

Social Fluidity in Children’s Face-to-Face Interaction Networks

Daniela Gawehns^[0000–0002–9678–9012] and
Matthijs van Leeuwen^[0000–0002–0510–3549]

LIACS, Leiden University, The Netherlands
{gawehnsd, m.van.leeuwen}@liacs.leidenuniv.nl

Abstract. During free play, children decide with which peers they want to spend their time. The distribution of attention and social effort amongst their peers can be indicative for the inclusivity of the group. In this appraisal paper, we explore a recently introduced measure to estimate social mixing in groups, called social fluidity. We estimate social fluidity for 27 data sets of children’s face-to-face interaction networks collected during playtime breaks and discuss how this measure can be used by field experts to characterize group behavior as well as individual behavior. We therefore draft a step-by-step plan to analyze proximity data based on the behavioral model and the social fluidity measure introduced by Colman et al. [2].

Keywords: Face-to-face Interaction Networks · Temporal Networks · Social Fluidity.

1 Characterizing children’s interaction patterns

School is a place to learn. Reading, writing, and other skills are taught during structured classroom activities. Besides these academic skills, school is also a place to acquire social and emotional skills, which children need to learn and practice by interacting with others. Insights in pupils’ social interactions during unstructured playtime inform developmental researchers about the dynamics in peer networks. This information is important for the development of interventions that aim at more inclusive school environments for pupils with less access to the social world (e.g., hard-of-hearing or autistic children). In turn, an inclusive environment benefits all children’s learning of social, emotional as well as academic skills, and fosters an inclusive society [3].

Studying peer interactions with observational techniques is challenging as children might engage in different types of play and group forming when adults are present to observe them [4]. To study children’s behavior with least interference possible, proximity sensors are used to record interactions between pairs of children. In this paper, we analyze data of face-to-face interactions collected during 27 playtime breaks in pre-primary and primary schools in Portugal [4] [5] and the Netherlands. We are curious to find out how the classrooms’ behavior differ from each other and if we can compare them in a way that is interpretable

by field specialists, robust to choices during preprocessing, and in agreement with qualitative observations of play behavior.

A candidate measure to compare group behaviors is social fluidity, a measure that was recently introduced by Colman et al. (2020) [2]. The authors state that social fluidity can be used to compare groups of human and non-human animals (also called systems) on how individuals distribute their attention amongst other individuals. They characterize these systems as either "allegiant" or "gregarious". Allegiant systems are systems of low social fluidity, characterized by close relationships with few other individuals, as seen in socially grooming animals. On a playground, we would see children spending most time with few of their closest friends and interact often with them. Gregarious systems, on the other hand, are those of high social mixing, where individuals make contact with many partners and distribute their effort more evenly. On a playground, one would observe a group of children spreading their attention equally amongst their classmates, with little clique forming.

Field scientists' observe differences between groups of children in how they spread their attention amongst their peers. The suggested measure fits these observations and is easily translated into the field researchers' vocabulary, which makes it a good candidate for comparing groups of children. We expect most groups to be somewhere in between purely gregarious (i.e., even spread of attention) or purely allegiant (i.e., only cliques) systems, with individual children showing interaction patterns that are different from the rest of the group.

From a network science perspective, we are faced with a temporal network analysis problem. While the suggested measure explicitly models neither the duration of interactions or nor the order of interactions, we hope to infer more about the behavioral patterns of the classrooms than a count of interactions, unique interactions partners, or triangles of interactions would give us.

In this brief appraisal, we describe how we use the measure on our own data and compare our results to what Colman et al. found. First, we introduce social fluidity and reason why it is a promising approach for the data at hand. Then we explore the robustness of the measure, present initial validation results and compare it with an alternative density measure. In the final section, we discuss under which circumstances social fluidity can be used to characterize the data at hand and draft a standard analysis plan, which allows for network and node level analyses.

2 Social Fluidity in temporal networks

Colman et al. [2] define social fluidity as "variability in the amount of social effort the individual invests in each member of their social group" and estimate it for more than 50 different animal systems, including several human face-to-face interaction networks. Face-to-face interaction data can be represented as a temporal network, with people or animals as nodes and their interactions as edges. We first introduce the behavioral model underlying the social fluidity measure and its assumptions, and then describe the algorithmic solution to

estimate the value of the social fluidity measure as suggested in the original paper.

