Association Between Oral Microbiota and Close-Range Proximity in a Primary School

Lorenzo Dall'Amico[1], Xiangning Bai[2,3], Sandra Marie Weltzien[4], Simon Rayner[5,6], Daniela Paolotti[1], Isabelle Sylvie Budin Ljøsne[7], Andreas Matussek[2,8], Anne-Sofie Furberg[4,9], Ciro Cattuto[1,*], Christopher Sivert Nielsen[10,11]

December 27, 2024

Abstract

The microbiota is the ensemble of microorganisms inhabiting the human body and it deeply influences human health and well-being. Recent studies showed its interplay with social behavior, suggesting that part of the microbiota might be socially transmissible. In this work, we investigate the association between close-range proximity and the oral microbiota composition in a group of children attending primary school. Unlike most related studies, our cohort comprises non-cohabiting individuals, and we use high-resolution proximity sensors to objectively measure social interactions. Our analysis shows that prolonged interactions between children pairs correlate with a higher similarity between their microbiota. These findings complement what is known for co-habiting individuals, suggesting that microbiota might be transmitted in social settings.

Introduction

The human microbiota is a complex ecosystem of microorganisms inhabiting the human body. It is predominantly composed of bacteria but also includes viruses, archaea, and fungi that engage in a continuous interplay of cooperation and competition [1]. It is known that the human microbiota significantly influences the host's phenotype, energy metabolism, immunity, and, ultimately, human health and well-being [2, 3, 4]. Recent studies also demonstrated its impact on psychological development and social behaviors in humans and animals [5, 6, 7, 8, 9]. As such, the human microbiome is considered the second human genome [10].

Social interactions and human health are deeply intertwined and are known to influence one another. Seminal studies have shown that obesity [11] and depression [12] are shaped by social interactions and can spread in a community. Likewise, a growing body of literature suggests that social interactions can influence the composition of human microbiota [13]. In humans, microbiota transmission can occur along a vertical direction of inheritance – primarily from the mother – and via a horizontal route via interactions with peers [14]. With time, the latter transmission mode dominates, and different dietary habits, environmental factors, and social contexts lead to differences in the microbiota between the child and the parents [15].

Horizontal transmission of microbiota was first documented in various non-human mammals, including mice [16], baboons [17, 18], chimpanzees [19], and lemurs [20], among others. To date, only a handful of studies have investigated the correlations between human microbiota and close-range social interactions. For instance, Ref. [21] showed that the skin microbiota of people moving to a new house rapidly aligns with that of the previous tenants. Similarly, Ref. [22] showed that family members and their dogs living in the same household share skin microbiota. Other studies [23, 24] have shown that it is possible to infer cohabiting couples

[1]ISI Foundation, Turin, Italy; [2]Department of Microbiology, Division of Laboratory Medicine, Oslo University Hospital, Oslo, Norway; [3]Division of Infectious Diseases, Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden; [4]Faculty of Health Sciences and Social Care, Molde University College, Molde, Norway; [5]Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway; [6]Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [7]Department of Food Safety, Norwegian Institute of Public Health, Oslo, Norway; [8]Department of Microbiology, Division of Laboratory Medicine, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [9]Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway; [10]Department of Chronic Diseases, Norwegian Institute of Public Health, Oslo, Norway; [11]Department of Pain Management and Research, Oslo University Hospital, Oslo, Norway. *Corresponding author: ciro.cattuto@isi.it

from their gut microbiota, and Ref. [25] found that spouses have more similar gut microbiota and share more bacterial taxa than siblings. These works suggest that, in adult life, prolonged interactions and similar lifestyles play a predominant role in shaping microbiota. Moving to non-cohabiting individuals, a study in a group of villages in Honduras found greater similarity of the gut microbiota among individuals who had more frequent interactions [26, 27], and Ref. [8] showed that sociability increases the diversity of the human gut microbiota. Ref. [28] investigated the relation between the oral microbiota and the social network in a group of Agta hunters and gatherers in the Philippines. They introduced and identified the *social microbiome* as a portion of the oral microbiome transmitted through social interactions. The authors also observed a higher diversity in the oral microbiota composition of the most central nodes in the social network.

Although the correlation between social interactions and the microbiota structure has been observed in many studies, research in this field is still in its early stages because of the many technical challenges it poses. One challenge is the accurate measurement of close-range proximity relations. In previous studies, contacts were typically defined in terms of known, long-standing relations or proxies of those relations. The most frequent contacts were usually identified by direct observation [18, 19] or using surveys, contact diaries, and self-reported information on interpersonal affinity between individuals [27]. These measurements are subjective and tend to be biased towards long-lasting relations [29, 30]. In many cases [15, 16, 20, 22, 24, 25], membership in a family/household/village was used as a proxy for interaction. While information on these associations is readily available and reliable, using such strong relationships does not allow the study of the microbiota's similarities (or lack thereof) among individuals who have social contacts reaching beyond their closest social circle. One exception is Ref. [28] in which proximity contacts were objectively measured with radio sensing technology.

In this work, we investigate the relationship between social interactions and oral microbiota composition in a group of children attending primary school in Molde, Norway. We focus on short-range proximity interactions happening at school, and we measure them using the high-resolution proximity sensors developed by the `SocioPatterns` collaboration [31, 32]. We analyze the oral rather than the gut microbiota because changes in its composition are more likely to be observed in relatively short time frames [15]. Moreover, based on long-term studies of individuals in a confined environment [33], it was suggested that social influence might impact the oral microbiome more than the gut microbiome. It is also worth mentioning that, although oral microbiota appears to be more easily transmitted, a strong relationship has been shown between oral and gut microbiota [34].

Results

Data description

Our data describe a group of 35 children from two third-grade classes at Nordbyen primary school, Molde, Mid-Norway. During class hours, the children sat in pairs in fixed positions, except for group work and practical teaching sessions. They engaged in outdoor and indoor free activities during breaks and in the after-school care program. This cohort comprises individuals from different families, allowing us to investigate the correlation between close-range social contacts and oral microbiota among people who share a common environment, a school class in our case.

We characterized the participants' oral microbiota by performing saliva sampling during class hours and then applying 16S rRNA gene sequencing. The samples were taken at 8 time points from September to December 2022. Each measurement provides the relative abundance distribution of each taxon species in the host [35]. We evaluate the measurements' stability across time and individuals to identify the presence of clusters or trends. We use the Jaccard similarity coefficient to compare pairs of measurements. The results reported in Figure 1A show that the microbiota composition for each individual is remarkably stable over the sampling time of this study, in agreement with previous evidence at comparable time scales [15, 36]. The inset of Figure 1A further shows that different microbiota measurements of the same person have a significantly higher Jaccard similarity than microbiota measurements taken on different individuals (the Mann-Whitney signed test on the two distributions has a p value below machine precision). The Jaccard similarity acts on a binary version of the abundance vector, accounting only for the lack or presence of a given taxon. In the Supplementary material we investigate other similarity measures that also account for the count values and we obtain similar results for all metrics.

