An Exploration of Trait-Based Interactions in Preschool Social Networks

Complexity72h

Aishvarya Raj *[1], **Alberto Ceria**†[2], **Amina Azaiez**‡[3], **Clara Eminente**§[4], **Govind Gandhi**¶[5], **Guglielmo Ferranti**‖[6], **Gülşah Akçakır****[7], and **Iacopo Iacopini**††[8,9]

[1]The London Interdisciplinary School, London, United Kingdom
[2]Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands
[3]Department of Applied Mathematics, Panthéon Sorbonne University, 75013 Paris, France
[4]Department of Network and Data Science, Central European University, 1100 Vienna, Austria
[5]Department of Informatics, Indiana University Bloomington, IN, USA
[6]Department of Physics and Astronomy "Ettore Majorana", University of Catania, Via Santa Sofia 64, 95123 Catania, Italy
[7]Department of Communication, University of California, Los Angeles, CA, USA
[8]Network Science Institute, Northeastern University London, E1W 1LP London, United Kingdom
[9]Department of Physics, Northeastern University, Boston, MA 02115, USA

June 28, 2024

Abstract

Several studies have been devoted to the investigation of human interaction in different settings and across different age groups leveraging modern tracking techniques for face-to-face encounters. However, little attention has been devoted to understand how individual characteristics are associated to social human behaviour, which could be particularly important in younger age groups due to their potential effects in early childhood development. In this work, we analyse a dataset of human social interactions in a French preschool where face-to-face interactions among children were monitored using proximity sensors over an academic year. We leverage the complementary metadata obtained through parent surveys and linguistic tests at preschool, covering demographic information and home habits, to study the interplay between individual characteristics and contact patterns. Using a mixture of approaches, from random forest classifiers to network-based measures at both dyadic and group level, we identify sex, age, language scores, and number of siblings as the most important variables and find significant associations with interaction patterns in such categories. We examine how these variables influence dyadic interactions in and out of class, as well as in mixed and single-grade classes. At the group level, we investigate how group affinity affects group persistence. In addition, we find that network centrality, measured via hypercoreness, is higher in children that have siblings. This suggests that, despite no difference in total duration of contacts can be found according to possible exposure to sociality at home, children can be still embedded differently in the preschool network. This preliminary study agrees with existing literature on early social development while underscoring the necessity of integrating individual traits into the study of human interactions. Focusing on the 2-5 year old age group can also shed light on how an important phase of human cognitive development

*aishvarya.raj@lis.ac.uk
†a.ceria@liacs.leidenuniv.nl
‡amina.azaiez@univ-paris1.fr
§eminente_clara@phd.ceu.edu
¶gogandhi@iu.edu
‖guglielmo.ferranti@phd.unict.it
**gakcakir@ucla.edu
††iacopo.iacopini@nulondon.ac.uk

interacts with individual characteristics to produce the emergence of social preferences. Future research could leverage these insights to incorporate individual characteristics into mechanistic models of complex social systems.

Keywords Social networks · Preschool · Face-to-face interactions

1 Introduction

The study of early childhood interactions provides valuable insights into the impact of developmental factors on the functioning of human social systems. Through interaction with peers and caregivers, children between the ages of 3 and 6 develop skills in social and emotional regulation McClelland [2007]. Moreover, research in the fields of social-emotional competence and school-readiness Denham [2006], Birch [1997] demonstrate that social competence of children entering kindergarten is a strong predictor for later academic success. Researching social interaction in early childhood can also inform us about the development of social preferences and homophily. Several studies demonstrate that social links are more likely to be formed between individuals who share traits such as gender, race and social status McPherson et al. [2001], Mayhew [1995]. Furthermore, observable and hard-to-fake traits such as accent influence children's friendship and resource allocation preferences Cohen [2013].

Complex systems consisting of a number parts and heterogeneous interactions can be well represented by networks of nodes connected by edges. Since many different natural and synthetic systems can be characterised through individual interacting parts, networks provide a universal model to simulate and analyse them Barrat et al. [2008], Latora et al. [2017]. The study of networks began with graph theory in the field of discrete mathematics Euler [1741]. However, since then, it has found increasing application to a range of systems from the internet to ecosystems Sole and Montoya [2001], the brain Varela et al. [2001], and electric power grids Rosato et al. [2008]. One exciting area of application is to the social sciences, where networks have helped to understand the emergence of norms and institutions, the economy and social contagion Mitchell [1973]. An important factor in this domain is the existence of group interactions which points to the role of higher order networks using hypergraphs to capture a more complete picture of the dynamics to include group formation and dissolution Iacopini et al. [2023], Veldt et al. [2023].

The use of sensors, such as Radio Frequency Identification (RFID) tags, allows comprehensive real-time data to be collected for simultaneous close-range interactions. Furthermore, this method can be used to gain information beyond the formation and dissolution of groups; research in the field of proxemics has uncovered correlations between social and spatial distances. For example, smaller physical distances between interacting individuals often indicate more intimate relationships Cristani et al. [2011]. Within a closed environment such as a classroom or workplace, dyadic interactions for all individuals can be combined to generate a social network, linked through their interactions, where the strength of connections depends on the frequency and duration of interactions. Such networks can subsequently be analysed in order to learn more about an individual's importance or role within the network. The establishment of connections and their strengthenings can be illustrated by a filtering of connections over time, where turnover is quite high (50%) Génois et al. [2019].

Dominant nodes within a network can be identified through node centrality measures. Such measures can also be used to locate and analyse peripheral nodes or in the context of social systems, isolated members of the group. Notably, loneliness tends to spread in networks. However, as isolated individuals become less tied to others in a network, the formation of clusters of isolated individuals occurs at the periphery of the network. Furthermore, the spread of loneliness is prevented by the collective rejection of lonely individuals by the group Cacioppo et al. [2009]. It is often the case that minority individuals tend to be more isolated in networks. When examining founding teams of US entrepreneurs, the increased likelihood for female entrepreneurs to be solo founders compared to men, can be explained by gender-based homophily. Additionally, homophily on the basis of gender, race and status is not solely the result of social preferences; ecological constraints play a significant role in preventing minority individuals from finding similar others Ruef et al. [2003].

The research on social interaction patterns using data collected through these type of technologies often fall short in supplementing contact information with the social identity markers and individual characteristics beyond basic demographics like gender or age. In one such study, investigating the contact preferences of 6–12 years old children in a primary school in France, Stehlé et al. [2013] finds evidence for sex-based homophily. The observation that sex-based homophily increases with age is made more complex when considering weak ties. In this case, sex homophily decrease with grade for girls, yet increases with grade for boys Stehlé et al. [2013]. Furthermore, even when metadata on the participants are available, studies typically focus on one dimension at a time or, at most, interactions between two dimensions. However, relying on multidimensional traits to assess social interactions holds promise of gaining a better understanding of dynamics of behavioral networks Mastrandrea et al. [2015]. Multidimensional traits can reveal complex interdependencies and patterns that single or two-dimension analyses might overlook, providing deeper insights into how various social factors interplay and influence contact preferences at the dyadic and group level.