2.1 Behavioral Model

We have a temporal network, with N individuals i as nodes and edges between nodes whenever there is an interaction between individuals i and j . There is one network per timepoint t . The temporal information is reduced to the total number of observed interactions $s = \sum_t s_t$ as the strength of a node and the total number of unique interaction partners k as the degree of a node.

The core of the model is the marginal probability $\Psi(s)$ of finding an edge between a node i with s interactions and any other node j in the network:

$$\Psi(s) = 1 - \int \rho(x)(1-x)^s dx$$

Where $\rho(x)$ is the probability distribution of the probability x of a node i to interact with a node j . This distribution is assumed to follow a power law distribution, as observed empirically in a large number of animal systems (see also annex of [2]).

$$\rho(x) = \frac{\phi \epsilon^\phi}{1 - \epsilon^\phi} x^{-(1+\phi)} \text{ for } \epsilon < x < 1$$

The first term is a normalization factor, ϵ is introduced to avoid divergence. We are interested in estimating ϕ , which is part of the scaling factor and the main driver of the model.

Using the assumption of a power law distribution within the description of the marginal probability results in:

$$\Psi(s, \phi, \epsilon) = 1 - \frac{\phi \epsilon^\phi (1 - \epsilon)^{s+1}}{(1 - \epsilon^\phi)(s + 1)} {}_2F_1(s + 1, 1 + \phi, s + 2, 1 - \epsilon)$$

where ${}_2F_1$ stands for the Gauss hypergeometric function. To find ϵ , we can use that the individual probabilities of j interacting with any other node i in the network sum to one: $\sum_j (x_{j|i}) = 1$ and get:

$$N = 1 + \frac{(1 - \phi)(1 - \epsilon^\phi)}{\phi \epsilon^\phi (1 - \epsilon^{1-\phi})}$$

Which in turn can be solved for ϵ for given ϕ and N . The expected degree \bar{k}_i is

$$\bar{k}_i(s, \phi, N) = (N - 1)\Psi(s, \phi, \epsilon)$$

We now have the tools to estimate ϕ and ϵ , and to fit the model. The model's main driver is ϕ , characterizing the overall behavior of a group. As a marginal probability of finding an edge between a node with strength s and any other node in the network, it allows us to find an expected degree per observed strength in a network.

2.2 Algorithm

To estimate ϕ , we need to fit to model. This is done by taking the difference between observed and expected degree per node and minimizing the mean squared error of the entire group: $\sum_i (k_i - \bar{k}_i(s, \phi, N))^2$.

The expected degree per node is \bar{k}_i is $(N-1)\Psi(s, \phi, \epsilon)$, with $\Psi(s, \phi, \epsilon)$ describing the marginal probability of a node with strength s of interacting with any other node in the network.

Algorithmically, this is solved with Newtons' method, starting at a default value of ϕ_{zero} of 0.3 and running until a difference of less than 0.01 between observed and expected degree is achieved.

3 Data and Methods

Data was collected between 2013 and 2020 in schools in Portugal [4] and the Netherlands. Children were given light weight proximity sensors that were attached to their clothes and recorded a signal whenever two sensors were less than 1.5 meters away from each other, at a rate of 1 Hz. This translates into one network per second. The recordings are between 9 min and 30 min long and the groups consist of 10 to 41 children.

3.1 Methods

We explore two characteristics of the social fluidity measure: 1) Robustness to choices during data preprocessing, and 2) external and internal validity (i.e., does it measure what it is supposed to measure?)

Regarding robustness, we are curious to find out what the effect of preprocessing steps are on the value of ϕ . Sensor data needs preprocessing to remove sensor noise. Temporal smoothing and setting thresholds for a minimum length of interaction events are the most common preprocessing steps. Both reduce sensor noise and therewith the number of observed interactions and observed interaction partners, as the shortest interactions are removed. In consequence, this influences the estimated ϕ value. We focus on thresholds for minimum length of interactions and calculate ϕ for different thresholds (1, 10, 20, and 40 seconds).

Regarding validity, we compare our results with observational data gathered during data collection and replicate the internal validity assessment that Colman et al. performed by correlating ϕ with several other network characteristics. Furthermore, we correlated the count of triangles in the data as a measure for network density with the calculated ϕ values.

