

Proceedings of the Ninth International Space Syntax Symposium
Edited by Y O Kim, H T Park and K W Seo, Seoul: Sejong University, 2013

'HUMAN VERSUS MACHINE': Testing validity and insights of manual and automated data gathering methods in complex buildings

053

Kerstin Sailer

University College London
e-mail : k.sailer@ucl.ac.uk

Rosica Pachilova

rosica.pachilova.10@ucl.ac.uk
University College London

Chloë Brown

University of Cambridge

Abstract

With the advancement of information technologies, automated methods of gathering data on space usage patterns in complex buildings using sensors are gaining popularity. At the same time, typical Space Syntax studies still rely on traditional social science methods and manual data gathering, for instance through direct observations and user surveys. How insights generated by each approach compare to each other is still poorly understood. Therefore this paper reports findings from an in-depth two week long study of space usage in a university building, where both manual methods (direct observations, user surveys) and automated data gathering methods (RFID sensors recording locations and interactions of users) were employed in parallel. The main hypotheses to be tested are that automated data captured by RFID sensors delivers comparable findings (1), complementary findings (2) or contradictory findings (3) to direct observations and self-reported surveys. The user behaviour under investigation includes movement flows, patterns of occupancy, interactivity and interaction networks.

Results suggest that variable degrees of overlap can be established between the two approaches with rather few comparable findings. For certain space usage behaviours high levels of variance between the automated and manual datasets are found, pointing towards predominantly complementary and contradictory findings. It is shown that the goodness of the fit between automated and manual data depends on the way data is aggregated. This allows systematic reflections on the strengths and weaknesses of each of the approaches. In summary, evidence suggests that both human and machine based data gathering reveal crucial insights into behaviours of building users. Substituting manual methods with automated ones cannot be supported by the data of this study. Further suggestions for future studies of social life in complex buildings are made, thus contributing to the development of research methods in the field.

Keywords: *Observation methods; Snapshots; Surveys; RFID badges; complex buildings; Social Network Analysis*

Theme: *Modeling and Methodological Developments*

1. Introduction – The Problem of Capturing Social Reality

Space Syntax research is known for its interest in revealing the ‘social logic of space’ (Hillier and Hanson 1984) by combining an analysis of spatial configuration of cities or buildings with an investigation of patterns of usage and collective behaviours.

Traditionally this data on human behaviours is gathered by using observation methods as outlined in the commonly used Space Syntax Observation Manual (Grajewski 1992). For research in complex buildings, typical methods for observations include tracing the routes taken by all space users in a defined area over a specified period of time, thus capturing movement flow or mapping locations and occurrences of typical activities of space usage such as sitting, standing, moving, interacting at one moment in time (called ‘snapshots’). The Manual recommends carrying out those observations at each hour of the day that the specific building in question is in use and observing each area at least twice on two different working days. Data is normally presented in an aggregated way, e.g. as hourly movement flow or density of activities. Both approaches, doing repeat observation rounds and aggregating data, are supposed to ensure data validity, increase accuracy and avoid the effect of distorting patterns by special events.

From the point of view of sociological research methodology, movement traces and snapshots inside buildings can be seen as methods of direct observation and spot sampling (Bernard 2000; Reiss 1971) due to their systematic, structured and quantitative nature as opposed to the more generic term ‘participant observation’, which encompasses a whole variety of different techniques including qualitative, ethnographic and anthropological research (Kawulich 2005).

The main reason for conducting observations with snapshots and movement traces lies in providing data on usage and behaviours that would otherwise not exist. Another important advantage of the method is the very accurate mapping of exact locations of activities. Subtle behavioural differences can be distinguished, for instance between someone standing in front of an office (e.g. looking at a sign), someone standing in the door-frame of an office (e.g. dropping in for a quick question), or someone standing inside an office (e.g. for a longer conversation). These differences can be crucial in understanding the spatial affordances of buildings.

In contrast, various disadvantages and limitations of direct observations need to be considered. Firstly, observers intervene in the field through their presence and could potentially change behaviours, as famously noted by Whyte in his seminal work ‘Street Corner Society’, where the observed population became aware of being observed and reflected on their actions (Whyte 1943). Secondly, sampling and external validity can be an issue (Bernard 2000). Both snapshots and movement traces sample temporally, i.e. a random time is chosen during a specified time slot. Whether and how this data is representative and therefore generalizable is a difficult question. Even though repeating observations can enhance rigour and coherence of the data, often a considerable number of observations is needed to ensure validity (Bernard and Killworth 1993). Thirdly, recording behaviours accurately can become problematic, for instance when people engage in multiple behaviours at the same time (Bernard 2000). In a snapshot for instance it could be debatable whether someone listening passively to a conversation is included in the mapping of an interaction or not. Last but not least, direct observations are time-consuming and involve a lot of manual data processing, since notes from the field (often in tally sheets and on paper) need to be digitised for further analysis.