In parallel with the microbiota sampling, we mea-

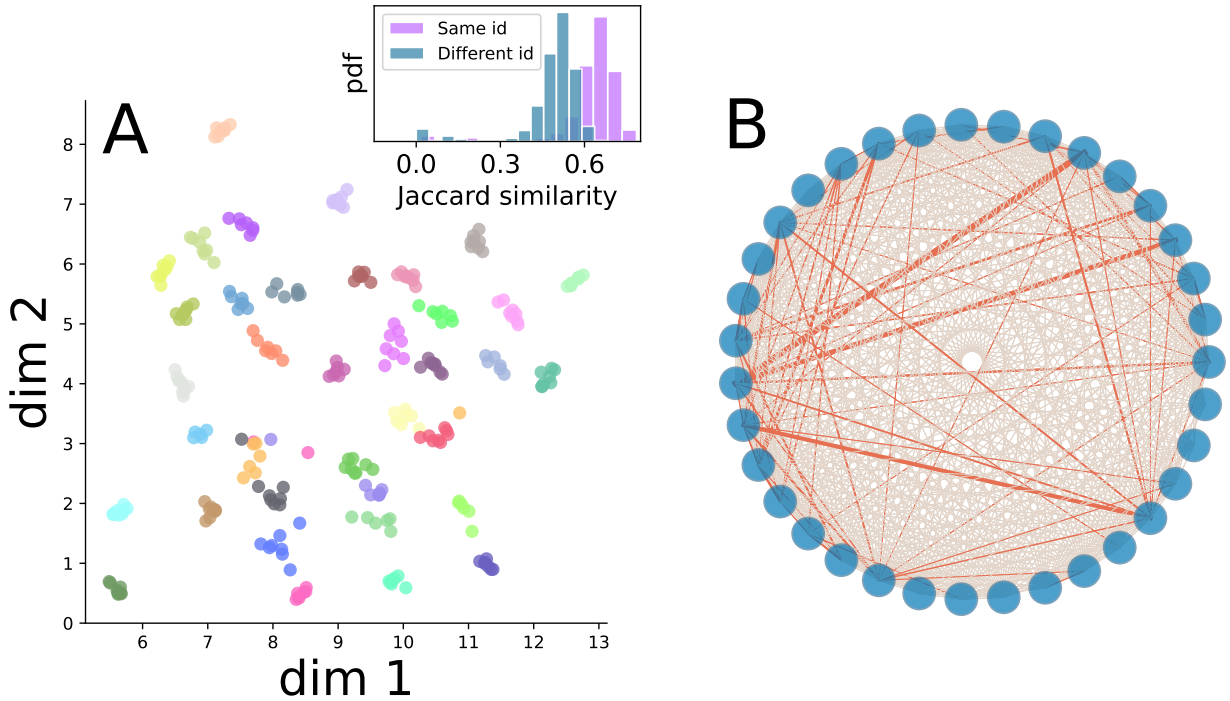

Figure 1: **Microbiota and proximity network measurements**. ***Panel A****: Empirical microbiota measurements.* Every point represents one microbiota measurement for an individual (color-coded). The two dimensional representation is obtained with `UMAP` algorithm [37] and the Euclidean distance in the scatter plot is small for microbiota profiles with high Jaccard similarity. ***Inset of Panel A****: Histogram of microbiota similarity across samples and across individuals.* The two histograms represent the Jaccard similarity between all measurements involving the same person for all individuals ("Same id", purple) and between all measurements involving different individuals ("Different id", blue). ***Panel B****: Aggregated proximity network.* The blue nodes represent children, while the edges stand for interactions between them. Longer cumulative interaction times are represented with thicker lines. We highlighted in red those with a weight exceeding 2.5 hours and in gray all others.

sured a short-range interaction network among children with close-range proximity sensors, developed by the `SocioPatterns` collaboration [31, 32]. These sensors are a well-established, unobtrusive, scalable technology that has been used to quantify human close-range proximity in a variety of real-world settings spanning schools [38, 39, 40], offices [41, 42], hospitals [43] and households in low-resource settings [44, 45, 46, 47], among others. Like the microbiota sampling, the proximity measurements were collected in 3 waves. Due to accumulating damage to the proximity sensors (see Methods section), only the first two waves achieved satisfactory coverage of the participant's contacts. Hence, we restrict our analysis to these two measurement waves and to the five microbiota samples collected during the same period. We encode the proximity network with an undirected, weighted graph, obtained by aggregating all the temporal information collected with the proximity sensors. The graph nodes represent children and the edge weights are the total interaction time between pairs of children over the experiment period. We work with a static graph representation of contacts because it is more representative of recurrent relationships between the children. This choice is further legitimated by the relatively stable nature of microbiota measurements over the considered period, as depicted in Figure 1A. We show the resulting graph in Figure 1B, highlighting in red the edges with cumulative interaction time exceeding 2.5 hours. Figure 2 summarizes the collection schedule for the data used in our analysis.

Microbiota similarity as a function of the social interaction strength

We explore the relationship between the microbiota composition and the proximity network. We aim to test the hypothesis that the microbiota composition is more similar for children pairs with longer interaction times. In a preliminary analysis, we did not observe a strong correlation between microbiota similarity and interaction times. However, such correlation can be observed by adopting a coarser ap-

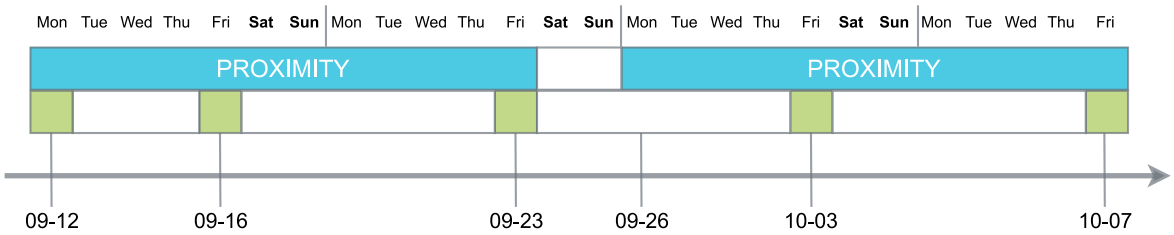

Figure 2: **Data collection timeline relevant for this study.** The arrow at the bottom indicates time and reports the dates in 2023 relevant for data collection. The light blue bars indicate the time intervals during which proximity sensors were used. Saliva samples were collected on the days highlighted in green, on the bottom row.

proach, detailed in the following steps.

1. *Identify the network's strong ties.* We build an unweighted and undirected graph in which every node is a child and an edge connects two nodes if their total interaction time exceeds a threshold τ. We analyze the results as a function of the threshold τ. For $\tau = 0$, all nodes are interacting and there is no network structure.

2. *Measure the microbiota similarity between nodes connected by strong ties.* For each strong tie, we measure the similarity $\langle s \rangle$ between the vectors encoding the taxa abundance distribution in the nodes forming the edge. We show the results for the Jaccard similarity, and test alternative similarity definitions in the Supplementary material.

3. *Compare the results against a null model.* We design a null model in which the association between the edge weights and the microbiota similarities is randomized. This null model preserves the network topology and, in particular, the weight and degree distributions.