We ran a combination of R and Python scripts to analyze the data. The codebase of the Colman et al. paper is available on Github [1] and is used to estimate ϕ . The data preprocessing and visualizations are done in R. We estimated ϕ with the default starting value ϕ_{zero} of 0.3 on 27 data sets and with four different thresholds for minimum interaction lengths. We also recorded the estimated degree per node in all data sets for further exploration.

4 Degrees and Strengths as a basis

We start our exploration with a similar graph as in the original work, where the number of observed interactions and the number of interaction partners per individual are plotted against each other. Fig.1 shows how degree k and strength s are related to each other for every node in the data, filtered with a minimum interaction length threshold of 10 seconds. We can see that our data resembles somewhat what we expect under the assumptions of the behavioral model, which is an arc from the lower left to the upper right. When removing the four children with more than 200 interactions (Fig.2), the plot looks more spread out, with a pronounced cluster of children who have few interaction partners and few interactions. Overall, we can state that there are large differences in how children spread their attention amongst their peers. This makes a measure based on the relation between these two counts a good candidate to characterize attention spreading in groups.

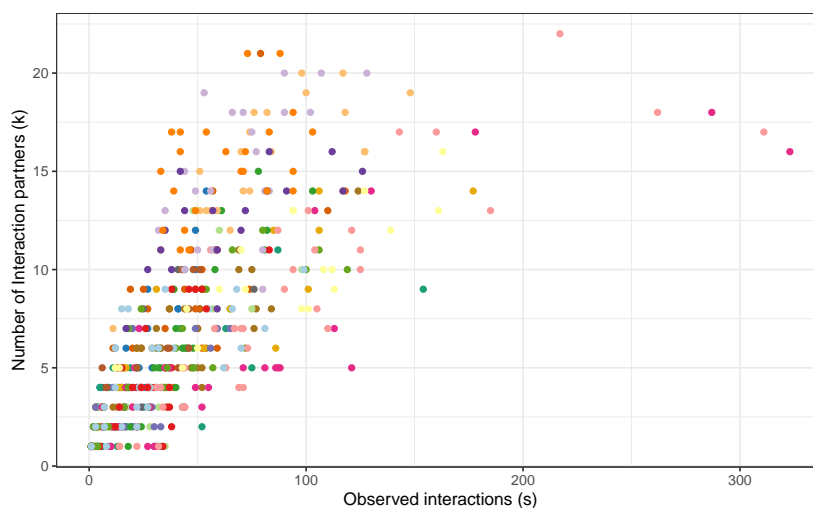


Fig. 1. K - S plot of all children in all 27 data sets. Only interactions that lasted at least 10 seconds are included.

5 Model Assumptions

We estimated ϕ for all 27 data sets at four different thresholds of minimum interaction length (i.e., 1, 10, 20, and 40 seconds). Out of 108 combinations, 28 times the script to estimate ϕ did not converge. Out of those 28 runs, three times there were no interactions of the minimum length in the data, i.e., all data were filtered out. For the remaining 25 runs, the algorithm to estimate ϕ did

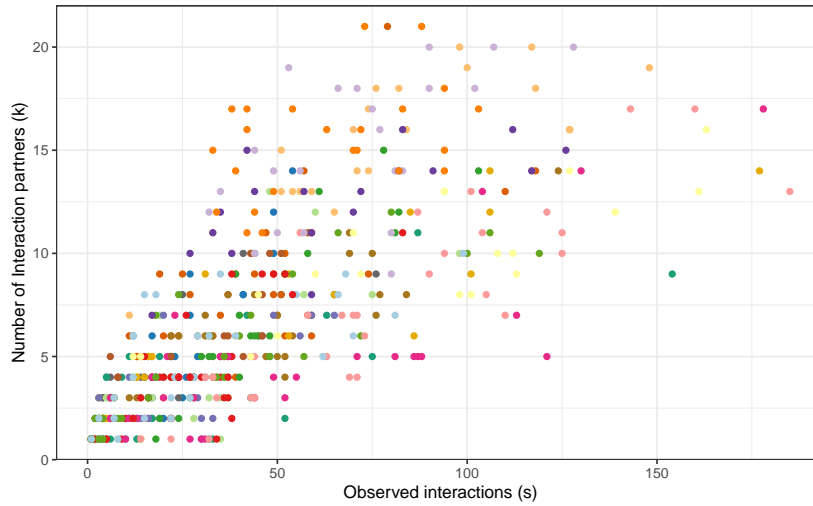


Fig. 2. K - S plot of children with less than 200 observed interactions. Only interactions that lasted at least 10 seconds are included.