In addition, the controlled experiments conducted in educational institutions of a higher level with young children or adult participants (e.g., Mastrandrea et al. [2015], Stehlé et al. [2013]) do not fully inform us about the factors contribute to formation and dynamics of behavioral networks in the early childhood period. Considering prominent differences in structure and rules both in and out of class settings (e.g., seating arrangements) on top of overall developmental level, one anticipates to find substantial differences in the preferences. Primary school students and beyond typically spend the entire class time seated near peers, either by choice or through assigned seating arrangements. Preschoolers, on the other hand, are usually freer to move around and interact with a broader range of classmates even in class.

Research on early childhood social dynamics utilising wireless sensor networks has gained some attention in recent years. The collection of movement and speech data with sensing technologies in pre-school settings has been used to learn about the relationship between language development and social interactions Elbaum et al. [2024]. In fact, movement data alone can provide a rich description of social dynamics since this data captures both information on proximity and synchrony, which are distinct indicators used distinguish between friendships and ephemeral interactions to achieve temporary goals Horn et al. [2024], Santos et al. [2015]. Findings on age-related differences in expectations of inter-personal distancing Paulus [2018] show that children develop reason abilities about inter-personal space at an early age, supporting the use of sensor methods in pre-school classrooms. Network approaches to studying peer to peer relationships can reveal links between developmental disabilities such as autism spectrum disorder and lower social connectedness and isolation Chen et al. [2019], Locke et al. [2013]. Similar results were found by Chamberlain et al., in a study on the involvement on autistic children in typical classrooms; children with high functioning autism or autistic spectrum disorder experienced lower centrality, acceptance, companionship, and reciprocity, but not lower levels of lonelinessChamberlain et al. [2007]. In recent work exploring the links between socio-demographic traits and homophily in pre-school children using a range of individual and group-based indices found that children choose to interact similar others, particularly with respect to sex and linguistic development features. Additionally, sex-based homophily increased with age Dai [2022]. Previous work on the structure of social groups has shown that tendencies towards cluster formation differs between children and adults. In particular, it was found that the levels of transitive organisation increased from the age of 3 to 11. This is said to reflect cognitive development in children giving rise to interpersonal preferencesLeinhardt [1973].

The author in Dai [2022] already investigated the relationship between network structure and some additional features by means of homophily, the tendency of individuals to preferably connect to others with similar characteristics McPherson et al. [2001]. In particular, they investigated whether networks aggregated over the timespan of four months showed homophilic behavior with respect to gender, dominant language of the child, occupation category of the mother, occupation category of the father, education level of the mother, education level of the father, vocabulary size and syntactic development level. The dimension showing the most significant level of homophily is gender, especially during time spent out of class. Moreover, they found that the level of homophily is very close to the baseline for all dimensions during class time, probably due to the the fact that during class-time children are assigned fixed seats.

In this study, we investigate the interplay between individual characteristics and social interactions in a French preschool, utilizing time-resolved data on face-to-face proximity collected through wearable sensors over different months throughout the time-span of a year. First, we identify the relevant socio-demographic and linguistic features that exhibit variability among the pupils, and the consistent interaction patterns during both in-class and out-of-class periods. We do this by examining both dyadic and higher-order signatures of social interactions, accounting thus for the existence of social groups of different sizes that be combined in non-trivial patterns. Finally, we analyze the differences in how pupils interact across the different contexts, based on the identified classes of individual characteristics.

2 Materials and Methods

2.1 Dataset

This work utilizes data from the DyLNet project Dai et al. [2022]. The goal of the original project was to observe the co-evolution of social networks and language development of children in pre-school age. The dataset comprises information on a total of 164 children and their interactions within school setting. Social interactions were estimated using spatial proximity sensors. Proximity data was collected in the span of 10 months. Specifically, recordings were made during one week in each of the 10 months and in 9 sessions per week. Alongside the information on interactions, the dataset also contains the metadata collected through a survey administered to the parents and language tests administered to the children. The survey is composed of basic socio-demographic information on the children as well as additional information regarding their attitudes, favorite activities and home environment. The language tests were designed to assess the level of language development of the participants.

2.2 Individual characteristics

Parents of the children participated in the study were asked to fill in questionnaires in order to provide a basic socio-demographic characterization of the children (gender, age) as well as other information regarding their at home environment and activity preferences. Information relevant to assess the link between language and social interactions was collected through the questionnaire and a series of language tests. Parents were asked to indicate the main language spoken in the household and whether the children could understand and speak a language other than French. Additionally, children were administered individual language tests in order to evaluate their receptive lexical skills, short-term memory and their receptive syntactic skills, and the individual scores from these tests were recorded in the dataset.

We first inspected the descriptive statistics of the survey data to identify features that can augment our understanding of social behavior patterns when combined with the high-resolution contact data. Among the participating children, the distribution of sex is comparable, with 81 female and 83 male children. The age gap spans nearly three years, ranging from a minimum of 24 months to a maximum of 59 months. Based on their school level (i.e., 1^{st}, 2^{nd} and, 3^{rd} grade), the children were divided into 7 classes. There are two mixed-grade classes (3 and 5), with one combining 1^{st} and 2^{nd}-graders, and the other combining 2^{nd} and 3^{rd} graders. Figure 1(a) illustrates the age and grade composition of each class. As expected, the mixed classes show the highest age variability, with ages ranging from 26 to 44 months in Class 3 and 41 to 57 months in Class 5. Overall, we observe that sex balance was maintained across the classes, except for one class where there was a 70-30% split, with male students in the majority.

The survey questions regarding children's personality traits were restricted to only two aspects: sociability and talkativeness. Figure 1(b) demonstrates the distribution of parent responses to the question of if the child is social or shy, categorized by sex. We observe that parents perceive a higher rate of male pupils as shy compared to females, whereas females are more likely to be categorized as sociable. The information on talkativeness lacks variability and so is not included in the further analyses. As illustrated in 1(c), the number of siblings is another variable that exhibits variation among the group of pupils who participated in the original study, with majority having at least one sibling.

The dataset also includes two indicators per both lexical and syntactic skills, as well as short-term memory span evaluations administered at two different time points, resulting in a total of 10 measures. The linguistic evaluations include test items specifically designed for the level of each grade and 10 anchor questions presented to the children whichever their grade, and chosen to be rather adapted to 3rd grade pupils. Since the test items measuring the same type of skills are highly correlated, we aggregated them by averaging the sum of distinct scores from the two consecutive years for each skill. The aggregated linguistic scores show a correlation, yet we still observe variation across different score ranges. Conversely, short-term memory span does not correlate with linguistic skills. Unsurprisingly, these developmental measures are also correlated with age but at moderate levels.

Parents were also asked to choose from the proposed activities those that their children prefer, distinguishing between activities done during the day and those done at night. Figure 2a shows the counts of responses to favorites activities. In Figure 2b), the correlation between these activities is displayed. We first notice that the sample of children is quite heterogeneous and most of the variables are poorly correlated. Additionally, some variables exhibit low variability and might not be of interest for this study.

