

---

# EXPLORING THE INTERPLAY OF INDIVIDUAL TRAITS AND INTERACTION DYNAMICS IN PRESCHOOL SOCIAL NETWORKS

---

COMPLEXITY72H

✉ **Gülşah Akçakır**<sup>\*1</sup>, ✉ **Amina Azaiez**<sup>†2</sup>, ✉ **Alberto Ceria**<sup>‡3</sup>, ✉ **Clara Eminente**<sup>§4</sup>, ✉ **Guglielmo Ferranti**<sup>¶5</sup>, ✉ **Govind Gandhi**<sup>||6</sup>, **Aishvarya Raj**<sup>\*\*7</sup>, and ✉ **Iacopo Iacopini**<sup>††8,9</sup>

<sup>1</sup>Department of Communication, University of California, Los Angeles, CA, USA

<sup>2</sup>Department of Applied Mathematics, Panthéon Sorbonne University, 75013 Paris, France

<sup>3</sup>Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

<sup>4</sup>Department of Network and Data Science, Central European University, 1100 Vienna, Austria

<sup>5</sup>Department of Physics and Astronomy “Ettore Majorana”, University of Catania, 95123 Catania, Italy

<sup>6</sup>Department of Informatics, Indiana University Bloomington, IN, USA

<sup>7</sup>The London Interdisciplinary School, London, United Kingdom

<sup>8</sup>Network Science Institute, Northeastern University London, E1W 1LP London, United Kingdom

<sup>9</sup>Department of Physics, Northeastern University, Boston, MA 02115, USA

June 28, 2024

## ABSTRACT

Several studies have investigated human interaction using modern tracking techniques for face-to-face encounters across various settings and age groups. However, little attention has been given to understanding how individual characteristics relate to social behavior. This is particularly important in younger age groups due to its potential effects on early childhood development. In this study, conducted during the Complexity 72h Workshop, we analyze human social interactions in a French preschool, where children’s face-to-face interactions were monitored using proximity sensors over an academic year. We use metadata from parent surveys and preschool linguistic tests, covering demographic information and home habits, to examine the interplay between individual characteristics and contact patterns. Using a mixture of approaches, from random forest classifiers to network-based metrics at both dyadic and higher-order (group) levels, we identify sex, age, language scores, and number of siblings as the variables displaying the most significant associations with interaction patterns. We explore these variables’ relationships to interactions within and outside classrooms and across mixed and single-grade classes. At the group level, we investigate how group affinity affects group persistence. We also find that higher-order network centrality (hypercoreness) is higher among children with siblings, indicating different group embedding despite similar total contact duration. This study aligns with existing literature on early social development and highlights the importance of integrating individual traits into the study of human interactions. Focusing on 2-5-year-olds offers insights into emerging social preferences during critical phases of cognitive development. Future research could use these findings to enhance mechanistic models of complex social systems by incorporating individual traits.

---

\*gakcakil@ucla.edu

†amina.azaiez@univ-paris1.fr

‡a.ceria@liacs.leidenuniv.nl

§eminente\_clara@phd.ceu.edu

¶guglielmo.ferranti@phd.unict.it

||gogandhi@iu.edu

\*\*aishvarya.raj@lis.ac.uk

††iacopo.iacopini@nulondon.ac.uk

# 1 Introduction

Complex social systems, consisting of numerous interacting individuals, can be effectively represented by networks of nodes connected via edges Albert and Barabási [2002], Newman [2003], Latora et al. [2017]. Since many different natural and man-made systems can be characterised through individual interacting parts, complex networks have emerged as a universal representation to analyse and model them Barrat et al. [2008], Vespignani [2012]. While the study of networks began with graph theory in the field of discrete mathematics Euler [1741], it has increasingly been applied to a wide range of systems. In the social sciences Wasserman and Faust [1994], Castellano et al. [2009], networks have helped to understand, among other things, the emergence of norms Baronchelli [2018], cooperation Axelrod and Axelrod [1984], the dynamics of social contagion Mitchell [1973], Centola and Macy [2007], and the formation of opinions and consensus [Nowak et al., 1990, Sznajd-Weron and Sznajd, 2000].

Over the past decades, advancements in automated data collection technologies have enhanced researchers' ability to feed empirical data into theoretical approaches, allowing the integration of insights obtained from observational studies of social systems into mathematical modeling. In particular, sensors like Radio Frequency Identification (RFID) tags enable the collection of comprehensive real-time data on close-range or face-to-face interactions Isella et al. [2011], Barrat et al. [2014], Mastrandrea et al. [2015a]. Research in the field of proxemics has uncovered correlations between social and spatial distances, where a closer physical distance between interacting individuals often indicates a more intimate relationship Cristani et al. [2011]. More generally, taken as a proxy for social interactions, these longitudinal data of dyadic contacts between individuals can then be studied within the framework of temporal (social) networks Holme and Saramäki [2012], where the strength of connections can map the frequency or the duration of interactions Génois et al. [2019]. The temporally-resolved nature of these empirical data collections enables also higher-order network representations where dyadic interactions are replaced by groups Battiston et al. [2020, 2021]—and links become hyperlinks connecting an arbitrary number of nodes Torres et al. [2021], Bick et al. [2023]. Such networks can subsequently be analysed in order to learn more about an individual's importance or role within the network, in particular with respect to the context of interaction and the research question behind the study. For example, dominant nodes within a network can be identified through network 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.