not converge when using the default starting value of 0.3 for ϕ . The algorithm did, however, converge when setting the starting value to a higher value, which resulted in ϕ values that were much higher than the other ones. Three ϕ values were also at a starting value of 0.3 much higher than all others ($\phi = 166/33/45$). The presented results exclude those data. In the data that were not successfully processed to extract ϕ , we looked more closely at five data sets after preprocessing with a minimum length threshold of 10 seconds. We found that in those data, between 18 % and 30 % of the individuals were in contact with only one other individual. The network sizes were not outside the range of the other, successfully converging data sets. When plotting k against s in Fig.3, we can see that most nodes in the non-converging data sets are in the lower left corner with many children having very few contact partners and contacts overall. Compared with the converging data (the grey dots), there seems to be an imbalance towards those individuals. These data do not follow the assumed power law distribution and forcing the data into the assumed distribution results in ϕ values that are not comparable to other networks. From a data exploration perspective, these imbalanced data sets will be interesting to look further into to find out if and how the groups and circumstances of data collection differed structurally from the balanced data sets.

6 Robustness

Fig.4 presents the spread of the ϕ values according to different values for minimum length. At minimum length 1, no interactions are filtered out of the data. The ϕ values decrease as the cut-off value for a minimum length of interaction increases.

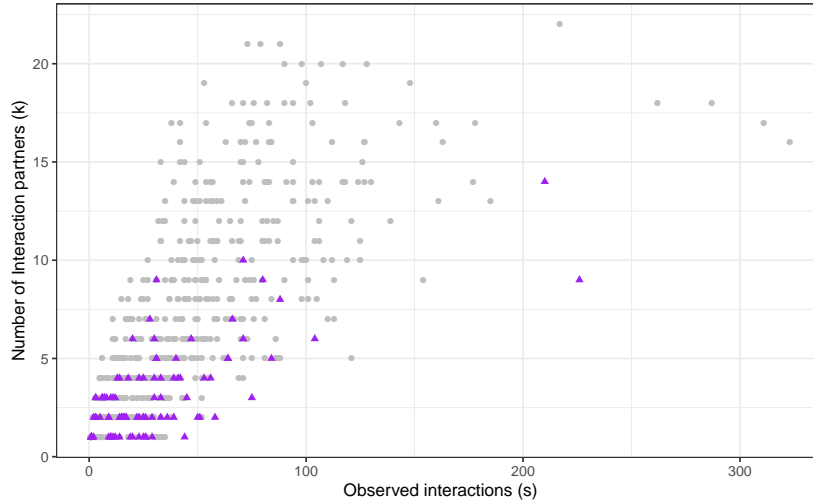


Fig. 3. K - S plot of all children in all data sets. The light grey dots are the children of the converging data sets, the purple triangles are the children in the non-converging data sets.

This is not surprising as ϕ is strongly correlated with mean degrees (Pearson Correlation is 0.92) and moderately correlated with the mean strengths (Pearson Correlation is 0.49). With a higher cut-off value, the value of the mean degree and mean strength decreases, as more interactions are filtered out of the data. The spread is highest around cut-off values 1 and 10 (SD of 0.22 at 1, SD of 0.27 at 10). We note, however, that as the cut-off value increases, also the number of successful converging runs decreases, i.e., comparing the spread across cut-off values has only limited use.

7 Validity

For a first exploration of the validity of the social fluidity measure on network level, we settled for a minimum length of 10 seconds as it resulted in most successfully converging runs while keeping out a sensible amount of sensor noise. This gave us $n=21$ data points to calculate the same correlations as in the original work. Colman et al. found a significant correlation neither between mean strength and social fluidity, nor between network size and social fluidity. ϕ did however correlate with mean degrees and the dispersion index (a measure for the distribution of edge weights). As dispersion index and mean degree did correlate with each other, they conclude that ϕ summarizes these two network features.