2.3 From proximity data to networks

Spatial proximity and duration information was collected using autonomous RFID Wireless sensors installed on participants. These device tracked face-to-face proximity information with a resolution of $5s$. The experimental set-up included in-situ checks and controlled experiments, in order to validate the information recorded through the sensors (i.e. making sure it corresponds to non spurious interactions). More details on the validation pipeline are given in Dai et al. [2020].

Once validated, the proximity data was aggregated in mutually observed pairs to create an undirected network. For the scope of this study, the data was further aggregated with a time resolution of $10s$, associated to a proximity of participants within 2 meters. Additional checks were performed to be sure that this choice did not affect the fundamental network properties of the resulting contact network.

The obtained network was then used to investigate the social interaction patterns properties of

Then, we assigned to each link, e.g. the one between nodes/children i and j, a corresponding weight w_{ij}, accounting for the total amount of time spent together by the two children. The individual propensity of a node i to interact with its peers is quantified by the node strength s_i, which is the total amount of time spent interacting with any other node. Finally, the number of different individuals that a node i interacts with is quantified by its degree k_i.

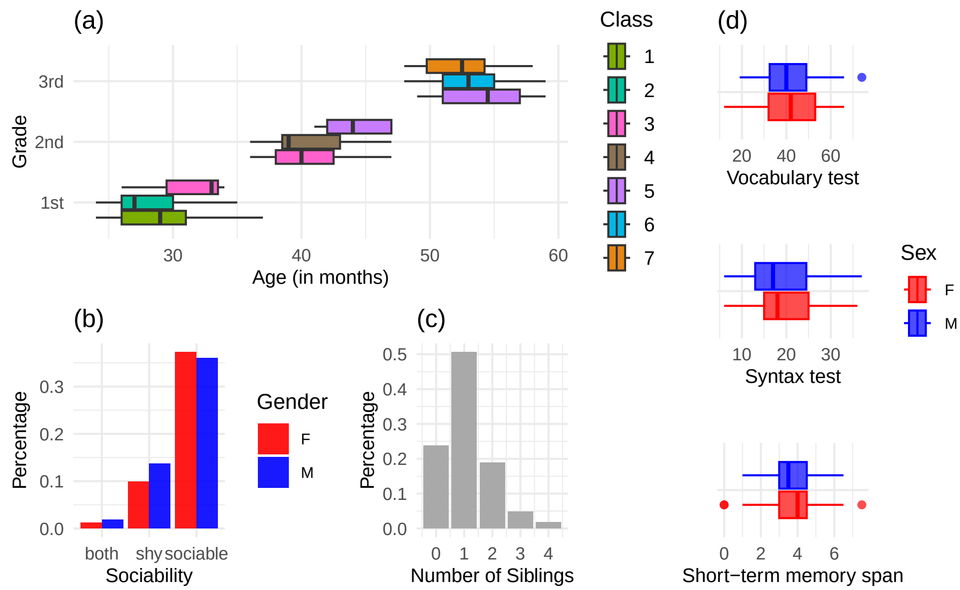

Figure 1: Descriptives of selected features. (a) Boxplots showing the distribution of children's age by their grade and class they were assigned to. Notice the presence of classes with mixed grades. (b) Barplots showing the percentage of answers given by the parents when asked is their child rather sociable or shy. (c) Distribution of number of siblings that the children have. (d) Distributions of aggregated test scores across sex.

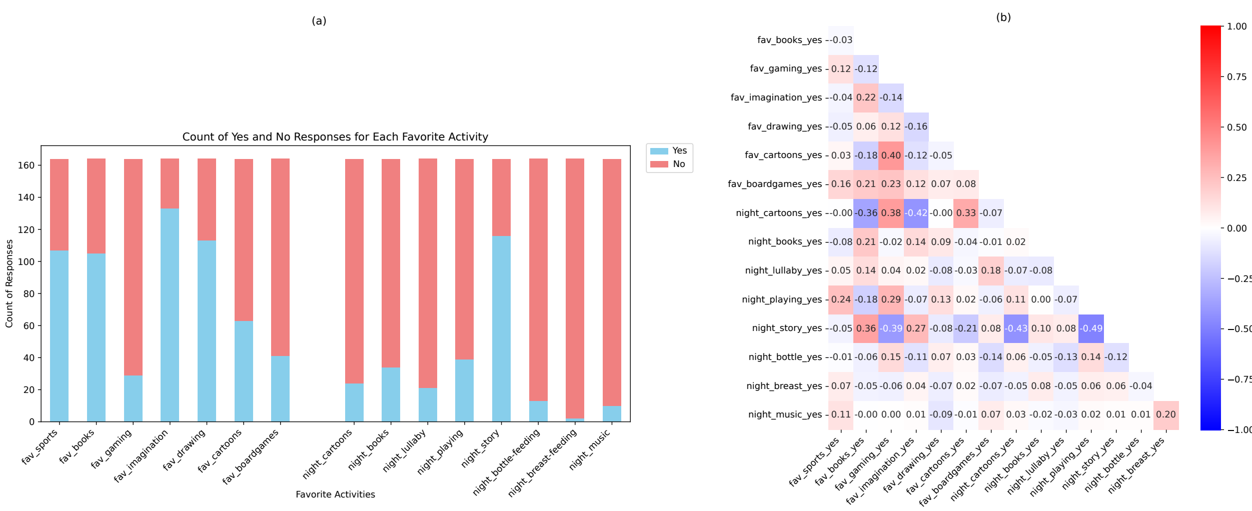

Figure 2: Descriptive analysis of favorite activities. (a) Count of responses to favorite activities. (b) Correlation matrix of favorite activities.

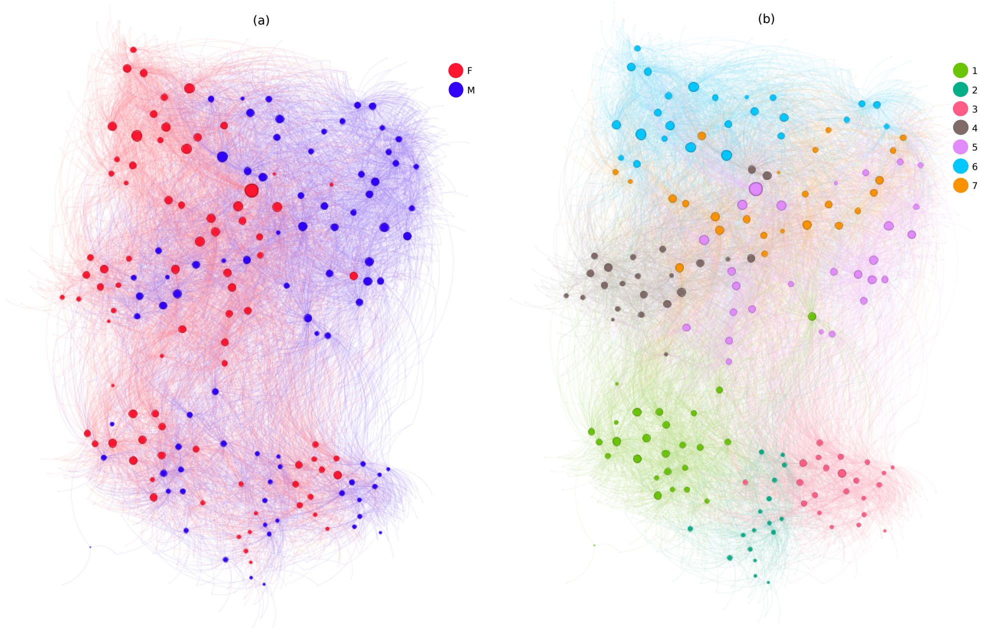

Figure 3: Aggregated network over time in an out-of-class context. The size of the nodes corresponds to the degree. The hypergraph is displayed in a bipartite representation, with nodes representing pupils connected to nodes corresponding to interactions. (a) Blue nodes represent male pupils, and red nodes represent female pupils. (b) Colors indicate class affiliation. The location of nodes remains the same between the two networks.

