Focus on dynamics: a proof of principle in exploratory data mining of face-to-face interactions

Daniela Gawehns[1], Guida Veiga[2], and Matthijs van Leeuwen[1]

[1]Leiden Institute of Advanced Computer Science, Leiden University
[2]Comprehensive Health Research Center, University of Évora

Keywords: Dynamic Networks, Information Theoretic Data Mining, Exploratory Data Analysis, Spatial Proximity, Minimum Description Length

Extended Abstract

In the context of behavioral sciences, face-to-face interaction data is often analyzed as aggregated information of stacked static analyses (e.g., [6] [5] [2]). We suggest a method to explore data collected with RFID sensors that exploits the temporal information of that data.

Introduction: local structures and global summaries

Face-to-face interactions can be represented as a changing network over time (i.e., a temporal network), where individuals are represented by nodes and contact moments by edges. A node's behavior is easily summarized by aggregates such as averages or counts. While such aggregates over time or across nodes are easily calculated statistics, we will loose information on the local level to afford a summary on a global level. Pattern mining methods, such as the one that we propose, can extract relevant local structures from data without losing the granularity of the given dataset. Furthermore, these local structures can be used to give the user an easily interpretable summary of the data.

Usecase: Proximity data of pre-schoolers

Our usecase is based on a set of proximity data collected by developmental researchers [6]. The data was gathered in groups of ~25 pre-school children during their recess breaks. The domain experts knew that there were children in the groups that spent most of the recess time alone. We were instructed to take this information into account. Furthermore, we had to deal with a small group size compared to a high temporal resolution (i.e., per-second proximity information).

Methods: Pattern mining in dynamic networks

The Minimum Description Length (MDL) principle is leveraged in a model selection method that can be described as "selection by compression". The description length (i.e., the size of the compressed data) of different models is compared and the model resulting in the best compression is selected. Inspired by an existing method for MDL-based pattern mining in dynamic networks [4], we developed our own encoding scheme to match the peculiarities of the data (i.e., lonely children, small group size, high resolution) and to make better use of the temporal information in the data.

Preliminary results: exploring subgroups

Figure 1 visualizes the first 20 subgroups (i.e., those that covered at least seven time steps) that we extracted from the data. On the y-axis, the group ID is given, with a low ID number indicating that the subgroup was introduced early-on in the search process to the model. Time is given in time steps of 20 second intervals on the x-axis. The subgroups are categorized into "loners", "groups of two", and "larger groups". We can now link the children's scores on several questionnaires, such as the Empathy Questionnaire (EmQue) [3] and the Strengths and Difficulties Questionnaire [1], to the subgroup categories. For example, we can say that the two children that keep mostly to themselves (subgroups one and four) both have an empathy score of 2.7 and a social competence score of 2.3 and 2.4 respectively. This enriched structural information will be the basis for further exploration of the data.

Future work: interactive visualization

On-going work includes the validation of the method on additional datasets. We are also working on an interactive visualization to include background variables (e.g., gender and age), as well as scores on social and emotional competences when exploring the resulting groups.

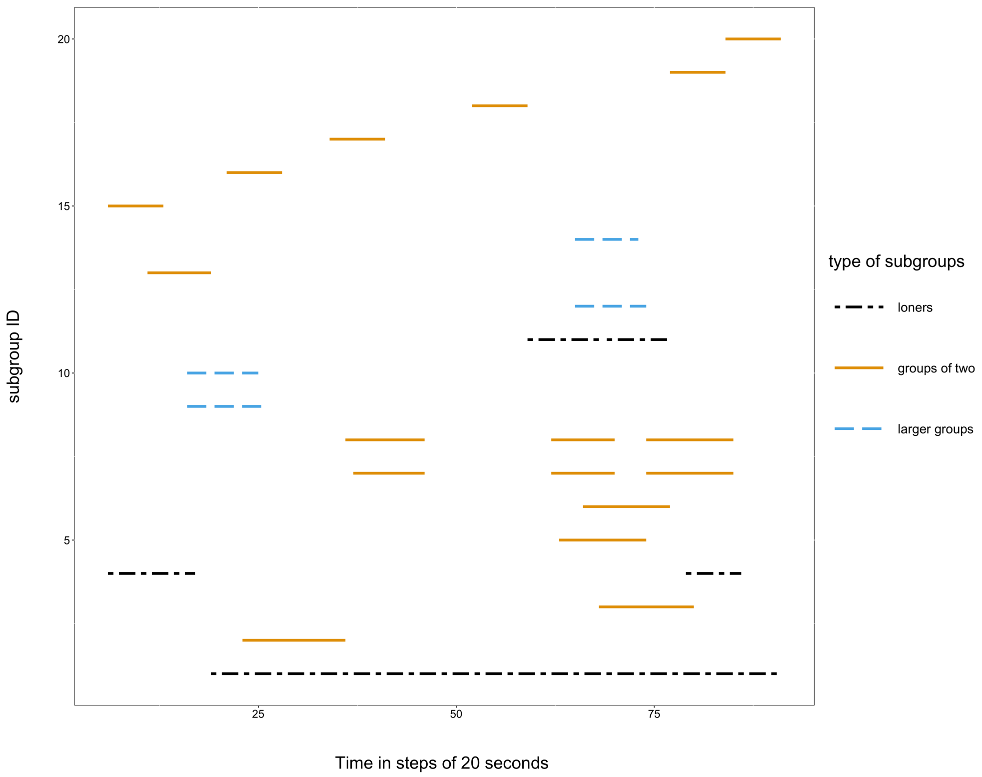

Figure 1: Extracted subgroups from proximity data. The graph shows 20 subgroups covering at least seven time steps (i.e., at least 140 seconds). Three children (groups 1, 4 and 11) spent long periods of time alone. The two re-occuring groups of two (groups seven and eight) show temporally similar behavior.

References

[1] R. Goodman. "The strengths and difficulties questionnaire: A research note". In: *Journal of Child Psychology and Psychiatry* (1997). DOI: 10.1111/j.1469-7610.1997.tb01545.x.

[2] D. S. Messinger, E. B. Prince, M. Zheng, K. Martin, S. G. Mitsven, S. Huang, T. Stölzel, N. Johnson, U. Rudolph, L. K. Perry, B. Laursen, and C. Song. "Continuous measurement of dynamic classroom social interactions". In: *International Journal of Behavioral Development* (2019). DOI: 10.1177/0165025418820708.

[3] C. Rieffe, L. Ketelaar, and C. H. Wiefferink. "Assessing empathy in young children: Construction and validation of an Empathy Questionnaire (EmQue)". In: *Personality and individual differences* (2010). DOI: 10.1016/j.paid.2010.03.046.

[4] N. Shah, D. Koutra, T. Zou, B. Gallagher, and C. Faloutsos. "TimeCrunch: Interpretable Dynamic Graph Summarization". In: *Proceedings of the 21st ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 2015.

[5] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, Lorenzo Isella, Jean-Franc ois Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems. "High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School". In: *Plos One* (2011). DOI: 10.1371/journal.pone.0023176.

[6] G. Veiga, W. de Leng, R. Cachucho, L. Ketelaar, J. N. Kok, A. Knobbe, C. Neto, and C. Rieffe. "Social Competence at the Playground: Preschoolers During Recess". In: *Infant and Child Development* (2017). DOI: 10.1002/icd.1957.