Figure 3 reports the result of this analysis, showing the dependence of microbiota similarity $\langle s \rangle$ on the interaction threshold τ. The results are compared against the 90%, and 80% confidence intervals, computed on 1500 realizations of the null model. By restricting to sufficiently strong ties, having $\tau > 2.5$ hours, we observe that the microbiota similarity of interacting pairs is significantly higher than the random null model one, thus confirming our hypothesis. The tests on other similarity metrics described in the Supplementary material show a remarkable agreement between the Jaccard similarity and the Spearman correlation. These two definitions of similarity can deal with the broad distribution of taxa abundance observed in the data. In fact, the Jaccard similarity depends on whether each taxon is present or not, while the Spearman

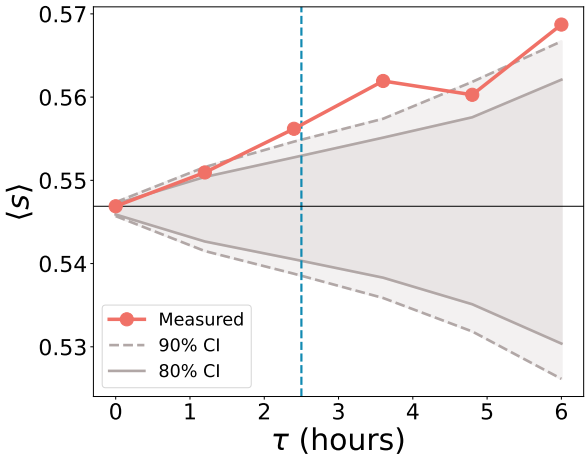

Figure 3: **Average microbiota similarity between node pairs forming strong ties as a function of the definition of strong tie.** The red solid line is the average Jaccard similarity $\langle s \rangle$ computed over the node pairs with cumulative interaction time exceeding τ (expressed in hours). The two style-coded gray lines are the 90%, and 80% confidence intervals obtained by the random null model over 1500 realizations. The vertical dashed blue line is located at 2.5 hours, the threshold used for color coding Figure 1B.

correlation is non-parametric. Lower agreement is observed for the other similarity (or dissimilarity) definitions we investigated, namely the cosine similarity, the Jensen-Shannon divergence, the Bray-Curtis dissimilarity, and the log loss. These definitions depend on the number of measured counts and they reject the null model with a lower confidence compared to the Jaccard similarity and the Spearman correlation. This implies that the presence or lack of a given taxon is more informative than its relative abundance.

Inference of the proximity network from the microbiota

We further explore the relationship between oral microbiota similarity and close-range proximity by

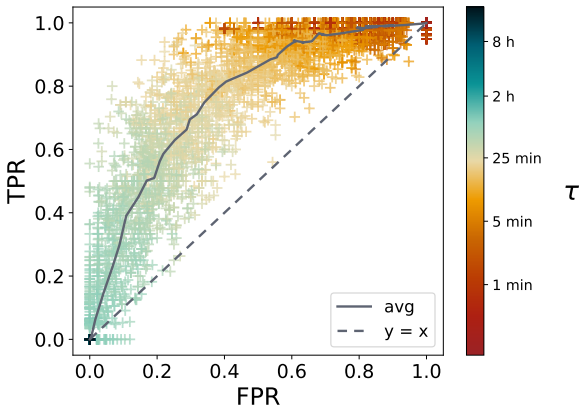

Figure 4: **ROC curve for the binary classifier used to infer the proximity links based on microbiota profiles.** We train a support vector machine to predict whether the weight (duration in proximity) of a given edge (ij) exceeds (or not) a threshold τ. The feature vector used as input is a one-hot encoding of the taxa shared by nodes i and j. The classifier is validated by training on a random selection of 90% of the links and tested on the remaining 10%. 100 random realizations of the training/test split are used, for each of the 250 logarithmically spaced values of τ we scan. Every cross in the plot corresponds to one realization of the training and test sets. The x coordinate is the false positive rate (FPR) and the y coordinate is the true positive rate (TRP). Markers are color-coded according to the value of the threshold τ. The gray solid line (avg) is obtained by averaging, for each value of τ, the TPR and FPR values over all realization of the training/test split and it has a AUC = 0.77. The dashed line is the curve $y = x$ for a random-guess binary classifier.

studying whether we can predict the proximity network using the oral microbiota information alone. Specifically, given a proximity duration threshold τ, we train a binary classifier that predicts, based on the microbiota profiles of two individuals i and j, whether a link between i and j with weight (duration) exceeding τ is present in the measured proximity network. Missing links in the network are treated as zero weight edges. In other words, given a threshold duration τ, we consider the complete graph of relations between individuals, and assign a link (ij) to class "1" if nodes are i and j were observed in proximity for a total duration longer than τ, and to class "0" otherwise. As an input to the classifier, we use a binary feature vector $\boldsymbol{y}$ describing the taxa sharing between nodes i and j: entry y_a of $\boldsymbol{y}$ is 1 if taxon a is shared by i and j and 0 otherwise. For the classifier, we choose a support vector machine [48]. We validate the performance of the classifier by training it on a random sample of 90% of the links and testing it on the remaining 10%. Figure 4 shows the true positive rate (TPR) versus the false positive rate (FPR) for 100 random realizations of the training and test sets, for varying values of the threshold τ. The classifier achieves an area under the curve (AUC) of 0.77, demonstrating that the proximity network can actually be inferred from the microbiota profiles. Consistently with Figure 3, we observe that the false positive rate drops to almost zero for $\tau \approx 2$ h.

Identification of taxa more strongly associated with network structure

In this section, we identify a set of taxa whose co-occurrence in pairs of individuals correlates with sustained close-range physical proximity. We consider a complete graph in which all possible edges are present. Recalling the definition of the vector $\boldsymbol{y}$ introduced in Section Inference of the proximity network from the microbiota, given a taxon a, we assign to each edge a binary attribute $y_a = 1$ if both nodes hosted the taxon a in any measurement and $y_a = 0$ otherwise. We correlate this binary attribute with the cumulative interaction time between the corresponding nodes. Also in this case, we attribute a weight equal to zero to the edges that were not measured by the sensors. Repeating this operation for all taxa a, we identify those for which the interaction duration of the edges formed by nodes sharing taxon a ($y_a = 1$) is significantly longer than for those that do not ($y_a = 0$).

We employ the Mann-Whitney signed test to assess the significance of the difference between the two distributions. This is a non-parametric test that can deal with the broad distribution of the interaction duration. Since it is a signed test, not only can it determine whether the two distributions are different, but also which one is larger. We test two hypotheses: (a) that edges in which the nodes share a taxon ($y_a = 1$) have *larger* weights and (b) that they have *smaller* weights. We expect only hypothesis (a) to provide statistically significant results.