Significant efforts have been put forward towards developing more realistic models of social interaction starting from data collected through these types of technologies Stehlé et al. [2010], Zhao et al. [2011], Perra et al. [2012], Starnini et al. [2013], Vestergaard et al. [2014], Karsai et al. [2014], Nadini et al. [2018], Le Bail et al. [2023], both in pairwise and non-dyadic settings Petri and Barrat [2018], Gallo et al. [2024a], Iacopini et al. [2023]. Despite the success of these data-driven modeling approaches, they often neglect the individual characteristics that can influence the observed social dynamical. The integration of heterogeneous traits at the level of nodes could instead lead to a better characterisation of the driver of social interactions, enhancing our understanding of these systems and our ability to inform realistic models.

While research on social interaction patterns often falls short of supplementing contact information with social identity markers and individual characteristics beyond basic demographics like gender or age, there are exceptions Kontro and Génois [2020], Dai et al. [2022], Génois et al. [2023]. In one such study Stehlé et al. [2013], investigating the contact preferences of 6–12 years old children in a primary school in France led to 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. [2015b]. 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 this study, we investigate the interplay between individual characteristics and social interactions in a French preschool Dai et al. [2022], utilizing time-resolved data on face-to-face proximity collected through wearable sensors over different months throughout the time-span of a year. The study of early childhood interactions provides valuable insights into how developmental factors impact the functioning of human social systems. In fact, through interaction with peers and caregivers, children between the ages of 3 and 6 develop skills in social and emotional regulation McClelland et al. [2007]. Moreover, research in the fields of social-emotional competence and school-readiness demonstrate that social competence of children entering kindergarten is a strong predictor for later academic success Denham [2006], Birch and Ladd [1997]. Getting a deeper understanding of 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 et al. [1995]. Further insights can thus be gained by examining how these phenomena emerge during the critical period

when children develop social skills and are influenced by their peers. Controlled experiments conducted in educational institutions of a higher level with young children or adult participants (e.g., Mastrandrea et al. [2015b], Stehlé et al. [2013]) do not fully inform us about the factors contributing 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 loneliness Chamberlain 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 preferences Leinhardt [1973].

In the following report, 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 look for differences in interaction patterns when accounting for individual traits.

## 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 outside of school, and home environment. The language tests were designed to assess the level of language development of the participants.

The authors of Ref. Dai [2022] already investigated the relationship between network structure and some additional features by means of homophily. 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. They found that 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 fact that during class-time children are assigned fixed seats.

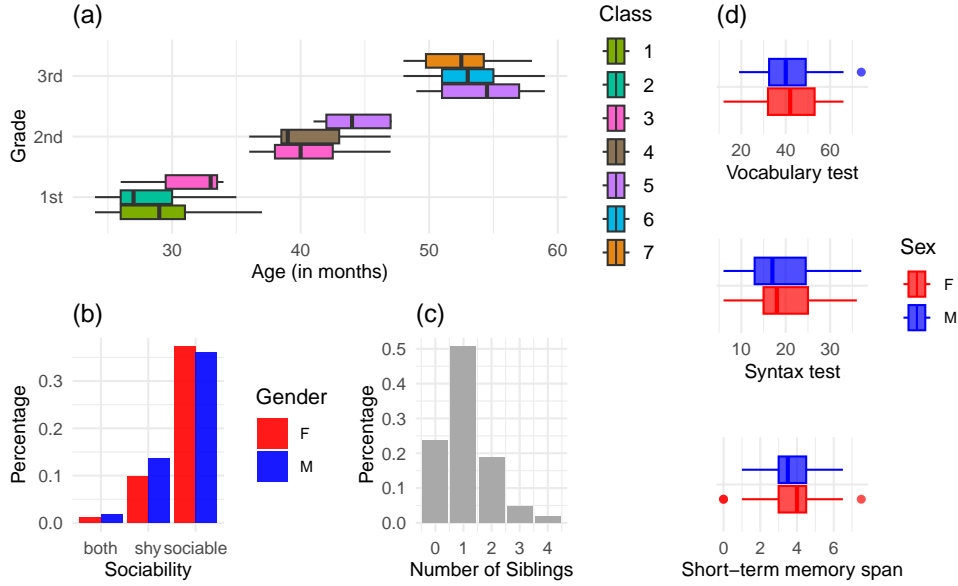


Figure 1: Descriptive statistics of individual socio-demographic and linguistic features. (a) Distributions of childrens' age by their grade and class they were assigned to. Notice the presence of classes with mixed grades. (b) Percentage of answers given by the parents when asked is their child rather sociable or shy, disaggregated by sex. (c) Distribution of number of siblings that the children have. (d) Distributions of aggregated scores in development skill tests disaggregated by sex.

## 2.2 Individual characteristics

The uniqueness of this dataset lies in the metadata that comes attached with the individual nodes. Parents of the children participating 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 environment at home 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 behavioural 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. 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<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> grade), children were divided into 7 classes. Two classes have mixed grades (3 and 5), with one combining 1<sup>st</sup>- and 2<sup>nd</sup>-graders, and the other combining 2<sup>nd</sup>- and 3<sup>rd</sup>-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 Figure 1(c), the number of siblings is another variable that exhibits variation among the group of pupils who participated in the original study, with the 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 3<sup>rd</sup> grade pupils. Since the test items measuring the same