User surveys are another traditional research method stemming from sociological enquiry. They are frequently employed in Space Syntax research, for instance in the analysis of office buildings, to shed light on interaction and collaboration patterns (as used for instance in: Penn, Desyllas,

and Vaughan 1999; Sailer and Penn 2007, 2009). Surveys, especially when administered online are an efficient and time-saving way to establish insights into peoples' behaviours. Using standardised questions has the advantage of comparability, thus eliminating interviewer bias (Bernard 2000). The main disadvantages of the method, in contrast, include the difficulty to achieve good return rates, the problem of participants interpreting questions differently (Bernard 2000) and the issue of response or recall bias, i.e. participants completing the questionnaire normatively from the point of view of what seems socially desirable, or simply not remembering correctly (Van de Mortel 2008; Bradburn et al. 1978).

Despite their disadvantages, snapshots, movement traces and interaction surveys provide crucial insights into collective behaviours and the social life emerging in buildings. Therefore, they have become a standard approach in Space Syntax research on buildings and space usage.

Recently, the relevance of those traditional and manual methods of data gathering has been challenged by the increasing importance and popularity of using automatically generated and technology-derived data, often also called 'Big Data' (boyd and Crawford 2012) in line with the emerging discipline of 'Computational Social Science' (Lazer et al. 2009). While the majority of big data traverses spatial boundaries and those datasets with a spatial component are particularly relevant at the urban level (see for instance: Batty 2012; Batty et al. 2012; Hossmann, Efstratiou, and Mascolo 2012), some studies are known to investigate patterns of social life inside buildings with the use of technology utilising wearable badges or sensors (Wu et al. 2008; Olguin et al. 2009; Lopez de Vallejo 2009; Heo et al. 2009; Choudhary et al. 2010). However, accuracy and reliability of sensor data, especially on indoor location tracking are at times reported as problematic (Lopez de Vallejo 2009).

Ultimately, this raises a series of questions: How accurately do automated methods record events occurring in space? Which approach – automated or manual data-gathering – best captures the social reality of space usage patterns and behaviours in buildings given that both have their own methodological challenges, advantages and disadvantages? And finally, could automated data-gathering techniques ease and speed up research and thus potentially replace traditional social science data-gathering methods?

To compare and test the validity and quality of insights generated from both manual and automated data-gathering techniques, an experiment using Radio Frequency Identification (RFID) technology was conducted alongside direct observations and an online survey in a two-week long study of movement, occupancy, static activities and interaction patterns conducted in July 2012 in a university building in Cambridge.

The main hypothesis to be tested is whether RFID data delivers 1) comparable and overlapping findings, 2) complementary findings, or 3) contradictory and disparate findings to direct observations and surveys.

The argument will proceed in the following steps: chapter 2 will introduce the RFID technology used in the project. Chapter 3 will present the case study and explain the methodology in more detail. Chapter 4 will highlight the main findings of the project and a final chapter 5 will discuss these reflecting on validity, reliability and quality of insights of the two distinct approaches.

2. Sociopatterns – Exploring Social Life in Buildings with RFID Technology

The RFID system employed in this project is based on the interdisciplinary research collaboration 'SocioPatterns' (<http://www.sociopatterns.org/>), aiming at uncovering fundamental patterns in social dynamics and coordinated human activity. The associated SocioPatterns sensing platform (explained in more depth in: Cattuto et al. 2010) uses active RFID devices embedded in wearable

badges. The badges send signals roughly 8-9 times per second; these are detected by readers installed in the environment and in a peer-to-peer fashion badges detect other badges in close proximity. The proximity setting is tuneable and can range from 1 to 5 metres. Since the badges are worn around the neck and since the human body shields the radio signal, the system mines face-to-face interactions between participants based on the assumption that humans so close to and facing each other would be interacting in some way. The main advantage of this method and system is the accuracy of the temporal data. Using a 20 second interval to establish face-to-face proximity between participants, a probability of more than 99% for an interaction is reported (Panisson et al. 2012).

The SocioPatterns system has been deployed in a variety of buildings and settings¹ including conferences (Panisson et al. 2012), museums and galleries (Van den Broeck et al. 2012), schools (Stehlé et al. 2011) and hospitals (Isella, Romano, et al. 2011). The system was mostly used to gain insights into the dynamics of human behaviour and the structure of social networks of face-to-face interaction with a focus on the modelling of the potential for diseases to spread (Isella, Stehlé, et al. 2011; Isella, Romano, et al. 2011).