Figure 3 (a) and (b) shows networks in the out-of-class context aggregated over time where the location of nodes is the same. In Figure 3 (a), nodes are colored according to the sex of the child, 3 (b), nodes are colored according to the class of affiliation. As we can see in Figure 3 (b), the network is clustered, with pupils in the same class interacting preferentially with each other. Additionally, the network is polarized with respect to grade: lower grade classes are located at the bottom, while higher grade ones are located at the top. Furthermore, Figure 3, reveals a clustering according to sex. Interestingly, different behaviors are observed depending on the age of the pupils. Children in classes 1, 2, 3, and 4 tend to cluster first by class, then within their classes by sex. Conversely, children in classes 5, 6, and 7 exhibit a different pattern, clustering first by sex and then by class. This qualitative observed pattern suggests that younger children prioritize class-based groupings over gender, while older children exhibit stronger gender-based clustering, indicating a developmental shift in social dynamics. This change reflects previous work which finds that sex-based homophily increases with age in young children Stehlé et al. [2013], Dai [2022]. Moreover, both younger and older children prefer to associate with others similar in age.

3 Results

In this section we analyse the interplay of individual characteristics with behavioural social patterns starting from the simplest analysis at the level of individual nodes and then moving to more complex dyadic and group measures that take into account the features of the nodes at different categorical levels.

3.1 Classification of interaction time using metadata

As a first step, the distribution of interaction time was divided into four quantile-based classes, emphasizing the first and last quantiles to represent the most and least socially active children, respectively [Figure 4 (a, d)]. A Random Forest classifier [Breiman, 2001] was trained on a subset of the metadata described in section 2.2 to predict the quantile of each pupil, with the results illustrated in Figure 4 (b) and (e). In order to improve the stability of our results, given the relatively small sample size, we employed stratified k-folding [Hastie et al., 2009] for training and averaged the confusion matrices [Figure 4 (b) and (e)] over 20 independent training realizations. The red panels (a, b, c) refer to 'in-class' interactions, while the green panels (d, e, f) pertain to 'out-of-class' interactions. Our analysis reveals that the metadata for each child contains a weak signal, enabling partial differentiation between the first and last quantiles [Figure 4(b) and (e)]. Additionally, we identified the most important features for classification and confirmed the consistency of the top five features (age, school-class and language test scores), as shown in Figure 4 (c) and (f). This analysis suggests that individuals characteristics that do not explicitly include levels of sociality are sufficient to detect the amount of interactions in preschool, in particular for the two most interesting subset of most and less active children.

3.2 Pairwise interaction analysis

From the metadata we see the importance of age in predicting clusters of children, we would like to investigate this relationship independently from the contact patterns of the children. Previous works Veenman [1996], Leroy-Audouin and Suchaut [2007] have looked at the effect of age differences in classrooms on childrens' performance but have not touched upon the social effects and differences associated with it. From a social perspective, the most obvious change in mixed-grade classrooms is that students of different ages are forced to interact with each other. When this restriction is removed, do these children interact differently? We can explore the differences in removing this restriction by looking at the in-class and out-of-class interactions of social networks formed by the classes. We construct an undirected, weighted network from all the in-class interactions happening between children of the same class aggregated over all 10 observed weeks. The weights between two nodes (children) are the total duration they have interacted with each other during the observed period. In Fig. 5(a) we plot the duration of in-class interactions against the age difference of the children involved in the interactions and we see that the interactions are highest for kids with less than half a year's difference but quickly falls off beyond that. We are interested in knowing if the network structure and/or the distribution of weights matter in the observed pattern. We use two reshuffling methods on the networks to test this. The hard reshuffling removes all edges and reassigns them to random pairs of nodes (the Erdos-Renyi null model). This destroys the network structure and we see that it does not follow the duration distribution observed in the data. The soft reshuffling removes all the weights (without changing the network structure) and reassigns them randomly to the edges. This does not change the interaction pattern at all, indicating that the effect of how much you interact with somebody instead of another person in your network pales in comparison to the effect of network structure. We repeat the same construction and reshuffling methods but for out-of-class interactions. While the distribution resembles the in-class interactions, the effect of how much you interact with people in your network matters significantly more relative to the null model. In Fig. 5(b), we see children with less than half a year's age difference interact more than expected compared to the soft reshuffling null model, which preserves network structure while shuffling the weights across edges. The interactions between children of larger age differences are also much less than expected from the soft reshuffling model.

With notable differences, we proceed to ask do children with siblings interact outside their age group more than children without siblings? That is, do children who are exposed to other children over (roughly) a year's age gap through their family tend to show more interactions with varied age groups during school as well? In 5(c), we plot the duration of out-of-class interactions against the age difference of interacting pair of nodes for three cases - 1) both nodes have no siblings, 2) both nodes have siblings, 3) one of them has a sibling. We see no difference across these three classes across interactions with varying age differences except for nodes that have no siblings who interact more with children within a year apart from each other. The plot also indicates that having a sibling does not seem to be associated with children interacting outside their age group.

3.3 Higher-order interaction analysis

So far we have discussed, on the one hand, the interplay between the individual characteristics of a child and their individual propensity to interact, and on the other, the propensity of children of similar ages to spend more time together than those of larger age gaps. This second analysis focused on the amount of time spent together by pairs of children. Children's interactions can involve more than 2 participants Lambiotte et al. [2019], Battiston et al. [2020], Torres et al. [2021], Bick et al. [2023]. Such group (non-dyadic) interactions have been observed in many social settings, such as offices, conferences, primary schools and high-schools, and it has been shown that the higher-order representations offered by hypergraphs as opposed to pairwise graphs can lead to interesting emerging phenomenaBattiston et al. [2021]. Given the individual metadata, it is natural to extend our previous question to ask how individual traits of group members influence the overall time spent by them in a given gathering. To address this question, we deduced the group interactions among children from the proximity data discussed above. Interaction data are recorded in the form of tuples $((i, j), t)$ between pairs of children occurring at a specific time t, where (i, j) is the couple of nodes (children) interacting and t is the time at which the interaction was observed. It is however, possible to deduce the group interactions by looking at the cliques formed by pairwise interactions occurring at the same timestamp, (following the procedure of Cencetti et al. [2021]). For example, if at a specific timestamp t we observe that in the interaction data nodes i, j are all mutually connected by a link, i.e. we observe $((i, j), t)$, $((j, k), t)$ and $((i, k), t)$, then the three nodes are considered as jointly interacting in a group at time t. The so-obtained network is then a higher-order network, where group interactions among children are represented as hyperlinks connecting the nodes. We then assigned to each hyperlink a weight, accounting for the total number of times that the hyperlink was observed in the interaction data.

