14 years old, while age disassortativity was observed for adult groups.

#### 4.4 Daily activity profiles

We collected data continuously over 24 hours a day, and we obtained a clear trend of the temporal activity of the village. Family members congregate in the early morning, during lunch time and dinner time. These results agreed with those obtained in a study in rural Kenya (Kiti et al. [28]), thus suggesting a typical family behaviour in African rural villages. A growing contact activity from the morning reaching a maximum in the afternoon in individuals not belonging to the same household was observed. We suppose that people congregate in the afternoon for work or other engagements (e.g., play for the youth). However, we have contact information only for the study participants and we are not aware of any other potential contacts they had with people outside of the study.

### 5 Conclusion

The study has a number of limitations to be considered before findings can be generalized. Overall, it was performed based on data for a relatively small population and only for one village, and the high-resolution contact network data we used spans less than one month. Due to the spatially restricted data collection site and the relatively small participation rate (58.5%), it will not be possible to generalize with certainty these results to other

rural locations or urban localities that have different social, economic, and demographic characteristics. Another limitation is that only bivariate analyses were performed. Future studies should test the multivariate relationships between social interactions and individual's attributes. Despite these limitations, the data collection infrastructure used in this study provided the level of information needed to understand the social context of the village in detail. Proximity sensors captured the dynamics of contacts by collecting high resolution temporal data characterized by timescales comparable with those intrinsic to social dynamics, without the influences of recall bias as instead is the case with paper diaries. Bias due to non-compliance, such as not wearing correctly a sensor and behavior change when participants wore the sensors were potential problems. In our study the participants wore the sensor enclosed in a pouch and pinned to the front of a blouse/shirt in order to minimize attracting attention to themselves. There was relevant anecdotal evidence that these biases did not affect the data collection, and this could be attributed to familiarity as individuals become accustomed to wearing the sensor over time.

The proximity interactions detected by the sensors can provide key information on transmission opportunities of infections within the village. It is well known the extreme vulnerability of African communities to threats of infectious diseases (Fenollar and Mediannikov [15]) and the weaknesses of the disease surveillance systems in several African countries (Marston et al. [33]). Specifically, in Malawi much needs to be done for responses to pandemics with the necessity for substantial improvements in preparedness in key areas such as local surveillance and control guidelines (Sambala and Manderson [42]). The successful transmission of an infectious disease in a population is dependent on many factors, and one of the key factors is the frequency of contacts among infected and susceptible individuals (Horby et al. [19]; Cauchemez et al. [7]). The collection of reliable proximity data can provide crucial information on network properties, such as the presence of superspreaders who are more likely to spread infections based on the number and duration of their interactions. Mathematical models of infectious disease transmission can help to understand the transmission dynamics and to investigate the impact of intervention strategies, and these models generally incorporate age as the key structural feature governing transmission patterns (Mossong et al. [35]). We obtained contact matrices stratified by age that have a crucial role for the evaluation of empirically driven mathematical models that aim to inform interventions strategies. In rural areas of the developing world, access to healthcare is scarce, therefore, it is paramount to be able to understand transmission of communicable diseases and to identify and control measures such as vaccination and social distancing.

Our study represents a first step to track proximity interactions in rural Malawi which we hope will provide the basis for more detailed and expansive studies. Describing and understanding close proximity interactions can provide key information on behavioural patterns of individuals such as social isolation and social conflicts. Previous studies showed how social interaction dynamics can negatively influence mental health leading to depressive symptoms (Elmer & Stadtfeld [13]), including in early adolescence (Pachucki et al. [39]). Although family members are centrally important as a primary source of care and source of support, household-level networks can be informative on negative relationships and family conflicts that might be detrimental to proper development and quality of life of children. Moreover, the social network at village-level can show the existence of separation between individuals or population groups and identify social segregation.

For future work, it would be interesting to consider more inter-connected villages, in order to understand the social contacts and to simulate the spread of infection in a wider and more inter-connected community. A larger and more representative sample would be desirable, and the study design should take into account a longer period of data collection for each village. Future studies should aim at characterizing social contact patterns across different spatial regions in Malawi and elsewhere (e.g., high-income countries), particularly in the urban setting, which is rapidly growing, in order to generate more generalizable insights into network characteristics of different regions.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-021-00302-w>.