We find a high correlation with Mean Degree ($r=0.92$), a less pronounced correlation with Mean Strength (0.49), and negative correlations with Network Size ($r=-0.45$) and Dispersion Index (-0.37). While the first two are in accordance with the original findings, the second two are not.

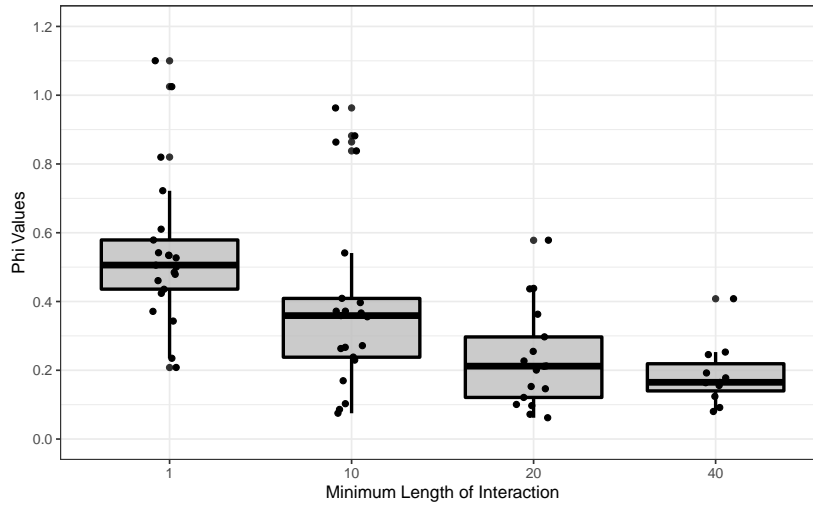


Fig. 4. Spread of ϕ values depending on minimum length of interaction

The difference in results is possibly caused by the smaller number of networks in our sample (21 as compared to 57) and a smaller range of ϕ values. In the original exploration these ranged from 0.2 to 1.9, while the spread in our data set ranges between 0.1 and 1.1.

We also counted triangles in the networks, to see if a more simple method to estimate connectedness would yield similar results. Ignoring the temporal aspect and counting all interactions made as edges, we found no correlation between ϕ and the number of triangles ($r = -0.07$). Even though ϕ is correlated with the average number of connections per node (i.e., the mean degrees), it isn't correlated with the number of triangles made by those connections.

We have limited qualitative observations and could not make out a relationship with observed play behaviors, for example if football was played or not. In one group, the ϕ value changed from .397 at the first assessment day to 0.075 and 0.086 at the second assessment day, when two break times were recorded with the sensors. For these three assessments, we know that children played in groups and on a small playground; there is no extraordinary observation that could explain the rather high value at the first assessment. Unfortunately, the qualitative observations are not structured enough to relate the kind of games played, group forming, or the availability of playground toys to the estimated ϕ value. We are in discussions with the field specialists to see if using a semi-structured observational protocol would be feasible during data collection. Such additional information could serve as external validation.

8 Characterizing children's interaction with Social Fluidity - a suggested analysis pipeline

As a basis for future work and discussion, we suggest the following pipeline to use the concept of social fluidity to analyze and explore human face-to-face interaction data in unstructured and time limited settings, such as playtime breaks.

- Step 1: choose an appropriate cut-off score for the necessary preprocessing. This will depend on the type and length of interaction expected to happen during the assessment.
- Step 2: check the assumption of a power law distribution of contact partners given the total number of contacts per node. This can be done by plotting k and s against each other.
- Step 3 : Estimate ϕ and interpret the network level behavior
- Out of curiosity, we add: Step 4: Look at the difference between expected and observed degree for a node level analysis

We walk through these analysis steps for three classrooms. The remaining classrooms will be assessed in collaboration with the field experts.

We chose 10 seconds as a minimum length cut-off score, keeping the balancing act between filtering and information loss in mind, and being fully aware that this is an arbitrary choice.

To check the assumption of a power law distribution, we plot all three groups in Fig.5 and add the calculated expected degrees as a visual aid. We can see that the nodes are spread across a diagonal, and not so much like the expected arch. However, we do not see the large clustering in the lower left, as in the groups where the search did not converge, nor do we see groups where the majority of children have only one or two interaction partners. We therefore continue our analysis and estimate ϕ for the three groups. Out of curiosity, per child, the difference between observed and expected degree is calculated.