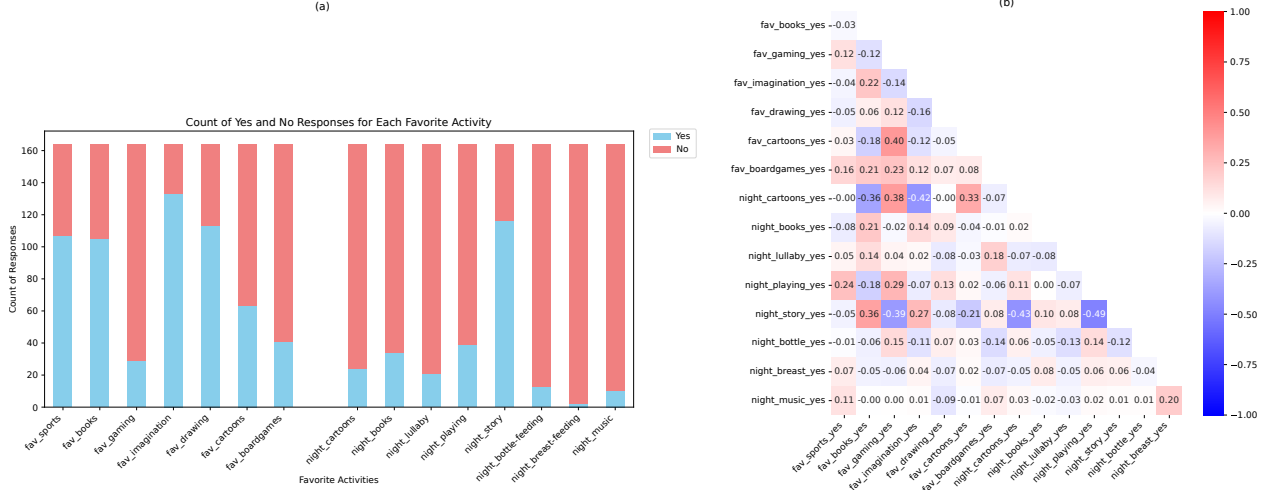


Figure 2: Descriptive analysis of preferred home activities among individuals. (a) Number of positive and negative responses associated to the different out-of-school activities during day (leftmost bars) and before nighttime (rightmost bars). (b) Correlation matrix of favorite activities.

type of skills are highly correlated (not shown), we aggregated them by averaging, for each skill, the sum of distinct scores from the two consecutive years. These two resulting aggregated linguistic scores show a correlation, yet we still observe considerable variation across different score ranges. Conversely, short-term memory span does not correlate with linguistic skills. Probably unsurprisingly, these developmental measures, even though separately designed for each grade, turn out to be correlated with age, but at moderate levels. The boxplots for the three resulting development skill test measures are reported in Figure 1(d), disaggregated by sex.

Parents were also asked to indicate their children’s preferred activities from a set of proposed ones, distinguishing between daytime and nighttime activities. Aggregated counters associated to the given answers are reported in Figure 2(a), while the correlations between them is displayed in Figure 2(b). We first notice that the sample of children is quite heterogeneous, and most of the variables are poorly correlated. Additionally, we also see from (a) that some variables exhibit low variability and may not effectively distinguish social behaviors. Nevertheless, we decided to retain all these features to avoid arbitrary selections, while we did not use dimensionality reduction techniques to preserve their individual interpretation.

### 2.3 From proximity data to networks

Information on spatial proximity and contact duration was collected using autonomous RFID Wireless sensors installed on participants. These devices tracked face-to-face proximity 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 Ref. 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 Iacopini et al. [2023]. 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 among pupils. 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. The individual propensity of a node  $i$  to interact with its peers can be quantified via 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 3 shows pictorial visualization of the resulting networks measured during the out-of-class context, aggregated over time. The positioning of the nodes is the same in both panels, but in Figure 3(a) nodes are colored according to the sex of the child, while in Figure 3(b) nodes are colored according to the class affiliation. We can see from this second case how, even during the out-of-class time, where in principle pupils are free to interact beyond their classmates, the network presents clusters: pupils from the same class preferentially interact 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

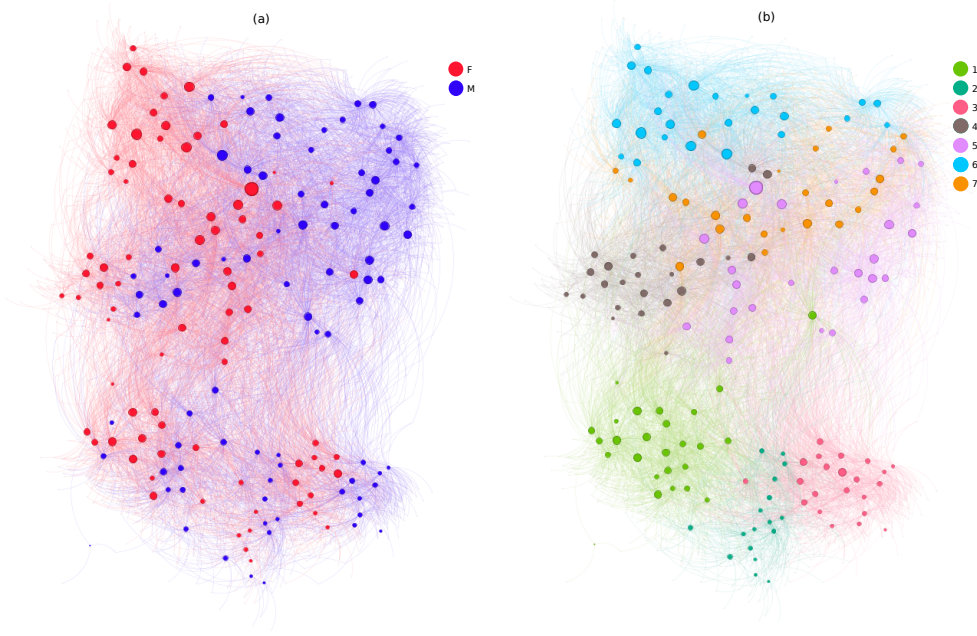


Figure 3: Visualizations of the temporal aggregated network during the out-of-class context. The size of the nodes corresponds to their degree. The layout is induced by the bipartite representation of interactions, where coloured nodes representing pupils are connected through “phantom nodes” representing group interactions (hyperedges). (a) Colors indicate the sex of pupils: blue for male and red for female. (b) Colors indicate class affiliation. The location of nodes remains the same between the two networks. The visualization has been produced using Gephi Bastian et al. [2009].