**Additional file 1.** Supplementary information (DOCX 758 kB)

### Acknowledgements

We want to thank the Center for Child Development team members: Nicolò Tomaselli, Walter Dzerouniam, Jiajing Feng, Faith Milongo and the UNICEF Malawi Country Office and UNICEF Innovation.

### Funding

The study was funded by UNICEF Malawi. LO, DP, JPR and CC acknowledge support from the Lagrange Project funded by the CRT Foundation.

### Abbreviations

UNICEF, United Nations International Children's Emergency Fund; HAS, Health Surveillance Assistant; MOH, Ministry of Health; ID, identification; NMI, Normalized Mutual Information.

### Availability of data and materials

The data supporting this work is available in a fully anonymised format at: <http://www.sociopatterns.org/datasets/>.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

GL, OBLN and CC conceived of and designed the study. OBLN led the data collection. LO and JPR analyzed the data. LO wrote the manuscript. DP, CC, SH, GL, JPR and OBLN critically revised and edited the draft. All authors read and approved the final manuscript.

### Author details

<sup>1</sup>ISI Foundation, Turin, Italy. <sup>2</sup>Department of Economics, University of Zurich, Zurich, Switzerland. <sup>3</sup>College of Medicine, Lilongwe, Malawi. <sup>4</sup>University of Turin, Turin, Italy.

## Publisher's Note

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

Received: 20 December 2020 Accepted: 30 August 2021 Published online: 08 September 2021

### References

1. Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO (2013) The diffusion of microfinance. *Science* 341(6144):1236498. <https://doi.org/10.1126/science.1236498>
2. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
3. Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. *Science* 323(5916):892–895. <https://doi.org/10.1126/science.1165821>
4. Bradley RH, Putnick DL (2012) Housing quality and access to material and learning resources within the home environment in developing countries. *Child Dev* 83(1):76–91. <https://doi.org/10.1111/j.1467-8624.2011.01674.x>
5. Cassidy L, Barnes GD (2012) Understanding household connectivity and resilience in marginal rural communities through social network analysis in the village of Habu, Botswana. *Ecol Soc* 17(4):11. <https://doi.org/10.5751/ES-04963-170411>
6. 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):e11596. <https://doi.org/10.1371/journal.pone.0011596>

7. Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, Ferguson NM, Swerdlow D, Pennsylvania H1N1 working group (2011) Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc Natl Acad Sci USA* 108(7):2825–2830. <https://doi.org/10.1073/pnas.1008895108>
8. Chami GF, Ahnert SE, Voors MJ, Kontoleon AA (2014) Social network analysis predicts health behaviours and self-reported health in African villages. *PLoS ONE* 9(7):e103500. <https://doi.org/10.1371/journal.pone.0103500>
9. Chikaphupha KR, Kok MC, Nyirenda L, Namakhoma I, Theobald S (2016) Motivation of health surveillance assistants in Malawi: a qualitative study. *Malawi Med J* 28(2):37–42. <https://doi.org/10.4314/mmj.v28i2.2>
10. Contisciani M, Power EA, De Bacco C (2020) Community detection with node attributes in multilayer networks. *Sci Rep* 10:15736. <https://doi.org/10.1038/s41598-020-72626-y>
11. Crampin AC, Floyd S, Ngwira BM, Mwinuka V, Mwaungulu JN, Branson K, Fine PE, Glynn JR (2008) Assessment and evaluation of contact as a risk factor for tuberculosis in rural Africa. *Int J Tuberc Lung Dis* 12(6):612–618
12. Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, Ross JV, Vernon MC (2011) Networks and the epidemiology of infectious disease. *Interdiscip Perspect Infect Dis* 2011:284909. <https://doi.org/10.1155/2011/284909>
13. Elmer T, Stadfeld C (2020) Depressive symptoms are associated with social isolation in face-to-face interaction networks. *Sci Rep* 10(1):1–12. <https://doi.org/10.1038/s41598-020-58297-9>
14. Engle PL, Black MM, Behrman JR, Cabral de Mello M, Gertler PJ, Kapiriri L, Martorell R, Young ME (2007) Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *Lancet* 369(9557):229–242. [https://doi.org/10.1016/S0140-6736\(07\)60112-3](https://doi.org/10.1016/S0140-6736(07)60112-3)
15. Fenollar F, Mediannikov O (2018) Emerging infectious diseases in Africa in the 21st century. *New Microbes New Infect* 26:S10–S18. <https://doi.org/10.1016/j.nmni.2018.09.004>
16. Fraser C (2007) Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* 2(8):e758. <https://doi.org/10.1371/journal.pone.0000758>
17. Funk S, Salathé M, Jansen VA (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J R Soc Interface* 7(50):1247–1256. <https://doi.org/10.1098/rsif.2010.0142>
18. Helleringer S, Kohler HP (2007) Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi. *AIDS* 21(17):2323–2332. <https://doi.org/10.1097/QAD.0b013e328285df98>
19. Horby P, Thai PQ, Hens N, Yen NTT, Mai LQ, Thoang DD et al (2011) Social contact patterns in Vietnam and implications for the control of infectious diseases. *PLoS ONE* 6(2):e16965. <https://doi.org/10.1371/journal.pone.0016965>
20. Hosegood V, Timaeus IM (2005) Household composition and dynamics in KwaZulu Natal, South Africa: mirroring social reality in longitudinal data collection. In: *African households: an exploration of census data*. ME Sharpe, New York, pp 58–77
21. 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(2):e17144. <https://doi.org/10.1371/journal.pone.0017144>
22. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LD, Bekker LG, Wood R (2011) Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. *Am J Epidemiol* 174(1):1246–1255. <https://doi.org/10.1093/aje/kwr251>
23. Johny J, Wichmann B, Swallow BM (2017) Characterizing social networks and their effects on income diversification in rural Kerala, India. *World Dev* 94:375–392. <https://doi.org/10.1016/j.worlddev.2017.02.002>
24. Kalomo EN, Besthorn FH (2018) Caregiving in sub-Saharan Africa and older, female caregivers in the era of HIV/AIDS: a Namibian perspective. *GrandFamilies Contemp J Res Pract Policy* 5(1):6
25. Karimi F, Génais M, Wagner C et al (2018) Homophily influences ranking of minorities in social networks. *Sci Rep* 8:11077. <https://doi.org/10.1038/s41598-018-29405-7>
26. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ (2014) Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. *PLoS ONE* 9(8):e104786. <https://doi.org/10.1371/journal.pone.0104786>
27. Kiti MC, Melegaro A, Cattuto C, Nokes DJ (2019) Study design and protocol for investigating social network patterns in rural and urban schools and households in a coastal setting in Kenya using wearable proximity sensors. *Wellcome Open Res* 22(4):84. <https://doi.org/10.12688/wellcomeopenres.152682>
28. Kiti MC, Tizzoni M, Kinyanjui TM, Koech DC, Munywoki PK, Meriac M, Cappa L, Panisson A, Barrat A, Cattuto C, Nokes DJ (2016) Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Sci* 5:21. <https://doi.org/10.1140/epjds/s13688-016-0084-2>
29. Landry SH, Smith KE, Swank PR (2006) Responsive parenting: establishing early foundations for social, communication, and independent problem-solving skills. *Dev Psychol* 42(4):627–642. <https://doi.org/10.1037/0012-1649.42.4.627>
30. Leal Neto O, Haenni S, Phuka J, Ozella L, Paolotti D, Cattuto C, Robles D, Combining G (2021) Wearable devices and mobile surveys to study child and youth development in Malawi: implementation study of a multimodal approach. *JMIR Public Health Surveill* 7(3):e23154. <https://doi.org/10.2196/23154>
31. Leszczensky L, Pink S (2015) Ethnic segregation of friendship networks in school: testing a rational-choice argument of differences in ethnic homophily between classroom-and grade-level networks. *Soc Netw* 42:18–26. <https://doi.org/10.1016/j.socnet.2015.02.002>
32. Lu H, Halappanavar M, Kalyanaraman A (2015) Parallel heuristics for scalable community detection. *Parallel Comput* 47:19–37. <https://doi.org/10.1016/j.parco.2015.03.003>
33. Marston BJ, Dokubo EK, van Steelandt A, Martel L, Williams D, Hersey S, Jambai A, Keita S, Nyenswah TG, Redd JT (2017) Ebola response impact on public health programs, West Africa, 2014–2017. *Emerg Infect Dis* 23(13):S25–32. <https://doi.org/10.3201/eid2313.170727>
34. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Annu Rev Sociol* 27(1):415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
35. 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 Med* 5(3):e74. <https://doi.org/10.1371/journal.pmed.0050074>
36. Nsamenang AB (2010) Fathers, families, and children's well-being in Africa. In: Lamb ME (ed) *The role of the father in child development*. Wiley, New York, pp 388–412