The estimated ϕ values are 0.37, 0.4, and 0.37 for classrooms 1, 2, and 3 respectively. Unfortunately, qualitative information on the network level is missing for one group and is very limited for the other two ("big playground" and "children play different things (no football)"). Class 1 and class 3 are the same children, measured during their morning break at two different days. On yet another day, we estimate a ϕ value of 0.23 for a lunch break with the same group of children. For the exploration of the remaining classrooms, we are curious to find out if a lower ϕ value for the lunch breaks as compared to morning breaks can be found structurally also for other classes. This would imply an effect of the time of the day on the children's mixing behaviors.

Out of curiosity, we explored if a node-level analysis was possible and useful, and calculated the difference between expected and observed degree per node. We have observational notes for children that are hard of hearing (hoh), and wanted to see if these children's interaction behavior could be characterized by the difference score. Fig.6 is a summary of the difference scores in the three

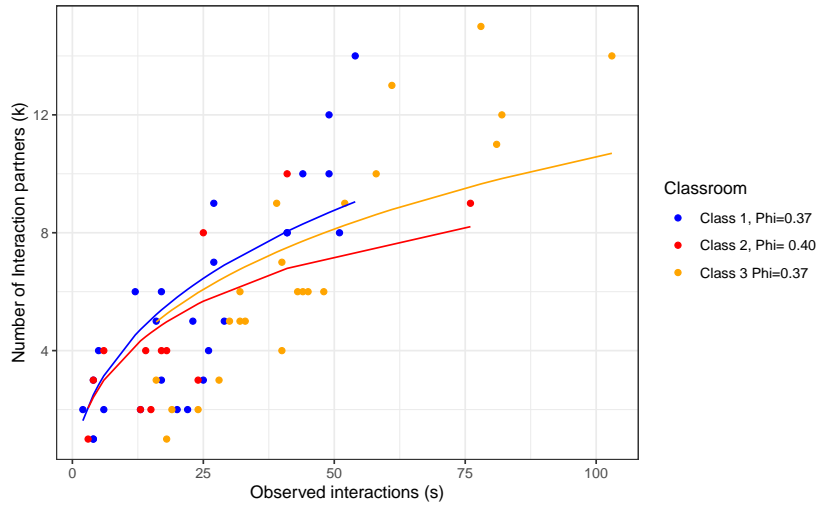


Fig. 5. K - S plot for a convenience sample of three classrooms

classrooms. The hoh-child is represented with a blue triangle, while the other children are marked with red dots. We can see that in classrooms 1 and 2, the hoh child has a negative difference score and is in the first quartile of the data. In class 3, the hoh child compares to the other children as average and has a positive difference score.

A negative difference score is possibly associated with "butterfly" behavior: a child has more contact partners than expected based on the overall number of contacts, which manifests in a child being in close proximity with many different peers, but not engaging in repeated interactions with them. The field researchers did observe such behavior for the hoh child in Class 2, but not for the hoh Child in Class 1.

A positive difference score implies less contact partners than expected based on the number of interactions made. This would manifest in many reoccurring interactions with the same peers. In Class 3, the hoh child played with a fixed group of friends and spent some time alone, which might explain the positive difference score.

We are aware that these observations have only anecdotal value. Further analysis of all classrooms and possibly extending the exploration to other node attributes, such as age, gender or scores on standardized tests is needed to draw a more substantial conclusion.

9 Conclusion and Work in progress

In this brief appraisal paper, we have investigated how the social fluidity measure and the underlying behavioral model can be used to characterize group forming