at the top. As already known Dai [2022], Figure 3(a), 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 qualitatively 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. We will then move to more complex dyadic and group measures that take into account the features of the nodes at different categorical levels.

The first and most fundamental quantity we consider is the total interaction time. Indeed, at the node level, the strength provides a simple and interpretable measure of the individual propensity to interact. Given the specific age group considered in this study, this represents a particularly interesting factor: it can be interpreted, in a dual manner, as an early indicator of both a tendency to interact and, conversely, as a proxy for a propensity to isolate.

#### 3.1 Classification of interaction time using metadata

We split the distribution of interaction time 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)]. We then train a Random Forest classifier [Breiman, 2001] on the metadata prepared as described in section 2.2 to predict the quantile of each pupil. 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. The resulting confusion matrices, obtained by taking the average over 20 independent training realizations, are reported in Figure 4(b) and (e) for the in-class and out-of-class settings, respectively.



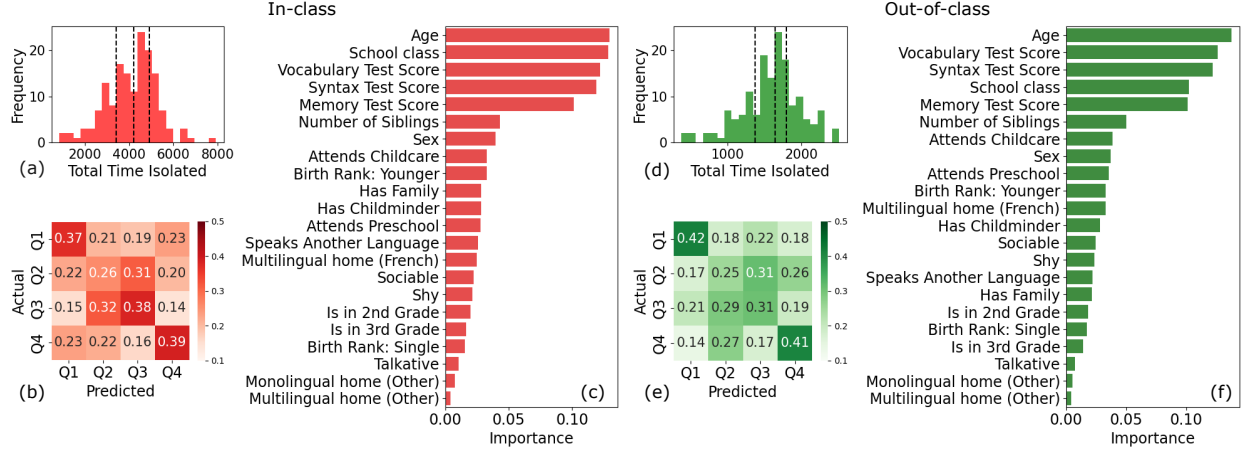


Figure 4: Classifying pupils into classes of social activity based on individual metadata during in-class (a-c) and out-of-class (d-f) time. (a,d) Pupils are divided into four classes of social activity based on individual time spent in social interactions. The 4 quantiles of the distributions are marked by dashed lines. (b, e) Confusion matrices associated to the classification task performed by a random forest model, that is able to distinguish the most and least socially active pupils. The bar plots (c,f) report the relevance the model assigns to each individual feature; the top 5 features are consistent across the two contexts.

Our analysis reveals that even though the metadata do not contain direct information on contacts, it is possible to use them to classify individuals according to their measured level of sociality. The signal, even though globally weak, enables good partial differentiation if one focuses on the two extreme quantiles—which are also the most interesting ones as they depict the most and less active individuals. The signal is stronger in the out-of-class setting [Figure 4(e)], possibly due to the removal of spatial and logistic constraints imposed by the room and the teacher in the classrooms. We identify the most important features for classification, shown in Figure 4(c) and (f). Notice how the top five are consistent in the two settings: age, school-class and language test scores. This preliminary analysis suggests that individual 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

We just saw that age can be a good predictor for duration of contact at the level of the single node. We would like to move a step further by taking into account the coupling at the level of links. Previous works 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 Veenman [1996], Leroy-Audouin and Suchaut [2007].

To begin, we construct a weighted undirected network from all the in-class interactions happening between children of the same class, aggregated over all the 10 weeks of observation. In Figure 5(a) we plot the distribution of duration of in-class interactions against the age difference of the children involved in the considered pair. We see a peak (longer interactions) for pairs of pupils with less than half a year of age difference, but this quickly diminishes beyond that. To access the effect of network structure and/or the distribution of weights in the observed pattern, we use two reshuffling methods. The “hard” reshuffling removes all edges and reassigns them to random pairs of nodes (conserving only the total number of edges). This procedure completely destroys the network structure, and consequently also the duration distribution observed in the data. By contrast, the “soft” reshuffling removes all the weights (without changing the network structure) and reassigns them randomly to the edges. This method does not affect the patterns of interactions, 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 then repeat the same construction and reshuffling methods but for out-of-class interactions. While the distribution resembles that of in-class interactions, the extent to which you interact with people in your network matters significantly more compared to the null model. In Figure 5(b), we see that pupils with less than half a year age difference interact more than expected when compared to both the hard and the soft reshuffling null models. Interactions between children of larger age differences are also much less than expected from the soft reshuffling model.