Out of the 603 taxa identified by the 16S analysis, we exclude the ones in which the set of edges in which the nodes share the taxon ($y_a = 1$) is too large or too small. We set a threshold $\delta_1 = 30$ and consider the taxa in which the number of edges with $y_a = 1$ is larger than δ_1 and smaller than $E - \delta_1$, where $E = 595$ is the total number of possible edges. Moreover, to ensure a sufficiently heterogeneous distribution of taxa across the population, we set a second threshold $\delta_2 = 5$ and consider a taxon if it is harbored by at least δ_2 and at most $n - \delta_2$ children, where $n = 35$ is the total number of children. The statistical tests are run independently

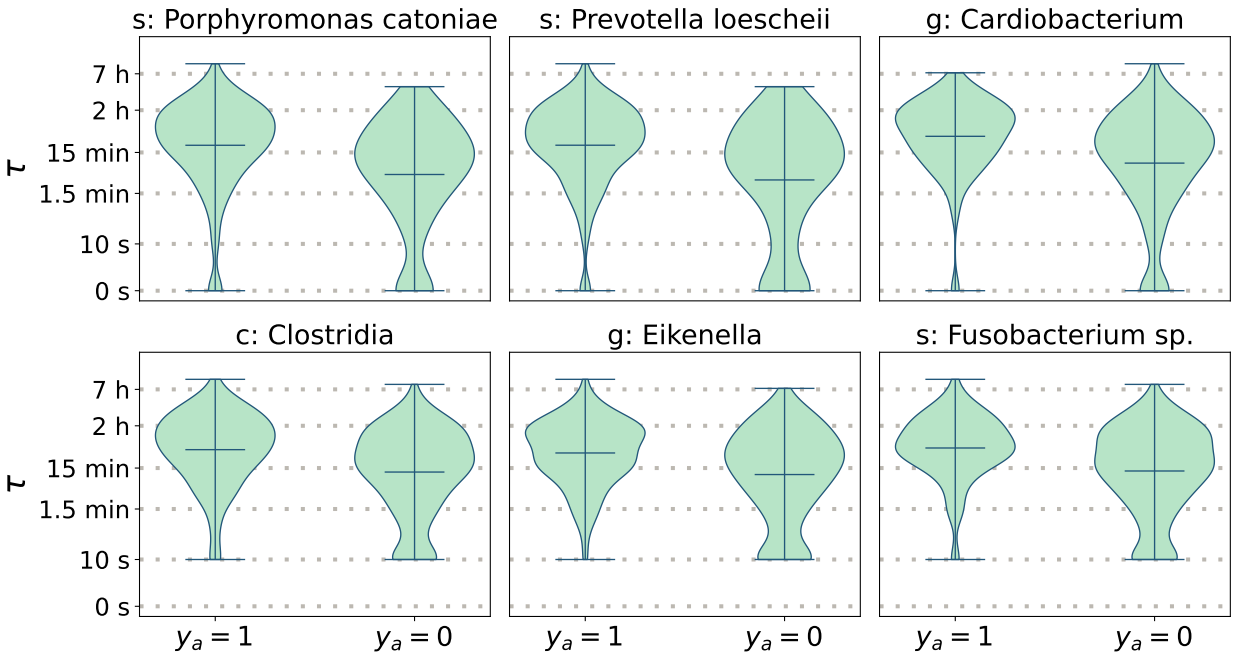

Figure 5: **Time spent in proximity by participants who share or do not share a given taxa**. Each panel corresponds to one taxon and shows in logarithmic scale the histogram of time spent in proximity by participants who share (left-hand violin plot, $y_a = 1$) or do not share (right-hand violin plot, $y_a = 0$) a given taxon. The title reports the last taxonomic level available with "s" being the *species*, "c" the *class* and "g" the *genus*. Note that we imposed $\log(0) = 0$ to enhance readability The plots show the two distributions for the 6 taxa for which the Mann-Whitney test is most significant, after the Bonferroni correction.

on each taxon and do not depend on the thresholds δ_1, δ_2, which only determine the inclusion criterion.

After this selection, we are left with 130 eligible taxa. Given that the test is repeated over all these taxa, we apply the Bonferroni correction to the statistical significance of the Mann-Whitney test. The test shows that edge with $y_a = 1$ have longer durations for several taxa. Figure 5 compares the weight distributions for the edges with $y_a = 1$ and $y_a = 0$ for the taxa in which the Mann-Whitney text p value is smallest. Table 1 reports the last available taxonomic level for all the taxa in which the p value is smaller than 0.05, after the Bonferroni correction. Moreover, we remark that there is no taxon for which the edges with $y_a = 0$ are associated with longer interaction times. These results show the correlation between the microbiota and prolonged close-range physical proximity. We note that some of the genera in Table 1 were also identified as part of the "social microbiome" of [28]. This is particularly the case of *Porphyromonas*, *Prevotella*, *Fusobacterium*, and *Moraxella*.

Association between network centrality and microbiota diversity

Previous studies showed that both in humans [8, 28] and animals [18], individuals with a larger social

p value	Level	Name
1.02E-09	s	Porphyromonas catoniae
1.00E-08	s	Prevotella loescheii
1.87E-08	g	Cardiobacterium
2.06E-06	c	Clostridia
3.62E-06	g	Eikenella
1.82E-05	s	Fusobacterium sp.
7.12E-05	c	Bacilli
1.57E-04	g	Pseudopropionibacterium
2.98E-04	s	Simonsiella muelleri
6.01E-04	c	Bacilli
6.08E-04	s	Cardiobacterium valvarum
1.40E-03	s	Leptotrichia hofstadii
1.89E-03	g	Moraxella
2.13E-03	o	Burkholderiales
2.90E-03	g	Gemella
1.10E-02	g	Fusobacterium
1.22E-02	s	Lachnospiraceae bacterium
1.53E-02	g	Oribacterium
2.63E-02	g	Kingella

Table 1: **List of taxa more strongly associated with the proximity network**. The first column indicates the significance of the Mann-Whitney test after the Bonferroni correction. The second column indicates the last available taxonomic level. The letters indicate "*c*" for class, "*o*" for order, "*g*" for genus, and "*s*" for species. The column "Name" is the taxa identifier.

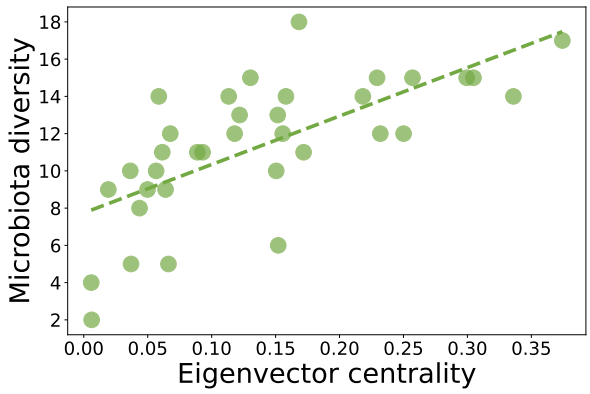

Figure 6: **Association between network centrality and microbiota diversity**. Each dot refers to an individual. The x-axis represents the node eigenvector centrality, while the y-axis the microbbiota diversity expressed as the number of taxa from Table 1 observed in each child. The dashed green line is the result of the linear regression between the two centrality metrics.

network have higher gut and oral microbiota diversity. We investigate this relation with our data, relating the diversity in the microbiota with a measure of centrality in the proximity network. Following Ref. [28], we consider the portion of microbiota associated with the network structure to quantify its diversity. In our case, this is the set of taxa listed in Table 1. We correlate the number of these taxa observed in a child with its eigenvector centrality and report the results in Figure 6 that shows a high correlation between these variables (Spearman $r = 0.76$, p value $= 10^{-7}$). In the Supplementary material, we explore alternative centrality measures and microbiota diversity definitions, obtaining in all cases significant correlations.