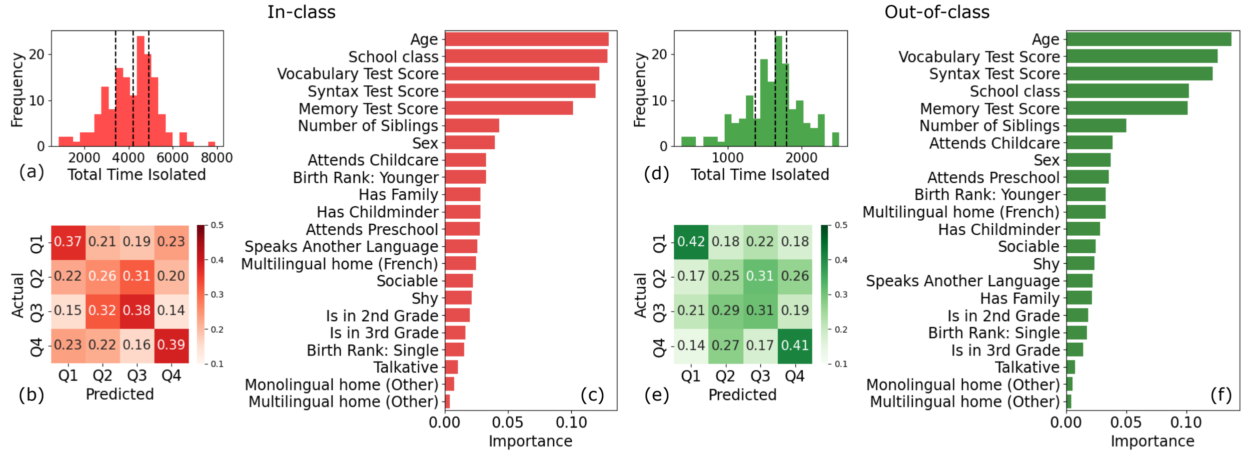

Figure 4: Classifying pupils based on duration of interactions during in-class (a-c) and out-of-class (d-f) time. Plots (a) and (d) show histograms for the interaction time divided into 4 quantiles marked by dotted lines; Confusion matrices (b, e) show our random forest model is able to distinguish the most and least socially active pupils; The bar plots (c) and (f) report the relevance the model assigns to each feature and shows the top 5 features are consistent across in-class and out-of-class context.

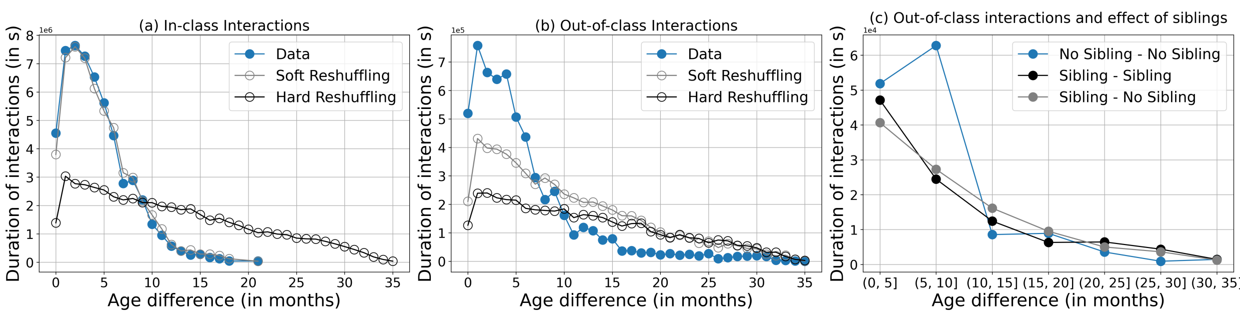

Figure 5: From the weighted interaction network of children aggregated across weeks, the duration of interactions is plotted against age difference of children interacting (a) during class time (b) during free time. The data is compared with 50 realizations each of hard and soft reshuffling models. The hard reshuffling does not preserve degree distribution nor the weight distribution while the soft reshuffling preserves degree distribution while shuffling the weights (duration of interaction) among edges. (c) Duration vs age difference for out-of-class interactions for pairs of nodes where both have siblings, both have no siblings, only one has siblings

Hyperlinks with null weight (i.e., never observed) were not considered in this analysis. We considered two individual characteristics that were included in the analysis of individual and pair activity, i.e. age and results in development skill tests. The results of the tests were normalized such that the score obtained by children is always between 0 and 1. Then, these results were stored in a 3-dimensional vector, where each entry corresponds to the result obtained by the child in the test on vocabulary, syntax and memory. The diversity in age and development skills test scores was quantified as the sum of the absolute distance of each node from the barycentric coordinate, divided by the number of members of the group. Similar to previous analysis, the analysis was split into in-class and out-of-class, to distinguish between the two different social contexts of children social interactions. As shown in Fig.6, we observe that longer-lasting groups tend to be more homogeneous in terms of age and development skill performances. We then investigated the relation between group duration and diversity in the considered individual characteristics of group members by comparing the results obtained from real data with a randomized reference model, where the node similarity of a group with a given number of members is randomly reshuffled with one of another group composed by the same number of children. Regarding in-class context, we see that the reshuffled model displays a decreasing trend of diversity as a function of group duration. This trend is much more evident in the case of age diversity than in the case of test performances. In the out-of-class context, we see clear differences of trends between real data and the reshuffled case. The group diversity after reshuffling indeed remains constant as the group duration increases, differing from the decreasing trend observed in the real data. We conclude that groups with larger duration tend to be less diverse in age and test performances.