Having accessed these differences between in- and out-of-class interactions, we now want to look at possible differences between the level of sociality when exposed to different “treatments”. From a social perspective, the most obvious

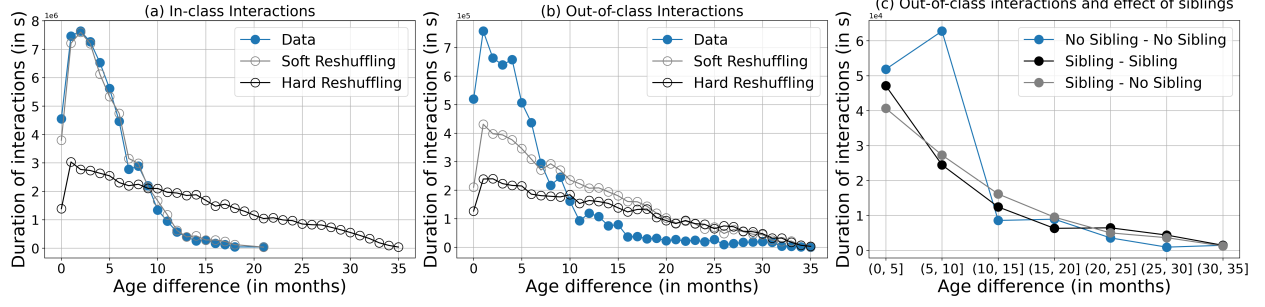


Figure 5: From the weighted interaction network of children aggregated across weeks, the duration of interactions is plotted against age difference of children interacting (a) during class time (b) during free time. The data is compared with 50 realizations each of hard and soft reshuffling models. The 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

change in mixed-grade classrooms is that students of different ages are forced, when in class, to interact with each other. When this restriction is removed, do these children interact differently? We can artificially explore the differences in “removing” this restriction by looking at interactions happening out-of-class but differentiating for children belonging to mixed- against single-grade classes. Notably, we do not observe significant differences in among the two. Even though a more rigorous check should be performed in further studies, this means that the distribution of duration of out-of-class contacts given an age difference, is not different if a pupil is exposed to larger age differences in class or not.

We can look for alternative signals by leveraging information we have about the environment at home. For example, do children with siblings interact outside their age group more than only children? That is, do children who are exposed at home to other children with an age gap of roughly a year or more tend to interact more with varied age groups at school as well? In Figure 5(c), we plot the duration of out-of-class interactions against the age difference of interacting pairs of nodes for three cases: (i) both nodes have no siblings, (ii) both nodes have siblings, (iii) only one node has siblings. While this preliminary investigation seem to indicate that having a sibling does not seem to be associated with more interactions outside your age group, pupils who have no siblings seem to have a different pattern in the way duration depends on age difference: instead of observing a smooth decay with age difference, single children display a more pronounced peak for interactions with children within a year apart from each other, followed by an abrupt drop.

### 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 latter analysis focused on the amount of time spent together by pairs of children. Nevertheless, as for many other systems, childrens’ interactions can involve more than 2 pupils at the same time 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 phenomena Battiston et al. [2021]. Given the individual metadata available in this study, it is then natural to extend the study conducted so far and investigate how individual traits of group members influence the overall time spent together in a given group gathering –beyond the pairs.

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 pair of nodes (pupils) interacting and  $t$  is the time at which the interaction was observed. Given the temporal granularity, it is however possible to naively construct the group interactions by looking at the cliques formed by pairwise interactions occurring at the same timestamp, (following the procedure of Refs. Iacopini et al. [2019], Cencetti et al. [2021]). For example, if at a specific timestamp  $t$  we observe that in the contact data nodes  $i$ ,  $j$  and  $k$  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 can then be considered what has been recently addressed as a higher-order network Battiston et al. [2020], where group interactions among children are represented as hyperlinks, that are links that connect an arbitrary number of nodes. Similarly to the pairwise case, we then assign to