Discussion

Our study shows a statistically significant association between the composition of oral microbiota and close-range proximity of non-cohabiting children in a primary school. Our findings are qualitatively consistent with previous studies involving cohabiting individuals [22] and with recent, larger-scale studies focusing on gut or oral microbiome [27, 15] in isolated social settings, including studies using wearable proximity sensors in isolated communities [28].

The distinctive feature of our study is the choice of a highly non-isolated social environment – a primary school in a developed country – and the adoption of a comparatively low-resource approach. We use 16S rRNA gene sequencing and low-cost proximity sensors deployed for just a few weeks in the chosen social setting, without observing any of the contacts children have off-school with household members and their communities.

We remark that while we did measure close-range interactions of children off-school, as children were instructed to wear the proximity sensors also before and after school hours, including the weekend, those interactions have a negligible weight in our analysis. Of course, this might be due to lower compliance with sensor wearing outside of school hours, but it is important to remark that the interaction strengths (durations) we observe are almost exclusively due to close-range proximity within the school environment.

We also show that information about the oral microbiota predicts sustained close-range proximity relations, with a prediction performance similar to that reported by other studies [27] in isolated settings. This suggests that the oral microbiota structure is significantly shaped by social interactions, at least for some taxa. We also identify a set of taxa whose presence in the oral microbiota correlates with prolonged social interaction. Possibly, these taxa are transmitted via social interactions at school, and we remark that some of the taxa of Table 1 have been associated with social interactions by other studies [28]. Moreover, we observe that the most central nodes in the proximity network have a higher number of these taxa: this is consistent with previous studies showing that individuals with a larger number of social interactions have a higher microbiota diversity [8, 28].

On the one hand, our results are consistent with the hypothesis that, at least for some taxa, the oral microbiota can be socially transmitted via sustained close-range proximity at school. On the other hand, it is important to remark that time proximity at school, measured during our observation period, is probably a proxy for the time in proximity spent by the same individuals during the whole school year, as well as for other types of unobserved kinships among children. Since social influence and latent homophily on unobserved attributes – socio-economic attributes, dietary habits, etc. – generally cannot be teased apart, care needs to be taken in interpreting our results as evidence for the social transmission of oral microbiota and, even assuming transmission is at play, it is important to stress that the interactions times we observe should not be regarded as causally relevant for transmission. In particular, the value of the threshold $\tau \approx 2$ hours, desumed from Figure 3, should not be regarded as a time scale for microbiota transmission,

but rather as the definition of a social tie sufficiently strong to be associated, in our study, with the composition of the oral microbiota.

The time scales relevant for our study also need to be discussed. We observe that the microbiota of all individuals is essentially stationary over the 4 months spanning the days on which saliva samples were collected. This is compatible with two opposed scenarios: one in which the time scale of the presumed transmission is much longer than the duration of our study, and another in which this duration is much shorter. In the former case (long time scale needed for transmission), the duration of our study is insufficient for observing any significant microbiota dynamics determined by close-range proximity. In the latter case (short time scale needed for transmission), the microbiota of all individuals may have already achieved a stationary state, because our data collection started months after the beginning of the school year. In both cases, this study only observed a snapshot of a complex process that likely involves latent homophily on several unobserved attributes.

To overcome some of the limitations discussed above, future studies might envision a longer data collection period, with the goal of exposing the temporal dynamics of oral microbiota and identifying the time scales relevant for changes in microbiota determined by close-range proximity. Increasing the sampling frequency or oral microbiota might also be necessary to expose microbiota dynamics, and a more fine-grained approach to characterizing the microbiota composition, going beyond the limitations of 16S rRNA gene sequencing, might be essential. Similarly, to deal with confounders and tease apart oral *microbiota association with social interactions*, such as the one observed in this study, and *microbiota transmission determined by social interactions*, future studies might enroll a larger cohort, possibly including the students' households, and collect richer information about study participants, such as socio-economic attributes, dietary habits, and more.

Methods

Ethics

The Regional Committee for Medical and Health Research Ethics (REK East Norway) reviewed the research protocol and concluded it did not fall under the scope of "The Act on Medical and Health Research", as the aim was not to directly generate knowledge about health and disease. Consequently, no REK approval was required for this study. The project received approval from the Data Protection Officer at Molde University College (September 2022). All re-identifiable data were stored at Molde University College and only de-identified data files were shared with research group members.

Enrollment

Our study was conducted in a Norwegian primary school with 292 children, subdivided into 7 grades. Before recruitment commenced, the details of the study were discussed with the faculty at the primary school to ensure they had a full understanding of what the study entailed. Parents/guardians were invited to a meeting where the study was presented, and questions could be asked. Written information about the study was provided to the parents/guardians in two versions: one for the adults and another adapted for the children, along with a consent form. Children were informed about the study orally in class. Inclusion required consent from both parents/guardians and assent from the child.

We invited 41 children aged 7-8 years (22 boys and 19 girls), 37 participated (21 boys and 16 girls). Due to sensor damaging, however, the proximity data were collected for 35 children that constitute the cohort analyzed in this study. The experiment was conducted from September to December 2022. The children had a five-day school week with four to five class hours daily. During October 2022, the school had an autumn break with the duration of one week. After the data collection was completed, all children in the two classes were invited to an afternoon trip to the local cinema and a class trip to an aquarium in a neighboring city, both financed by the study. No distinction was made between children participating in the study and non-participating children.

Experiment description

One dedicated researcher from our team served as the contact point for the children, their guardians, and the teachers. This researcher organized the collection of materials at the school throughout the whole data collection period, with visits to both classes every one to two weeks. Saliva samples were collected at 8 time points, with a one-week interval between the first five samples and a two-week interval between the last three. The samples were taken from September 12th to November 25th,

2022. Daily use of sensors overlapped this entire period. The proximity sensors had to be worn outside the clothing, making them visible to other children. They had a unique identifier to simplify their identification by each child. To avoid the possible stigma, children who did not participate were given the option to take saliva samples without storing them for analysis and to wear a sensor with the battery removed, i.e., without data recording. Four of the 41 children invited used this opportunity.

There were no compliance issues with either procedure. The children were highly motivated to give saliva samples and to use the sensors throughout the whole study period. Feedback from parents and teachers indicated that the children were self-motivated and independent. They also looked out for each other to ensure they remembered to wear the sensors, e.g., after gym when the sensors had to be taken off. The study did not impact the children's routine, except for $10-15$ minutes set aside for collecting saliva samples on specific days. Our third-grade cohort participated in an after-school care program in the mornings ($07:15-08:20$), and joined fourth-grade children in the program during the afternoons ($13:00-16:45$). Most of the children in our cohort also live near each other and frequently spend time together outside of school hours.

Saliva sampling

The children performed saliva sampling during class hours using the GeneFiX™ Saliva Microbiome DNA collection kits (Isohelix, Cell Projects Ltd, Harrietsham, England)[1] containing sample tubes with 1 ml stabilizing solution and funnel. They were required to not eat or drink during the 30 minutes before the procedure. All the steps in the saliva collection were performed simultaneously by all the children in each class, under the instruction of the dedicated researcher from our team. First, an envelope containing the collection kit and labeled with each child's name was handed out. Next, the children followed the manufacturer's instructions. Finally, the children placed the collection tube back into the envelope and delivered it to the researcher. The collection tube was premarked with a GeneFix™ seven-digit serial number and a barcode, which were linked to each child's study ID and the date of sampling. The saliva samples were stored at room temperature before DNA extraction. All samples were transported overnight from Molde University College to the laboratory at Oslo University Hospital HF in autumn 2022 and received in two batches on October 4th and November 30th.