37. Ozella L, Gauvin L, Carenzo L, Quaggiotto M, Ingrassia PL, Tizzoni M, Panisson A, Colombo D, Sapienza A, Kalimeri K, Della Corte F, Cattuto C (2019) Wearable proximity sensors for monitoring a mass casualty incident exercise: feasibility study. *J Med Internet Res* 21(4):e12251. <https://doi.org/10.2196/12251>
38. Ozella L, Gesualdo F, Tizzoni M, Rizzo C, Pandolfi E, Campagna I et al (2018) Close encounters between infants and household members measured through wearable proximity sensors. *PLoS ONE* 13(6):e0198733. <https://doi.org/10.1371/journal.pone.0198733>
39. Pachucki MC, Ozer EJ, Barrat A, Cattuto C (2015) Mental health and social networks in early adolescence: a dynamic study of objectively-measured social interaction behaviors. *Soc Sci Med* 125:40–50. <https://doi.org/10.1016/j.socscimed.2014.04.015>
40. Rock A, Barrington C, Abdoulayi S, Tsoka M, Mvula P, Handa S (2016) Social networks, social participation, and health among youth living in extreme poverty in rural Malawi. *Soc Sci Med* 170:55–62. <https://doi.org/10.1016/j.socscimed.2016.10.005>
41. Salathé M, Kazandjieva M, Lee JW, Lewis P, Feldman MW, Jones JH (2010) A high-resolution human contact network for infectious disease transmission. *Proc Natl Acad Sci USA* 107(51):22020–22025. <https://doi.org/10.1073/pnas.1009094108>
42. Sambala EZ, Manderson L (2017) Anticipation and response: pandemic influenza in Malawi, 2009. *Glob Health Action* 10(1):1341225. <https://doi.org/10.1080/16549716.2017.1341225>
43. Schatz EJ (2007) "Taking care of my own blood": older women's relationships to their households in rural South Africa. *Scand J Soc Health* 69:147–154. <https://doi.org/10.1080/14034950701355676>
44. Smieszek T, Castell S, Barrat A, Cattuto C, White PJ, Krause G (2016) Contact diaries versus wearable proximity sensors in measuring contact patterns at a conference: method comparison and participants' attitudes. *BMC Infect Dis* 16:341. <https://doi.org/10.1186/s12879-016-1676-y>
45. 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(8):e23176. <https://doi.org/10.1371/journal.pone.0023176>
46. Voirin N, Payet C, Barrat A, Cattuto C, Khanafar N, Régis C, Kim BA, Comte B, Casalegno JS, Lina B, Vanhems P (2015) Combining high-resolution contact data with virological data to investigate influenza transmission in a tertiary care hospital. *Infect Control Hosp Epidemiol* 36(3):254–260. <https://doi.org/10.1017/ice.2014.53>
47. Wilson-Agarwal JK, Ozella L, Tizzoni M, Cattuto C, Swan GJF, Moundai T et al (2019) High-resolution contact networks of free-ranging domestic dogs canis familiaris and implications for transmission of infection. *PLoS Negl Trop Dis* 13(7):e0007565. <https://doi.org/10.1371/journal.pntd.0007565>
48. Yoneki E, Crowcroft J (2014) Epimap: towards quantifying contact networks for understanding epidemiology in developing countries. *Ad Hoc Netw* 13:83–93. <https://doi.org/10.1016/j.adhoc.2012.06.003>

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)