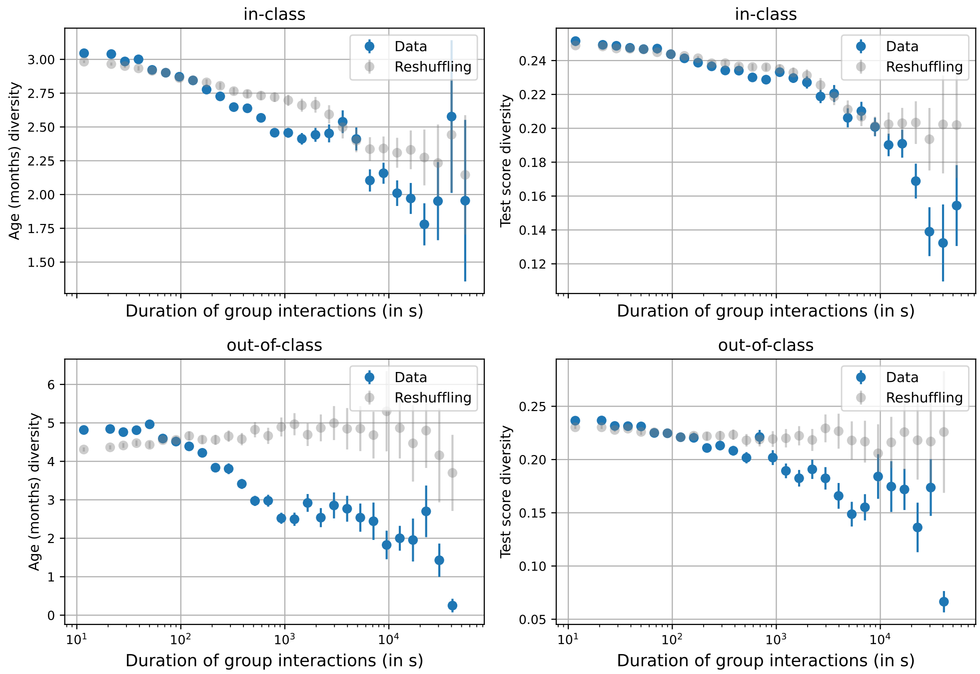

Figure 6: The duration of group interactions obtained from data (blue) and randomized model (yellow) is plotted against diversity in age group (left column) and language development test scores(right column) of children interacting during class (top row) and free time (low row). Results obtained from the random null model are compared with 10 realizations of the randomized null models.

3.4 Node attributes and network measures

We find that the overall the interaction time of female in out-of-class settings is higher than the one of male children. This is shown in Fig 7(b) where the two strength distributions are compared across the two sex categories (significance was checked using a 1-sided Mann–Whitney U test, $p < 0.01$). We also notice that total duration of out-of-class interactions increases with age. This is displayed in Fig 7(a) where the strength distribution for both female and male is plotted for different grade levels. Even if this is not shown in the figure, a temporal analysis shows that this effect seems to be more pronounced in the first quarter in the academic year. This could be due to the differences in the acquaintance network prior to the beginning of the data collection (stronger for 1st graders).

We then turn our attention to language development measures in association with duration of contacts in both in-class and out-of-class settings. We first discard the results of anchor questions —since they are positively correlated with grade, which we have already shown to be correlated with strength. In all settings we find (Fig 7(c-h)) a significant positive correlation for all test scores, in particular for memory and vocabulary tests (which are not correlated with one another).

Finally, we focus on differences in out-of-class contacts for children with and without siblings both in terms of duration of interactions and their patterns (how interactions are arranged across groups). In Fig 7(i) we compare distributions of contact duration across the two categories, finding no significant differences across the two. However, going beyond simple contact duration, differences emerge. This is shown in Fig 7(j), where we compare the hypercoreness values. Hypercoreness is a recently-developed measure of centrality for higher-order networks Mancastroppa et al. [2023] that quantifies the extent of nodes' interactions within groups, considering group size. High hypercoreness values correspond to nodes that interact within many large-sized groups that contain high-degree nodes.

More formally, we start by performing a (k, m)-hypercore decomposition Liu et al., Mancastroppa et al. [2023]. The (k, m)-hypercore decomposition is analogous to a k-core decomposition for graphs, involving the recursive removal of nodes with degree $k_i < k$ and edges with size $m_e < m$. The resulting (k, m)-hypercore is the maximal connected sub-network where all nodes belong to at least k distinct edges and all edges of size at least m. The hypercoreness

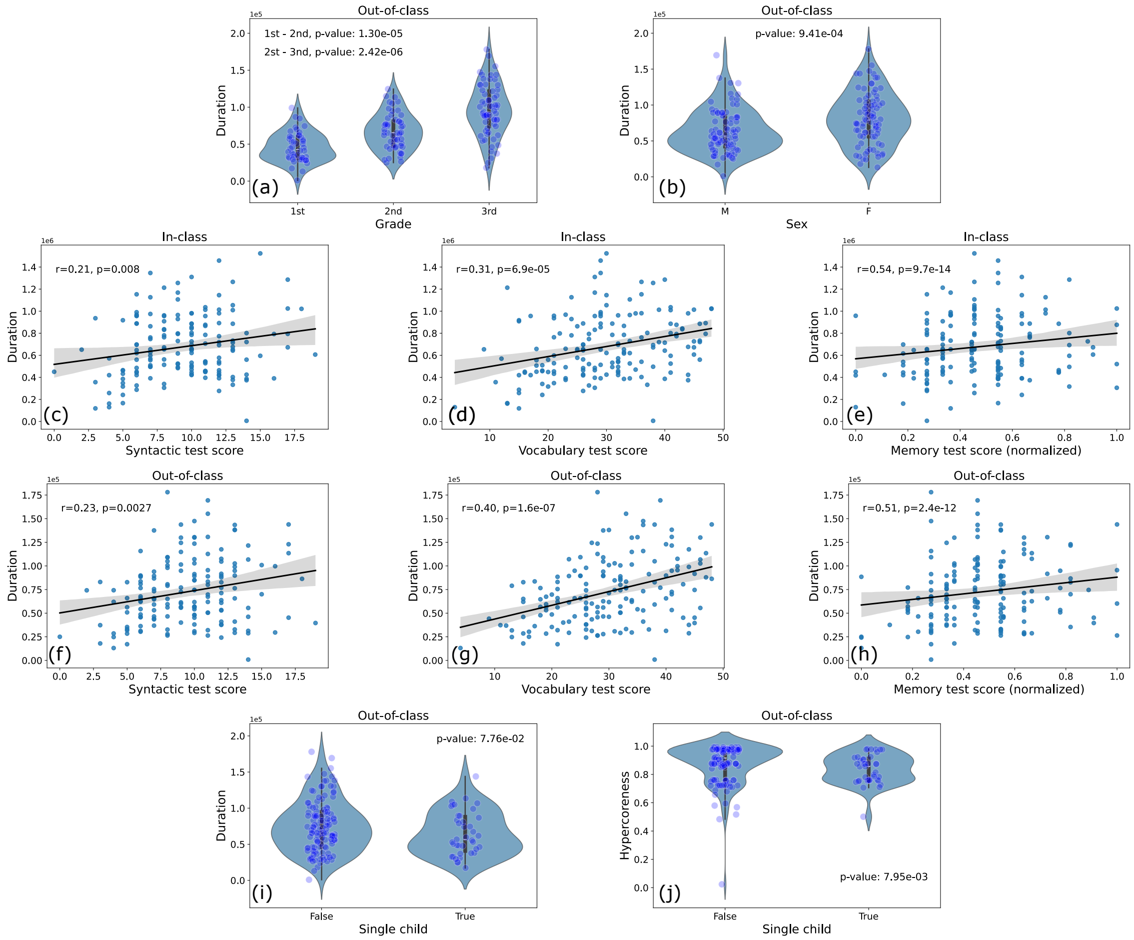

Figure 7: Distribution of contact duration across for children belonging to different grades (a) and sex categories (b). (c-h) Association between duration of contacts and language development test scores in different contexts. Distribution of contact duration (i) and hypercoreness centrality (j) for children with and without siblings.

centrality of a given node i is defined as:

$$R(i) = \sum_{m=2}^{M} \frac{C_m(i)}{N_m k_{max}^m}$$

where the m-core number, denoted as $C_m(i)$, is the value k such that i belongs to the (k, m)-hypercore but not the $(k+1, m)$-hypercore, k_{max}^m is the maximum value of k such that the (k, m)-hypercore is not empty and N_m is the number of edges of size m in the original network.