[1] www.isolhelix.com

16S sequencing and microbiota profiling

DNA from each saliva sample was extracted using the NAxtra™ nucleic acid extraction kit (Lybe Scientific) following the manufacturer's instructions, and then sent to the Norwegian Sequencing Center (Oslo, Norway) for 16S rRNA gene amplification, library preparation, and sequencing. In brief, the V3-V4 region of the 16S rRNA gene was amplified based on the two-step PCR procedure as described in the Illumina application note using primer sequences derived from Ref. [49]. Libraries were pooled and sequenced on an Illumina MiSeq platform (Illumina, San Diego, CA, USA) with 300 bp paired-end reads (v3 reagents). 20% PhiX control library was added to the libraries, and cluster density was reduced to 80% of regular levels. Base calling and production of demultiplexed FASTQ files were performed by running RTA v1.18.54.4 and bcl2fastq v2.20.0.422.

The demultiplexed FASTQ files were analyzed using QIIME2 version 2022.8 run via a SLURM script. The analysis comprised the following steps:

1. Reads were demultiplexed and adapters trimmed using cutadapt with forward and reverse sequences CCTACGGGNGGCWGCAG and GACTACHVGGGTATCTAATCC respectively. To remove low-quality calls at the ends of the reads, adapter-trimmed reads were truncated at 250 bp for the forward read and 190 bp for the reverse read.

2. Denoising and amplicon sequence variant (ASV) inference was carried out using DADA2.

3. Taxonomic classification of the ASVs was carried out using the feature-classifier classify-sklearn command with SILVA release 138 as the reference database.

4. A phylogenetic tree was estimated using the align-to-tree-mafft-fasttree command, which generates a MAFFT alignment from reads, and fasttree was used to estimate a phylogenetic tree from these alignments. The QIIME2 analysis SLURM script is available on GitHub at github.com/CBGOUS/.

Proximity network construction

The proximity network was measured using the `SocioPatterns` wireless proximity sensors [31, 32]. These are wearable, unobtrusive devices embedded in a badge worn by the participants using a lanyard. Each sensor emits low-power radio packets at known power on a radio channel within the ISM radio band at 2.4 GHz, at a rate of approximately 40 packets per second. When two sensors exchange at least one information packet, the receiving sensor records the unique identifier of the sender, the interaction time-stamp, and the signal attenuation value. Proximity interactions within approximately 1.5 meter are obtained by filtering the measurements according to their attenuation. Moreover, each sensor periodically records some status properties that are used during data cleaning to determine the correct functioning of the sensor. Among these, an accelerometer detects whether the sensor is moving or not. This signal can be used to filter out contact data collected while the sensor is not moving which is likely because it was not worn. This technology has been extensively tested and used for quantifying human contacts in a variety of contexts, including schools [38, 40], hospitals [43] and African rural villages [47, 46], among others.

Based on a preliminary analysis, we observed a rapid deterioration in the data quality collected by the proximity sensors over successive waves, primarily due to damage. Consequently, we opted to only utilize the initial two measurements as reliable inputs for our analysis. These two measurements show a high correlation in the contact patterns: the Spearman correlation between the weights is $r = 0.25$ with p value $7 \cdot 10^{-14}$ and the Spearman correlation between the eigenvector centralities is $r = 0.35$ with p value 0.05. These correlations, paired with the high compliance observed in these measurements, make them reliable data sources.

Acknowledgments

We thank the children for their participation in the study, the parents for their invaluable help ensuring that the sensors were worn every day, the rector and teachers for their support and practical assistance in executing the study.

Funding

This project was funded by the Norwegian Institute of Public Health and the Molde University College. LD, CC and DP acknowledge the support from the Lagrange project of Fondazione CRT. LD and CC acknowledge the support from Fondation Botnar (EPFL COVID-19 Real Time Epidemiology I-DAIR Pathfinder).

References

[1] Katharine Z Coyte, Jonas Schluter, and Kevin R Foster. The ecology of the microbiome: networks, competition, and stability. *Science*, 350(6261):663–666, 2015.

[2] Barbara A. Methé et al. A framework for human microbiome research. *Nature*, 486:215 – 221, 2012.

[3] Richard J. Lamont, Hyun Koo, and George Hajishengallis. The oral microbiota: dynamic communities and host interactions. *Nature Reviews Microbiology*, 16:745–759, 2018.

[4] Amar Sarkar, Cameron JA McInroy, Siobhán Harty, Aura Raulo, Neil GO Ibata, Mireia Valles-Colomer, Katerina V-A Johnson, Ilana L Brito, Joseph Henrich, Elizabeth A Archie, et al. Microbial transmission in the social microbiome and host health and disease. *Cell*, 187(1):17–43, 2024.

[5] Elizabeth A Archie and Jenny Tung. Social behavior and the microbiome. *Current opinion in behavioral sciences*, 6:28–34, 2015.

[6] Amar Sarkar, Siobhán Harty, Katerina V-A Johnson, Andrew H Moeller, Elizabeth A Archie, Laura D Schell, Rachel N Carmody, Timothy H Clutton-Brock, Robin IM Dunbar, and Philip WJ Burnet. Microbial transmission in animal social networks and the social microbiome. *Nature ecology & evolution*, 4(8):1020–1035, 2020.

[7] Eoin Sherwin, Seth R. Bordenstein, John L. Quinn, Timothy G. Dinan, and John F. Cryan. Microbiota and the social brain. *Science*, 366, 2019.

[8] Katerina V-A Johnson. Gut microbiome composition and diversity are related to human personality traits. *Human Microbiome Journal*, 15:100069, 2020.

[9] Katerina V-A Johnson, Karli K Watson, Robin IM Dunbar, and Philip WJ Burnet. Sociability in a non-captive macaque population is associated with beneficial gut bacteria. *Frontiers in microbiology*, 13:1032495, 2022.

[10] Elizabeth A Grice and Julia A Segre. The human microbiome: our second genome. *Annual review of genomics and human genetics*, 13(1):151–170, 2012.

[11] Nicholas A Christakis and James H Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

[12] J Niels Rosenquist, James H Fowler, and Nicholas A Christakis. Social network determinants of depression. *Molecular psychiatry*, 16(3):273–281, 2011.

[13] Brett B Finlay and CIFAR Humans. Are noncommunicable diseases communicable? *Science*, 367(6475):250–251, 2020.

[14] Catherine D Robinson, Brendan JM Bohannan, and Robert A Britton. Scales of persistence: transmission and the microbiome. *Current Opinion in Microbiology*, 50:42–49, 2019.

[15] Mireia Valles-Colomer, Aitor Blanco-Míguez, Paolo Manghi, Francesco Asnicar, Leonard Dubois, Davide Golzato, Federica Armanini, Fabio Cumbo, Kun D Huang, Serena Manara, et al. The person-to-person transmission landscape of the gut and oral microbiomes. *Nature*, 614(7946):125–135, 2023.