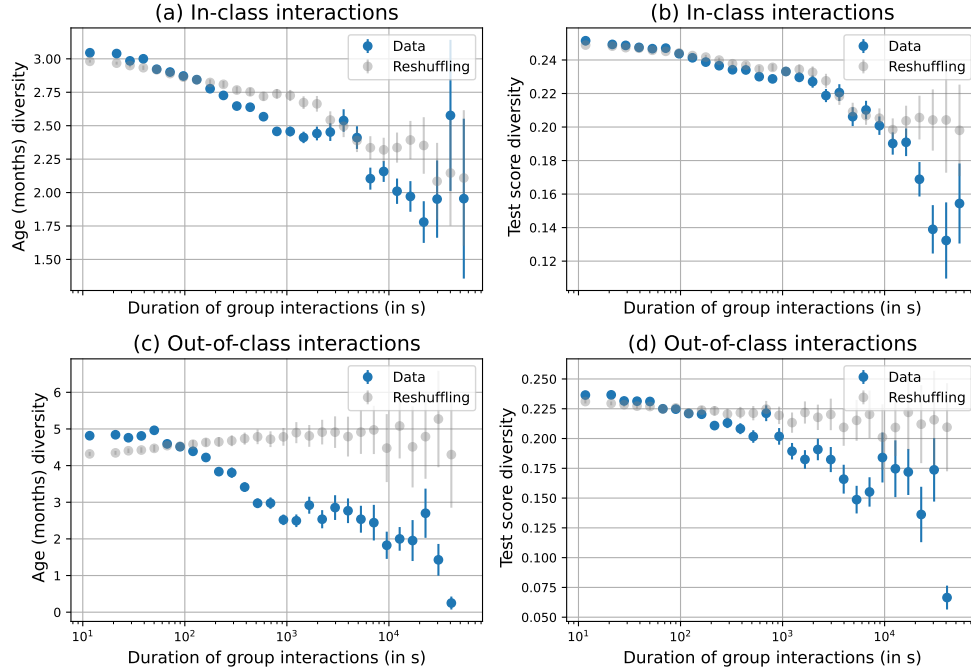


Figure 6: The duration of group interactions observed in pupils interacting during class (a,b) and out of class (c,d) is plotted against group diversity measured from individual characteristics: age (a,c) and language development test scores (b,d). Results obtained from empirical data (blue) are compared to those of randomized null model (gray), where alter group members are taken at random from a different group of the same size over 10 different realizations.

each hyperlink a weight which accounts for the total temporal duration the hyperlink was observed in the contact data. Hyperlinks with null weights (i.e., never observed) were not considered in this analysis.

We integrate individual traits belonging to different group member into a measure of affinity for the whole group. In particular, we considered two individual characteristics that were included in the previous analyses, i.e. age and results in development skill tests. While age is a simple measure given in months, test scores require some pre-processing. We first normalise the test results such that the score obtained by children is always between 0 and 1. Then, for each pupil, we store these results into a 3-dimensional vector, where each entry corresponds to the result obtained in the test on vocabulary, syntax and memory. We can now quantify *group affinity* in terms of how further apart are its members with respect to the considered quantity. A natural way of doing this is to consider individual distances from the barycenter of the group. We thus measure the diversity in age and development skills test scores by calculating the sum of the absolute distances of each node from the barycentric coordinate, and then dividing this sum by the number of group members. As per previous analyses, we split interactions into in-class and out-of-class, to distinguish between the two different social contexts in which pupils interact.

In Figure 6 we plot group diversity as a function of the group duration for the two different individual traits and the two different contexts of interaction considered. We observe that longer-lasting groups tend to be less distant, thus group affinity is higher, both in terms of age and development skill performances. Results obtained from empirical data are compared with randomized reference models, where the distance between a node and its group is computed instead using different nodes taken at random from another group of the same size. In the in-class context [Figure 6(a-b)], we see that the empirical groups that last for long present lower diversity than those from the reshuffled model. This trend is much more evident in the case of age diversity (a) than in the case of development skill tests (b), where appreciable differences only appear for the longer-lasting groups. Signals are stronger in the out-of-class context [Figure 6(c-d)], where we can see clear differences of trends between empirical data and the reshuffled cases. The group diversity after reshuffling remains constant as group duration increases, in striking contrast to the decreasing trend observed in real-world data. We conclude that, whenever pupils are left free to interact (out-of-class), they gather in groups whose duration depends on group affinity in terms of individual age and development skill test performances, independently from their size.