4 Discussion

We investigated the interplay of individual characteristics of children at preschool with their longitudinal patterns of face-to-face interactions automatically collected across different classes and contexts of interactions. Starting from the amount of social interactions as given by the total time spent in isolation, we showed it is possible to make predictions about the social activity of children using individual characteristics.

We subsequently used random forest methods to identify the key individual traits playing a role in interactions between students in and out of class as well as across mixed-grade classes. The main traits found were age, sex, number of siblings and test scores (vocabulary, syntax and memory scores). We identified an increasing preference with age for children to interact with same-sex others as well as a preference for children to associate with classmates outside of class. Furthermore, children were found to prefer to associate with others similar in age to themselves.

Our investigation into the age-based in-class and out-of class interaction patterns among children reveals that the age difference significantly influences the duration of in-class interactions, with the highest interactions occurring among children with less than half a year's age difference. To understand the role of network structure and weight distribution in these patterns, we employed two reshuffling methods: hard reshuffling (randomly reassigning edges) and soft reshuffling (randomly reassigning weights while preserving network structure). The hard reshuffling did not replicate the observed duration distribution, indicating the importance of network structure, whereas the soft reshuffling showed that interaction duration within the network was less critical. Out-of-class interactions mirrored the in-class patterns but highlighted a significantly greater importance of interaction duration even when the network structure is preserved. Further, we examined whether having siblings affected children's interactions across age groups. The analysis indicated no significant difference in interaction patterns based on sibling presence, except for children without siblings who showed more interactions within their age group. This suggests that exposure to mixed-age interactions in the family does not significantly influence social interactions in school. But more causal work is necessary to make that inference.

On the group level, we found differences across children with or without siblings in terms of centrality in the higher-order networks. Even though no significant differences emerge when comparing group durations, children with siblings display higher hypercoreness, which means that they engage in groups of larger sizes Mancastroppa et al. [2023]. Along this line, an interesting directions to investigate would focus on temporal hypercoreness Mancastroppa et al. [2024], leveraging the longitudinal nature of the dataset, or more complex local interactions patterns as hypermotifs Lotito et al. [2022].

A natural direction to explore in future work is the use of affinity measures for individual characteristics within group interactions based on the higher-order definition recently presented in Veldt et al. [2023], but generalised to account for labels that can take more than two values. More generally, our preliminary study calls for a more comprehensive and exhaustive investigation of homophilic Altenburger and Ugander [2018] —and monophilic Altenburger and Ugander [2018]— patterns of group formation and evolution. These signals could then be used to inform mechanistic model of higher-order social networks. In fact, recent studies have found consistent dynamical patterns of individual group transitions, group formation and disaggregation phenomena in both preschool and university settings during different activity types (in-class, out-of-class, and weekend). The observed phenomena could be replicated by a synthetic model describing the dynamics of individuals forming groups of different sizes and navigating through them, using a mechanism of short-term memory for group duration ("long gets longer" effect) and long-term memory for social contacts. Going beyond these simple signature of recurrent social contact, the individual preferences analysed in this study could be used to complement and further improve these mechanistic dynamics of social interactions.

Despite its prevalence, the French double-grade class system has been criticised and linked to poor performance outcomes due to teachers needing to switch attention between two teaching groups. However some benefits of this system presented were tutoring, imitation and joint supervision. Future work could explore how age differences, number of students and teaching abilities lead to positive or negative interactions and outcomesSuchaut [2010]. In the context of early cognitive development, an important direction of future work could couple neuro-cognitive measures with measures of sociability to learn more about how brain development changes the nature of social interactions.

5 Acknowledgements

This work is the output of the Complexity72h workshop, held at the Universidad Carlos III de Madrid in Leganés, Spain, 24-28 June 2024, `https://www.complexity72h.com`. We thank Alain Barrat for the insightful discussions and contributions during the workshop.

References

Cameron C. E. Connor-C. M.-Farris C. L. Jewkes A. M. Morrison F. J. McClelland, M. M. Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, 43(4):947–959, 2007. doi:https://doi.org/10.1037/0012-1649.43.4.947.

S. A. Denham. Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development*, 17(1):57–89, 2006. doi:https://doi.org/10.1207/s15566935eed1701_4.

Ladd G.W. Bletcher-Sass H. Birch, S. H. The teacher-child relationship and children's early school adjustment: Goodbye's can build trust. *Journal of School Psychology*, 35(1):61–79, 1997. doi:https://doi.org/10.1016/S0022-4405(96)00029-5.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

McPherson J. M. Rotolo-T. Smith-Lovin L Mayhew, B. H. Sex and race homogeneity in naturally occurring groups. *Social Forces*, 74(1):15–52, 1995. doi:https://doi.org/10.1093/sf/74.1.15.

Haun D. Cohen, E. The development of tag-based cooperation via a socially acquired trait. *Evolution and Human Behavior*, 34(1):230–235, 2013. doi:https://doi.org/10.1016/j.evolhumbehav.2013.02.001.

A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008. ISBN 9780521879507. URL `https://books.google.at/books?id=TmgePn9uQD4C`.

V. Latora, V. Nicosia, and G. Russo. *Complex Networks: Principles, Methods and Applications*. Complex Networks: Principles, Methods and Applications. Cambridge University Press, 2017. ISBN 9781107103184. URL `https://books.google.it/books?id=qV0yDwAAQBAJ`.

Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, pages 128–140, 1741.

Ricard V Sole and Ma Montoya. Complexity and fragility in ecological networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1480):2039–2045, 2001.

Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacques Martinerie. The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience*, 2(4):229–239, 2001.

Vittorio Rosato, Limor Issacharoff, Fabio Tiriticco, Sandro Meloni, S Porcellinis, and Roberto Setola. Modelling interdependent infrastructures using interacting dynamical models. *International Journal of Critical Infrastructures*, 4(1-2):63–79, 2008.

J Clyde Mitchell. *Networks, norms and institutions*. Mouton., 1973.

Iacopo Iacopini, Márton Karsai, and Alain Barrat. The temporal dynamics of group interactions in higher-order social networks. *arXiv preprint arXiv:2306.09967*, 2023.

Nate Veldt, Austin R Benson, and Jon Kleinberg. Combinatorial characterizations and impossibilities for higher-order homophily. *Science Advances*, 9(1):eabq3200, 2023.