[16] Aura Raulo, Bryony Allen, Tanya Troitsky, Arild Husby, Josh A. Firth, Tim Coulson, and Sarah C. L. Knowles. Social networks strongly predict the gut microbiota of wild mice. *The ISME Journal*, 15:2601 – 2613, 2020.

[17] Jenny Tung, Luis B Barreiro, Michael B Burns, Jean-Christophe Grenier, Josh Lynch, Laura E Grieneisen, Jeanne Altmann, Susan C Alberts, Ran Blekhman, and Elizabeth A Archie. Social networks predict gut microbiome composition in wild baboons. *elife*, 4:e05224, 2015.

[18] Laura E. Grieneisen, Joshua Livermore, Susan C. Alberts, Jenny Tung, and Elizabeth A. Archie. Group living and male dispersal predict the core gut microbiome in wild baboons. *Integrative and comparative biology*, 57 4:770–785, 2017.

[19] Andrew H Moeller, Steffen Foerster, Michael L Wilson, Anne E Pusey, Beatrice H Hahn, and Howard Ochman. Social behavior shapes the chimpanzee pan-microbiome. *Science advances*, 2(1):e1500997, 2016.

[20] Amanda C Perofsky, Lauren Ancel Meyers, Laura A Abondano, Anthony Di Fiore, and Rebecca J Lewis. Social groups constrain the spatiotemporal dynamics of wild sifaka gut microbiomes. *Molecular Ecology*, 30(24):6759–6775, 2021.

[21] Simon Lax, Daniel P Smith, Jarrad Hampton-Marcell, Sarah M Owens, Kim M Handley, Nicole M Scott, Sean M Gibbons, Peter Larsen, Benjamin D Shogan, Sophie Weiss, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science*, 345(6200):1048–1052, 2014.

[22] Se Jin Song, Christian Lauber, Elizabeth K Costello, Catherine A Lozupone, Gregory Humphrey, Donna Berg-Lyons, J Gregory Caporaso, Dan Knights, Jose C Clemente, Sara Nakielny, et al. Cohabiting family members share microbiota with one another and with their dogs. *elife*, 2:e00458, 2013.

[23] Ashley A. Ross, Andrew Charles Doxey, and Josh D. Neufeld. The skin microbiome of cohabiting couples. *mSystems*, 2, 2017.

[24] Ilana Lauren Brito, Thomas Gurry, Shijie Zhao, Katherine H. Huang, Sarah K. Young, Terrance Shea, Waisea Naisilisili, Aaron P. Jenkins, Stacy D. Jupiter, Dirk Gevers, and Eric J. Alm. Transmission of human-associated microbiota along family and social networks. *Nature Microbiology*, 4:964 – 971, 2019.

[25] Kimberly A Dill-McFarland, Zheng-Zheng Tang, Julia H Kemis, Robert L Kerby, Guanhua Chen, Alberto Palloni, Thomas Sorenson, Federico E Rey, and Pamela Herd. Close social relationships correlate with human gut microbiota composition. *Scientific reports*, 9(1):703, 2019.

[26] Edoardo M Airoldi and Nicholas A Christakis. Induction of social contagion for diverse outcomes in structured experiments in isolated villages. *Science*, 384(6695):eadi5147, 2024.

[27] Francesco Beghini, Jackson Pullman, Marcus Alexander, Shivkumar Vishnempet Shridhar, Drew Prinster, Adarsh Singh, Rigoberto Matute Juárez, Edoardo M Airoldi, Ilana L Brito, and Nicholas A Christakis. Gut microbiome strain-sharing within isolated village social networks. *Nature*, pages 1–9, 2024.

[28] Federico Musciotto, Begoña Dobon, Michael Greenacre, Alex Mira, Nikhil Chaudhary, Gul Deniz Salali, Pascale Gerbault, Rodolph Schlaepfer, Leonora H Astete, Marilyn Ngales, et al. Agta hunter–gatherer oral microbiomes are shaped by contact network structure. *Evolutionary Human Sciences*, 5:e9, 2023.

[29] Timo Smieszek, E U Burri, Ryan Scherzinger, and Roland W. Scholz. Collecting close-contact social mixing data with contact diaries: reporting errors and biases. *Epidemiology and Infection*, 140:744 – 752, 2011.

[30] Rossana Mastrandrea and Alain Barrat. How to estimate epidemic risk from incomplete contact diaries data? *PLoS Computational Biology*, 12, 2016.

[31] sociopatterns.org. Official site of the `SocioPatterns` organization.

[32] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PLoS ONE*, 5, 2008.

[33] Anders Esberg, Linda Eriksson, and Ingegerd Johansson. Site-and time-dependent compositional shifts in oral microbiota communities. *Frontiers in oral health*, 3:826996, 2022.

[34] Thomas SB Schmidt, Matthew R Hayward, Luis P Coelho, Simone S Li, Paul I Costea, Anita Y Voigt, Jakob Wirbel, Oleksandr M Maistrenko, Renato JC Alves, Emma Bergsten, et al. Extensive transmission of microbes along the gastrointestinal tract. *Elife*, 8:e42693, 2019.

[35] Huang Lin and Shyamal Das Peddada. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ biofilms and microbiomes*, 6(1):60, 2020.

[36] Steven N Steinway, Matthew B Biggs, Thomas P Loughran Jr, Jason A Papin, and Reka Albert. Inference of network dynamics and metabolic interactions in the gut microbiome. *PLoS computational biology*, 11(6):e1004338, 2015.

[37] Tim Sainburg, Leland McInnes, and Timothy Q Gentner. Parametric umap embeddings for representation and semisupervised learning. *Neural Computation*, 33(11):2881–2907, 2021.

[38] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE*, 6, 2011.

[39] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PloS one*, 10(9):e0136497, 2015.

[40] Julie Fournet and Alain Barrat. Contact patterns among high school students. *PLoS ONE*, 9, 2014.

[41] Mathieu Génois, Christian L Vestergaard, Julie Fournet, André Panisson, Isabelle Bonmarin, and Alain Barrat. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Network Science*, 3(3):326–347, 2015.

[42] Mathieu Génois and Alain Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7(1):1–18, 2018.

[43] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul a Kim, Brigitte Comte, and Nicolas Voirin. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE*, 8, 2013.

[44] Moses C Kiti, Michele Tizzoni, Timothy M Kinyanjui, Dorothy C Koech, Patrick K Munywoki, Milosch Meriac, Luca Cappa, André Panisson, Alain Barrat, Ciro Cattuto, et al. Quantifying social contacts in a household setting of rural kenya using wearable proximity sensors. *EPJ data science*, 5:1–21, 2016.

[45] Laura Ozella, Daniela Paolotti, Guilherme Lichand, Jorge P Rodríguez, Simon Haenni, John Phuka, Onicio B Leal-Neto, and Ciro Cattuto. Using wearable proximity sensors to characterize social contact patterns in a village of rural malawi. *EPJ Data Science*, 10(1):46, 2021.

[46] Lorenzo Dall'Amico, Jackie Kleynhans, Laetitia Gauvin, Michele Tizzoni, Laura Ozella, Mvuyo Makhasi, Nicole Wolter, Brigitte Language, Ryan G. Wagner, Cheryl Cohen, Stefano Tempia, and Ciro Cattuto. Estimating household contact matrices structure from easily collectable metadata. *PLOS ONE*, 19(3):1–13, 03 2024.

