



HAL
open science

Interaction reconstruction methods for large-scale RFID social experiments

Sicheng Dai, Màrton Karsai, Hélène Bouchet, Eric Fleury, Jean-Pierre
Chevrot, Aurélie Nardy

► **To cite this version:**

Sicheng Dai, Màrton Karsai, Hélène Bouchet, Eric Fleury, Jean-Pierre Chevrot, et al.. Interaction reconstruction methods for large-scale RFID social experiments. IC2S2 2019 - 5th International Conference on Computational Social Science, Jul 2019, Amsterdam, Netherlands. hal-02476480

HAL Id: hal-02476480

<https://hal.science/hal-02476480>

Submitted on 24 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interaction reconstruction methods for large-scale RFID social experiments

Sicheng Dai¹, Márton Karsai¹, H el ene Bouchet², Eric Fleury³, Jean-Pierre Chevrot², Aur elie Nardy²
¹ – DANTE, LIP, INRIA Rh one Alpes ; ² – LIDILEM, UGA ; ³ – INRIA Paris

Keywords: temporal networks, signal reconstruction, RFID sensors, sociolinguistic experiment, supervised learning

The emerging technology of radio-frequency identification (RFID) opens entirely new ways to record various aspects of human social interactions in a broad set of settings. First of all, this technology allows to log the temporal dynamics of face-to-face interactions by detecting the physical proximity of participants. The power of this method has been demonstrated earlier in several studies addressing the temporal and spatial evolution of social interactions with high resolution [1, 2]. As examples, such experiments brought us understanding about the structure and evolution of networks formed by face-to-face interactions [3, 4] or how epidemics are spreading in specific environments like in schools or hospitals [2, 5, 6].

Building on these advancements we have taken an uncharted direction by designing RFID tags, which capture not only proximity but also the verbal communication of participants [7]. We use our technology for a large-scale longitudinal sociolinguistic experiment, where we record the oral and face-to-face interactions of children between age 3 to 6 in a preschool setting for one week per month over three years. Our primary goal is to understand how the acquisition of linguistic patterns of children, coming from families of various socioeconomic, cultural, and linguistic background, co-evolve with their social network in time towards a consensus matching the linguistic standards.

There are several challenges in such an experiment starting with the reliable reconstruction of social interactions from the recorded RFID signals. Although this problematic has appeared earlier in other RFID studies, no systematic methods have been proposed so far, despite its importance to determine the outcome of any proximity study based on this technology. Packages exchanged between RFID sensors are sometimes lost due to noise or environmental conditions, which easily cause falsely interrupted or missed interactions or lead to the identification of fake ones. These can considerably bias the observations of the temporal social network and any dynamical process, like language or epidemics, evolving on the top of it. In this contribution our aim is to close this gap by exploring several signal reconstruction strategies using supervised learning methods trained on recorded and annotated data obtained from controlled RFID experiments. Our aim is to find the best method, which provides the highest reconstruction score of temporal interactions for the best approximation of the original social network.

More precisely, we consider the communication between pairs of RFID badges, which may receive a package from each other every five seconds with signal strength depending on the distance and orientation of the hosting people. First, using regression methods, we determine a baseline threshold of signal strength corresponding to a distance-range of 1 - 2 meters of realistic verbal face-to-face interactions. Using this threshold we convert our raw sequence to a sequence of pairs of binary signals indicating mutual simultaneous packages observed between sensors (see pairs of green arrows in Fig. 1b) if they capture a meaningful social interaction, or missing packages otherwise (red arrows in Fig. 1b). Reconstruction from a perfect signal would give us periods of interactions built from consecutive mutually observed packages, separated by

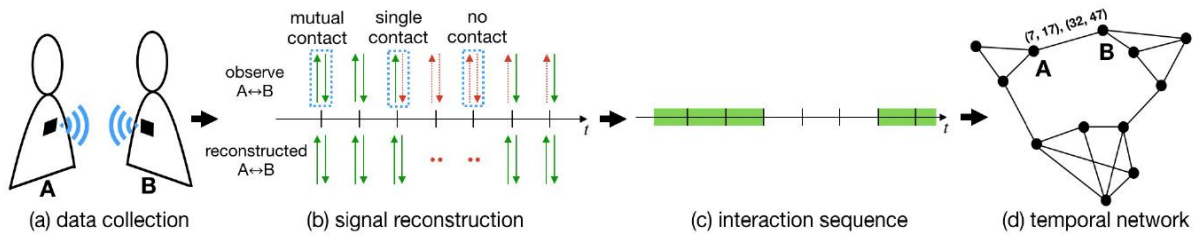


Figure 1: Temporal interaction network reconstruction. Starting from mutually observed packages of pairs of RFID tags (a), we train models using the raw and annotated data (b) to reconstruct interaction and non-interaction periods (c) between individuals to build a temporal network (d).

periods of non-interactions marked as periods with missing packages (see Fig. 1c). However, in reality packages can be lost in either direction during interactions, or falsely present during non-interaction periods, which makes the reconstruction process a great challenge.

To resolve this problem we follow three different strategies to distinguish between patterns of lost packages and non-interaction periods. As a reference measure we use a naive approach assuming that if there is only one missing package or pair of packages between two interaction periods, it is due to package loss thus the two surrounding periods can be merged. This method already gives us 87% reconstruction of the annotated real sequence. To reach better performance we apply two learning methods, which can identify short-term correlations in temporal sequences. We train on the annotated (ground truth) sequences of controlled measurements (a) a Hidden Markov Model (HMM) with short term time envelop before and after the actual package being re-constructed; and a Long Short-Term Memory (LSTM) model with backward memory. While HMM can reach 89% reconstruction precision, our best model is the LSTM, with which we can achieve 95% accuracy.

To demonstrate the importance of accurate interaction reconstruction, we build temporal networks from the raw data and by using the above mentioned reconstruction strategies. We use these networks to perform simulations of Susceptible-Infected spreading processes, a stereotypical model of information diffusion. We show that the speed and final outcome of the spreading is radically depending on the network being reconstructed precisely or not.

These results present methodological tools, which can be readily used [7] on any relational datasets collected by RFID tags. Nevertheless, they are only the first steps towards a larger project to understand how children and child-adult relationships in preschools influence the development of language acquisition and consensus formation among young children.

References

- [1] C. Cattuto, et al., *PLOS ONE* 5(7):e11596 (2010).
- [2] T. Obadia, et al., *PLOS Comp. Biol.* 11(3):e1004170 (2015).
- [3] L. Martinet, et al., *Social Network Analysis and Mining* 8(59):1–33 (2018).
- [4] J. Fournet, A. Barrat, *PLOS ONE* 9(9):e107878 (2014).
- [5] V. Gemmetto, A. Barrat, C. Cattuto, *BMC Infectious Diseases* 14:695 (2014).
- [6] P. Vanhems, et al., *PLoS ONE* 8(9):e73970 (2013).
- [7] *Project and source code will be disclosed upon acceptance.*