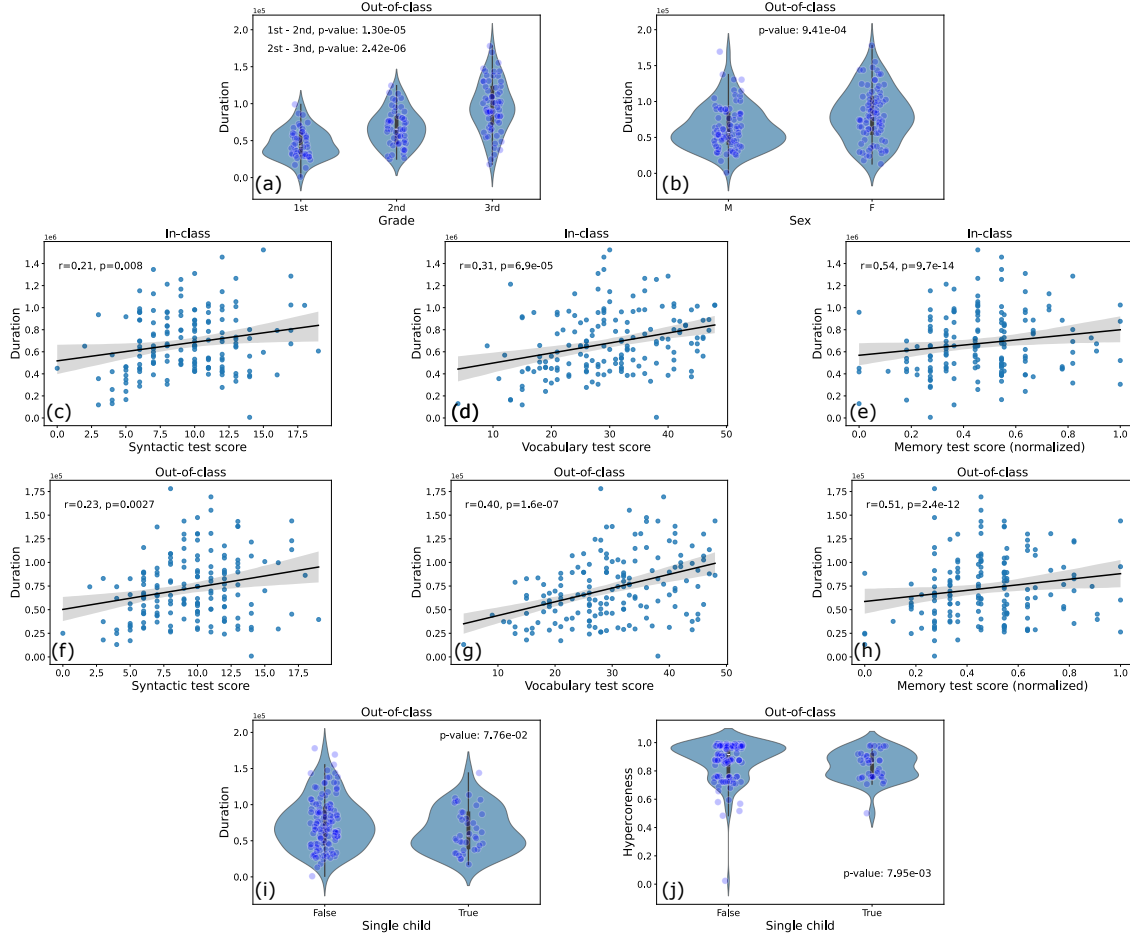


Figure 7: Distribution of contact duration across for children belonging to different grades (a) and sex categories (b). (c-h) Association between duration of contacts and language development test scores in different contexts. Distribution of contact duration (i) and hypercoreness centrality (j) for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

### 3.4 Node attributes and network measures

In this final section, we consider node-level measures of interaction computed for individual pupils and look for differences in their distributions when accounting for different categories of individual traits. In particular, we focus on the same characteristics used in the previous sections: sex, grade, development skill test scores, and number of siblings.

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 Figure 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 Figure 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 [Figure 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 Figure 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 Figure 7(j), where we compare the hypercoreness values of the nodes. Hypercoreness is a recently-developed measure of centrality for higher-order networks Mancastroppa et al. [2023, 2024] that quantifies the extent of nodes' interactions within groups, considering at the same time number of different groups and group sizes. High hypercoreness values correspond to nodes that interact within many large-sized groups that contain high-degree nodes<sup>9</sup>.

## 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 sociality as given by the total time spent interacting/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 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 childrens' interactions across age groups. The analysis indicated no significant difference in interaction patterns based on sibling presence, except for children without siblings who showed longer interactions within their age group. These preliminary results suggest that exposure to mixed-age interactions within the family does not significantly influence the age gap in social interactions at school—apart for a different decreasing trend on how contact duration decreases as the gap increases. Nevertheless, additional comprehensive causal work is necessary to make this inference.

At the group level, we found differences across children with or without siblings in terms of node centrality in the higher-order networks. In fact, even though no significant differences emerge when comparing group durations, children with siblings display higher hypercoreness Mancastroppa et al. [2023], meaning that they engage in more groups of larger sizes. 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].

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 outcomes Suchaut [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.

Another 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 Ref. Veldt et al. [2023], but generalised

<sup>9</sup>More formally, we start by performing a  $(k, m)$ -hypercore decomposition Liu et al. [2020], 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 groups (or hyper-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 groups and all groups of size at least  $m$ . The hypercoreness 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 groups of size  $m$  in the original network.

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 Petri and Barrat [2018], Gallo et al. [2024a], Iacopini et al. [2023], Gallo et al. [2024b]. 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) Iacopini et al. [2023]. 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.

## 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. The higher-order interaction analysis was performed using the XGI library Landry et al. [2023].

## References

- Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47, 2002. doi:<https://doi.org/10.1103/RevModPhys.74.47>.
- Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003. doi:<https://doi.org/10.1137/S003614450342480>.
- 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>.
- 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>.
- Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32–39, 2012.
- Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, pages 128–140, 1741.
- Stanley Wasserman and Katherine Faust. *Social Network Analysis : Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press, 1994. ISBN 0-521-38707-8.
- Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81(2):591, 2009. doi:<https://doi.org/10.1103/RevModPhys.81.591>.
- Andrea Baronchelli. The emergence of consensus: A primer. *R. Soc Open Sci*, 5(2):172189, 2018.
- R. Axelrod and R.M. Axelrod. *The Evolution of Cooperation*. Basic Books. Basic Books, 1984. ISBN 978-0-465-02121-5.
- J Clyde Mitchell. *Networks, norms and institutions*. Mouton., 1973.
- Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007.
- Andrzej Nowak, Jacek Szamrej, and Bibb Latané. From private attitude to public opinion: A dynamic theory of social impact. *Psychol. Rev.*, 97(3):362, 1990.
- Katarzyna Sznajd-Weron and Jozef Sznajd. Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(06):1157–1165, 2000.
- Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What’s in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology*, 271(1):166–180, 2011.
- Alain Barrat, Ciro Cattuto, Alberto Eugenio Tozzi, Philippe Vanhems, and Nicolas Voirin. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clin. Microbiol. Infect.*, 20(1):10–16, 2014. doi:<https://doi.org/10.1111/1469-0691.12472>.