[47] Jackie Kleynhans, SA-S-HTS Group, Lorenzo Dall'Amico, Laetitia Gauvin, M. Tizzoni, Lucia Maloma, S. Walaza, Neil Alexander Martinson, Anne von Gottberg, Nicole L. Wolter, Mvuyo Makhasi, Cheryl Cohen, Carlo Cattuto, and Stefano Tempia. Association of close-range contact patterns with sars-cov-2: a household transmission study. *eLife*, 12, 2023.

[48] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, 1992.

[49] Anna Klindworth, Elmar Pruesse, Timmy Schweer, Jörg Peplies, Christian Quast, Matthias Horn, and Frank Oliver Glöckner. Evaluation of general 16s ribosomal rna gene pcr primers for classical and next-generation sequencing-based diversity studies. *Nucleic acids research*, 41(1):e1–e1, 2013.

Supplementary material

S.1 Sensitivity analysis on the microbiota distance definition

We perform a sensitivity analysis on the definition of microbiota similarity. In the main text, the results are shown for the Jaccard similarity, which only accounts for the lack or presence of each taxon. We extend the analysis shown in the inset of Figure 1, comparing 5 additional definitions of similarity (or dissimilarity) that also account for the relative counts measured for each taxon. The abundance vectors are normalized before the comparison since the counts are expressed up to an unknown normalization factor. Figure S.1 shows that for all similarity (resp. dissimilarity) measures, the microbiota measurements of the same child are more similar (resp. less dissimilar) than the measurements related to different individuals. In all cases, the p value of the Mann-Whitney test applied to the two distributions is below machine precision.

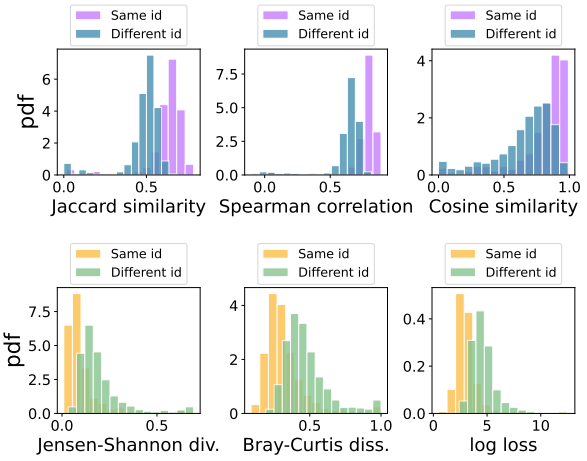

Figure S.1: **Histogram of microbiota similarity across samples and across individuals for different similarity metrics**. Each plot represents the histogram of the similarity (top row) or dissimilarity (bottom row) scores between all measurements involving the same person for all individuals ("Same id") and between all measurements involving different individuals ("Different id"). The x-axis indicates the similarity or dissimilarity definition adopted.

S.2 Sensitivity analysis of microbiota similarity between nodes forming a strong tie

Similarly to the previous section, we explore the results shown in Figure 3 for different definitions of microbiota similarity or dissimilarity. Figure S.2 shows the results for the 6 similarity (or dissimilarity) definitions investigated in Section S.1.

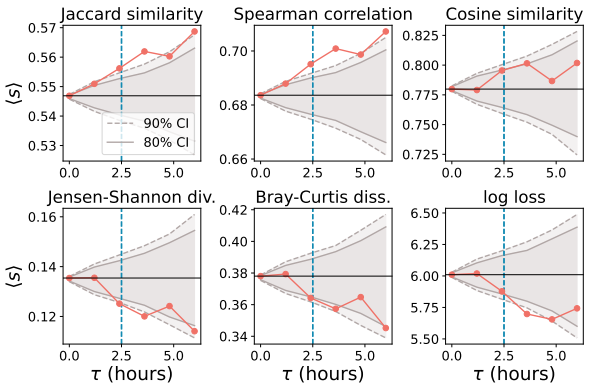

Figure S.2: **Average microbiota similarity between node pairs forming strong ties as a function of the definition of strong tie for different similarity definitions.** In all plots, the red solid line is the average similarity or dissimilarity $\langle s \rangle$ computed over the node pairs with cumulative interaction time exceeding τ (expressed in hours). The similarity definition is indicated in the title of each panel. The two style-coded gray lines are the 90%, and 80% confidence intervals obtained by the random null model over 1500 realizations. The vertical dashed blue line is located at 2.5 hours.

S.3 Sensitivity analysis on the network centrality and microbiota diversity definitions

In the main text, we investigate the correlation between the eigenvector centrality and the number of taxa in the socially transmissible portion of the microbiota. Here, we extend this analysis to other network centrality measures, namely the strength, betwenness, and closeness centralities. The results are reported in Table S.1. In the definition of the betweenness and the closeness centralities, the shortest paths are computed using the reciprocal of the cumulative non-zero interaction times as weights. All the considered centrality measures provide sta-

Centrality measure	r	p value
Strength	0.8	6.48E-09
Eigenvector	0.76	1.23E-07
Betweenness	0.62	6.00E-05
Closeness	0.71	2.24E-06

Table S.1: **Spearman correlation between node centrality and the number of taxa from Table 1 observed in each node, for different centrality definitions.** The first column reports the name of the centrality measure considered. The second and third columns indicate the Spearman correlation and p value, respectively.

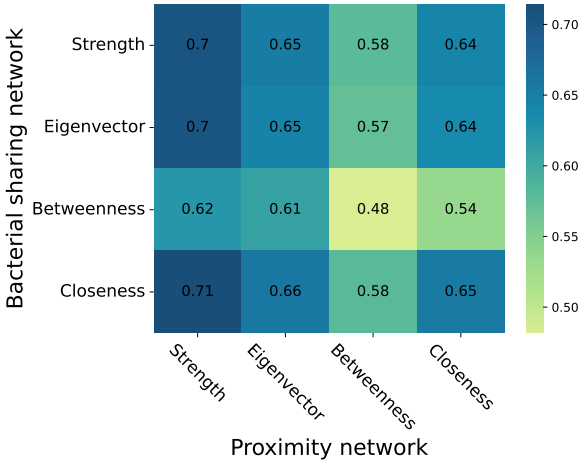

Figure S.3: **Spearman correlation between different node centrality metrics in the proximity and the bacterial sharing networks**. The number and color code indicate the Spearman correlation between the two corresponding centrality measures. The p values are, in all cases, smaller or equal to 0.003.

tistically significant results.

Following Ref. [28], we evaluate an alternative method to assess if the most central nodes act as super spreaders. In Ref. [28], the authors investigate the correlation between the node centralities in the social network and in the *bacterial sharing network*. In this network, nodes are associated with the participants, and edge weights are the number of socially transmissible taxa shared by the nodes forming the edge. Following the same approach, we create the *bacterial sharing network* from Table 1 and analyze the results for different centrality measures. The results are shown in Figure S.3 and they show significant and high Spearman correlations (p value ≤ 0.003 in all cases) for all the considered centrality measures. For the betweenness and closeness centralities of the *bacterial sharing network*, we use the reciprocal of the number of shared taxa to compute the shortest paths.