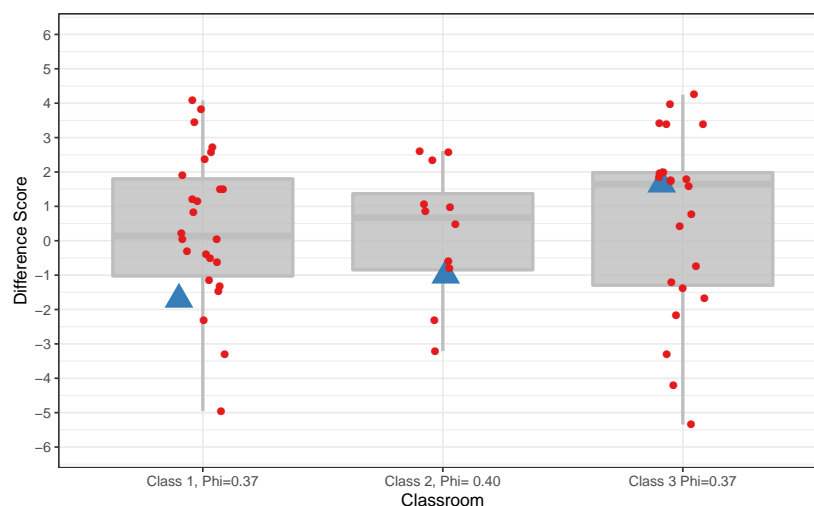


Fig. 6. Overview of the Difference Score between Expected and Observed Degree for three classrooms. Blue triangles represent the hard-of-hearing children, the red dots their classmates. Class 1 and Class 3 are the same children, observed at two different break times.

and social mixing of children during break time. We have shown that decisions at the pre-processing level impact social fluidity and that external validity can only be assessed with structured qualitative data.

As a limitation, we note that two important pieces of information are not directly taken into account in the suggested pipeline: the group size of the interactions, and the length of those interactions. Both indicators could help gain qualitative insight as certain types of play are associated with smaller or larger groups, and shorter or longer duration of interactions. We hope to include them in another way.

Another limitation is the limited robustness to changes of the thresholding parameter when preprocessing the data. Without prior knowledge about a minimum length of interaction in the study population, it remains hard to justify a threshold. A remedy would be to run several analyses in parallel and explore by how much the parameter settings will influence the high-level interpretation of the results. With such a multiverse analysis, the effect of the thresholding will be made transparent to the reader of the research report.

Work is in progress to refine the data analysis pipeline at both, the network- and node-level. Regarding the network-level analysis, we are curious to see if the characteristics of "gregarious" and "allegiant" can be translated into a notion of how inclusive a system is. Regarding the node-level analysis, we made a first step with analyzing the difference scores to characterize individuals by how much they under- or over-invest in other individuals. This approach certainly needs more exploration and research, but seems to be a promising way to

describe the dynamics of an individual child's interactions. In cooperation with the field specialists, we will have a closer look at other node attributes to find possible associations with the difference score. This cooperation will also show how intuitive and interpretable the suggested measure is for field experts.

Finally, in order to assess social fluidity as a measure to describe children's face-to-face interaction networks, we would like to suggest to preregister new assessments, including a detailed observational protocol before new validation data is collected.

Acknowledgements. Dr. Adva Eichengruen (Leiden University), Lisa-Maria van Klaveren (Leiden University) and Dr. Guida Veiga (University of Evora) collected the data and helped link these initial results to their research questions. Dr. Ewan Colman (University of Edinburgh) made code available and discussed its implementation.

References

1. Colman, E.: EwanColman/social-fluidity, <https://github.com/EwanColman/Social-Fluidity>, original-date: 2017-11-23T01:14:16Z
2. Colman, E., Colizza, V., Hanks, E.M., Modlmeier, A.P., Hughes, D.P., Bansal, S.: Social fluidity mobilizes contagion in human and animal populations . <https://doi.org/10.1101/170266>, <http://biorxiv.org/lookup/doi/10.1101/170266>
3. Qvortrup, A., Qvortrup, L.: Inclusion: Dimensions of inclusion in education **22**(7), 803–817. <https://doi.org/10.1080/13603116.2017.1412506>, <https://www.tandfonline.com/doi/full/10.1080/13603116.2017.1412506>
4. Veiga, G., Ketelaar, L., De Leng, W., Cachucho, R., Kok, J.N., Knobbe, A., Neto, C., Rieffe, C.: Alone at the playground **14**(1), 44–61. <https://doi.org/10.1080/17405629.2016.1145111>, <https://www.tandfonline.com/doi/full/10.1080/17405629.2016.1145111>
5. Veiga, G., de Leng, W., Cachucho, R., Ketelaar, L., Kok, J.N., Knobbe, A., Neto, C., Rieffe, C.: Social competence at the playground: Preschoolers during recess: Social competence at the playground **26**(1), e1957. <https://doi.org/10.1002/icd.1957>, <http://doi.wiley.com/10.1002/icd.1957>