- 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, 2015a.
- Marco Cristani, Giulia Paggetti, Alessandro Vinciarelli, Loris Bazzani, Gloria Menegaz, and Vittorio Murino. Towards computational proxemics: Inferring social relations from interpersonal distances. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 290–297. IEEE, 2011.
- Petter Holme and Jari Saramäki. Temporal networks. *Phys. Rep.*, 519(3):97–125, 2012. doi:<https://doi.org/10.1016/j.physrep.2012.03.001>.
- 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.
- 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.
- 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.
- 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>.
- Juliette Stehlé, Alain Barrat, and Ginestra Bianconi. Dynamical and bursty interactions in social networks. *Phys. Rev. E*, 81(3):035101, 2010. doi:10.1103/PhysRevE.81.035101.
- Kun Zhao, Juliette Stehlé, Ginestra Bianconi, and Alain Barrat. Social network dynamics of face-to-face interactions. *Phys. Rev. E*, 83(5):056109, 2011. doi:10.1103/PhysRevE.83.056109.
- Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Scientific reports*, 2(1):469, 2012.
- Michele Starnini, Andrea Baronchelli, and Romualdo Pastor-Satorras. Modeling human dynamics of face-to-face interaction networks. *Phys. Rev. Lett.*, 110(16):168701, 2013. doi:10.1103/PhysRevLett.110.168701.
- Christian L Vestergaard, Mathieu Génois, and Alain Barrat. How memory generates heterogeneous dynamics in temporal networks. *Phys. Rev. E*, 90(4):042805, 2014. doi:<https://doi.org/10.1103/PhysRevE.90.042805>.
- Márton Karsai, Nicola Perra, and Alessandro Vespignani. Time varying networks and the weakness of strong ties. *Sci. Rep.*, 4(1):1–7, 2014. doi:<https://doi.org/10.1038/srep04001>.
- Matthieu Nadin, Kaiyuan Sun, Enrico Ubaldi, Michele Starnini, Alessandro Rizzo, and Nicola Perra. Epidemic spreading in modular time-varying networks. *Sci. Rep.*, 8(1):1–11, 2018. doi:<https://doi.org/10.1038/s41598-018-20908-x>.
- Didier Le Bail, Mathieu Génois, and Alain Barrat. Modeling framework unifying contact and social networks. *Phys. Rev. E*, 107:024301, Feb 2023. doi:10.1103/PhysRevE.107.024301. URL <https://link.aps.org/doi/10.1103/PhysRevE.107.024301>.
- Giovanni Petri and Alain Barrat. Simplicial activity driven model. *Phys. Rev. Lett.*, 121(22):228301, 2018. doi:10.1103/PhysRevLett.121.228301.
- Luca Gallo, Lucas Lacasa, Vito Latora, and Federico Battiston. Higher-order correlations reveal complex memory in temporal hypergraphs. *Nature Communications*, 15(1):4754, 2024a.
- Iacopo Iacopini, Márton Karsai, and Alain Barrat. The temporal dynamics of group interactions in higher-order social networks. *arXiv preprint arXiv:2306.09967*, 2023.
- Inkeri Kontro and Mathieu Génois. Combining surveys and sensors to explore student behaviour. *Education Sciences*, 10(3):68, 2020.
- 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.
- Mathieu Génois, Maria Zens, Marcos Oliveira, Clemens M Lechner, Johann Schaible, and Markus Strohmaier. Combining sensors and surveys to study social interactions: A case of four science conferences. *Pers. Sci.*, 4:1–24, 2023. doi:<https://doi.org/10.5964/ps.9957>.

- 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 2015b. ISSN 1932-6203. doi:10.1371/journal.pone.0136497. URL <https://dx.plos.org/10.1371/journal.pone.0136497>.
- Megan M McClelland, Claire E Cameron, Carol McDonald Connor, Carrie L Farris, Abigail M Jewkes, and Frederick J Morrison. Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental psychology*, 43(4):947, 2007.
- Susanne 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.
- Sondra H Birch and Gary W Ladd. The teacher-child relationship and children's early school adjustment. *Journal of school psychology*, 35(1):61–79, 1997.
- 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.
- Bruce H Mayhew, J Miller McPherson, Thomas Rotolo, and Lynn Smith-Lovin. Sex and race homogeneity in naturally occurring groups. *Social Forces*, 74(1):15–52, 1995.
- 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, 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>.
- Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. In *Proceedings of the international AAAI conference on web and social media*, volume 3, pages 361–362, 2009.
- 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 française de pédagogie*, 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.
- Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. Simplicial models of social contagion. *Nat. Commun.*, 10:2485, 2019. doi:https://doi.org/10.1038/s41467-019-10431-6.
- 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.
- 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.
- Boge Liu, Long Yuan, Xuemin Lin, Lu Qin, Wenjie Zhang, and Jingren Zhou. Efficient  $(\alpha, \beta)$ -core computation in bipartite graphs. *The VLDB Journal*, 29(5):1075–1099, 2020.
- Quintino Francesco Lotito, Federico Musciotto, Alberto Montresor, and Federico Battiston. Higher-order motif analysis in hypergraphs. *Communications Physics*, 5(1):79, 2022.
- 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.
- Nate Veldt, Austin R Benson, and Jon Kleinberg. Combinatorial characterizations and impossibilities for higher-order homophily. *Science Advances*, 9(1):eabq3200, 2023.
- Kristen M Altenburger and Johan Ugander. Monophily in social networks introduces similarity among friends-of-friends. *Nature human behaviour*, 2(4):284–290, 2018.
- Luca Gallo, Chiara Zappalà, Fariba Karimi, and Federico Battiston. Higher-order modeling of face-to-face interactions. *arXiv preprint arXiv:2406.05026*, 2024b.
- Nicholas W Landry, Maxime Lucas, Iacopo Iacopini, Giovanni Petri, Alice Schwarze, Alice Patania, and Leo Torres. Xgi: A python package for higher-order interaction networks. *Journal of Open Source Software*, 8(85):5162, 2023.