Marco Cristani, Giulia Paggetti, Alessandro Vinciarelli, Loris Bazzani, Gloria Menegaz, and Vittorio Murino. Towards computational proxemics: Inferring social relations from interpersonal distances. pages 290–297, 2011. doi:10.1109/PASSAT/SocialCom.2011.32.

Mathieu Génois, Maria Zens, Clemens Lechner, Beatrice Rammstedt, and Markus Strohmaier. Building connections: How scientists meet each other during a conference. *arXiv preprint arXiv:1901.01182*, 2019.

John T Cacioppo, James H Fowler, and Nicholas A Christakis. Alone in the crowd: the structure and spread of loneliness in a large social network. *Journal of personality and social psychology*, 97(6):977, 2009.

Martin Ruef, Howard E Aldrich, and Nancy M Carter. The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *American sociological review*, 68(2):195–222, 2003.

Juliette Stehlé, François Charbonnier, Tristan Picard, Ciro Cattuto, and Alain Barrat. Gender homophily from spatial behavior in a primary school: A sociometric study. *Social Networks*, 35(4):604–613, October 2013. ISSN 03788733. doi:10.1016/j.socnet.2013.08.003. URL `https://linkinghub.elsevier.com/retrieve/pii/S0378873313000737`.

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, September 2015. ISSN 1932-6203. doi:10.1371/journal.pone.0136497. URL `https://dx.plos.org/10.1371/journal.pone.0136497`.

Batya Elbaum, Lynn K Perry, and Daniel S Messinger. Investigating children's interactions in preschool classrooms: An overview of research using automated sensing technologies. *Early childhood research quarterly*, 66:147–156, 2024.

Lisa Horn, Márton Karsai, and Gabriela Markova. An automated, data-driven approach to children's social dynamics in space and time. *Child Development Perspectives*, 18(1):36–43, 2024.

António J Santos, Joao R Daniel, Carla Fernandes, and Brian E Vaughn. Affiliative subgroups in preschool classrooms: Integrating constructs and methods from social ethology and sociometric traditions. *PloS one*, 10(7):e0130932, 2015.

Markus Paulus. Preschool children's and adults' expectations about interpersonal space. *Frontiers in Psychology*, 9:400891, 2018.

Jing Chen, Tzu-Jung Lin, Laura Justice, and Brook Sawyer. The social networks of children with and without disabilities in early childhood special education classrooms. *Journal of autism and developmental disorders*, 49:2779–2794, 2019.

Jill Locke, Connie Kasari, Erin Rotheram-Fuller, Mark Kretzmann, and Jeffrey Jacobs. Social network changes over the school year among elementary school-aged children with and without an autism spectrum disorder. *School Mental Health*, 5:38–47, 2013.

Brandt Chamberlain, Connie Kasari, and Erin Rotheram-Fuller. Involvement or isolation? the social networks of children with autism in regular classrooms. *Journal of autism and developmental disorders*, 37:230–242, 2007.

Sicheng Dai. *Study of dynamical social networks of pre-school children using wearable wireless sensors*. Theses, Université de Lyon ; East China normal university (Shanghai), May 2022. URL `https://theses.hal.science/tel-04010766`.

Samuel Leinhardt. The development of transitive structure in children's interpersonal relations. *Behavioral science*, 18 (4):260–271, 1973.

Sicheng Dai, Hélène Bouchet, Márton Karsai, Jean-Pierre Chevrot, Eric Fleury, and Aurélie Nardy. Longitudinal data collection to follow social network and language development dynamics at preschool. *Scientific Data*, 9(1):777, 2022.

Sicheng Dai, Hélène Bouchet, Aurélie Nardy, Eric Fleury, Jean-Pierre Chevrot, and Márton Karsai. Temporal social network reconstruction using wireless proximity sensors: model selection and consequences. *EPJ Data Sci.*, 9(1):19, 2020. doi:https://doi.org/10.1140/epjds/s13688-020-00237-8.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

Simon Veenman. Effects of multigrade and multi-age classes reconsidered. *Review of educational research*, 66(3): 323–340, 1996.

Christine Leroy-Audouin and Bruno Suchaut. Revisiting the pedagogical effectiveness of multigrade classes in france. *Revue francaise de pedagogie*, 160(3):103–118, 2007.

R. Lambiotte, M. Rosvall, and I. Scholtes. From networks to optimal higher-order models of complex systems. *Nat. Phys.*, 2019. doi:10.1038/s41567-019-0459-y.

F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri. Networks beyond pairwise interactions: Structure and dynamics. *Phys. Rep.*, 874:1–92, 2020. doi:10.1016/j.physrep.2020.05.004.

Leo Torres, Ann S Blevins, Danielle Bassett, and Tina Eliassi-Rad. The why, how, and when of representations for complex systems. *SIAM Rev.*, 63(3):435–485, 2021. doi:https://doi.org/10.1137/20M1355896.

Christian Bick, Elizabeth Gross, Heather A Harrington, and Michael T Schaub. What are higher-order networks? *SIAM Rev.*, 65(3):686–731, 2023. doi:https://doi.org/10.1137/21M1414024.

F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, M. Murray, T. Peixoto, F. Vaccarino, and G. Petri. The physics of higher-order interactions in complex systems. *Nat. Phys.*, 17(10):1093–1098, 2021. doi:10.1038/s41567-021-01371-4.

Giulia Cencetti, Federico Battiston, Bruno Lepri, and Márton Karsai. Temporal properties of higher-order interactions in social networks. *Sci. Rep.*, 11(1):1–10, 2021. doi:https://doi.org/10.1038/s41598-021-86469-8.

Marco Mancastroppa, Iacopo Iacopini, Giovanni Petri, and Alain Barrat. Hyper-cores promote localization and efficient seeding in higher-order processes. *Nat. Commun.*, 14:6223, 2023. doi:https://doi.org/10.1038/s41467-023-41887-2.

Boge Liu, Long Yuan, Xuemin Lin, Lu Qin, Wenjie Zhang, and Jingren Zhou. Efficient ($$\alpha $$, $$\beta $$)-core computation in bipartite graphs. 29(5):1075–1099. ISSN 1066-8888, 0949-877X. doi:10.1007/s00778-020-00606-9. URL `http://link.springer.com/10.1007/s00778-020-00606-9`.

Marco Mancastroppa, Iacopo Iacopini, Giovanni Petri, and Alain Barrat. The structural evolution of temporal hypergraphs through the lens of hyper-cores. *arXiv preprint arXiv:2402.06485*, 2024.

Quintino Francesco Lotito, Federico Musciotto, Alberto Montresor, and Federico Battiston. Higher-order motif analysis in hypergraphs. *Communications Physics*, 5(1):79, 2022.

Kristen M Altenburger and Johan Ugander. Monophily in social networks introduces similarity among friends-of-friends. *Nature human behaviour*, 2(4):284–290, 2018.

Bruno Suchaut. Efficacité pédagogique des classes à cours double à l'école primaire: le cas du cours préparatoire. *Revue française de pédagogie. Recherches en éducation*, (173):51–66, 2010.