3. Case Study, Methodology and Metrics Used

The case study reported in this paper was carried out from 9th-20th July 2012 in a building of the University of Cambridge. The building was constructed in 1999-2001 and is arranged on three floors around two central courtyards. It houses single, double and group offices as well as kitchens, common rooms and seminar rooms on all floors, while the majority of larger facilities, such as the canteen, library, seminar rooms and lecture theatres are located on the ground floor. Figure 1 shows a floor plan of the ground and first floor, which were the ones focused on in the study.



Figure 1: Ground and first floor of the University Building

The study combined the following methods of data collection: 1) Direct observations of two floors of the building, 2) Online questionnaire of interaction and collaboration patterns issued to a selected sample of study participants; 3) RFID sensors capturing face-to-face interactions and tracking locations of study participants.

¹ A full list of associated publications is available from here: <http://www.sociopatterns.org/publications/> (Last accessed: 16 April 2013)

Results from the direct observations were digitised and analysed in QGIS (QuantumGISDevelopmentTeam 2012). Statistical evaluations were done with Excel and JMP. Results on interactions networks (survey-based as well as RFID-based) were analysed with methods of Social Network Analysis (SNA) using Ucinet and Netdraw (Borgatti, Everett, and Freeman 2002). In this paper a segment model was created to calculate shortest paths between the office locations of participants using Segmen (Iida 2009). Shortest paths were calculated based on axial and segment topology (step depth), angle change and metric walking distance to create distance networks between participants, where the value of a tie between participants is the distance between their desk locations. In total four distance networks were created (called DIST_axtopo, DIST_topo, DIST_ang and DIST_metric). This approach is explained by Sailer and McCulloh (2012) in more detail.

In **direct observations** all areas on two floors of the building (ground and first) were covered. For each of the 10 working days of the data collection period, snapshots were completed at 8 different times of the day (within the following time periods: 9-10am, 10-11am, 11:30am-12:30pm, 12:30-1:30pm, 1:30-2:30pm, 3-4pm, 4-5pm and 5-6pm). The following activities were distinguished: sitting, standing, moving (primary activity) as well as interacting and talking on the phone (secondary activity). Movement traces were mapped for an exact time period of 3 minutes for all areas, once within each of the time periods outlined above. In both methods of direct observations, all building users were mapped (rather than just the participants of the study).

The following metrics were computed from the direct observation data:

- *Occupancy*: For each time period and day, the total numbers of people present in snapshots were counted for a selection of rooms (see figure 1), for the study area as a whole and for the whole two floors investigated.
- *Interactivity*: For each time period and day, the total numbers of people interacting in snapshots were counted and divided by the total number of people present to obtain a ratio of interactivity.
- *Movement Flow*: For each time period and day, the total numbers of people passing through the corridors (at specific points, called 'gates') as mapped by movement traces were counted and transformed to hourly movement flow.

The **online survey** was issued during the observation period to 61 study participants, of which 51 completed the survey (84% return rate). The participants were partially academic staff (n=48), located on the first floor of the building and partially members of the administration of the relevant department (n=13), located on the ground floor of the building. Participants in the survey were asked to identify 20 people within the sample they interacted with most and indicate frequency of interaction for planned face-to-face encounter (1), unplanned face-to-face encounter (2), email exchange (3), social media interactions (4) and not-work related socialising (5). Participants were also asked to select up to 20 people they collaborated with (6) indicating the nature of these collaborations (same project/team, co-authors, supervisory relation, exchanging knowledge / ideas); to select whom they consider friends (defined as those one would discuss personal matters with) (7) indicating strength of ties on a scale from 1-5; and to select one person that has a great influence on their work, again indicating tie strength on a scale from 1-5 (8). This means eight distinct networks were created from the survey data. In order to correlate the networks with the RFID derived networks, all networks were transformed from directed to undirected networks by summing up tie strengths. In addition, a network Survey_all was created adding up all tie strengths across the eight distinct networks.

For the **RFID experiment** the same 61 participants wore electronic badges throughout the

duration of the data collection period. Eleven RFID readers were installed in strategic locations (see figure 1). Each reader covered a catchment area of roughly 10-12 metres. The readers captured time and length of signals sent by the badges (therefore highlighting the rough location of participants), as well as close proximity between participants. To achieve a higher resolution on the exact location of participants during the experiment, 26 static badges were attached to the walls within individual rooms to sense occupancy of people in those rooms via the peer-to-peer signal of the badges.

Over the duration of the experiment 729,855 contacts were recorded (badge-to-badge and badge-to-static) of which 273,973 contacts were interactions among participants (badge-to-badge).

The following metrics were computed from the RFID data:

- *Occupancy*: For each time period (coinciding with observations) and each day, the number of distinct people seen within each room was computed. To match the structure of the data derived from direct observations, which capture locations of people at one exact moment in time, a five minute time slot was chosen exactly 15 minutes past the beginning of each observation round (RFID_T1_all). To test reliability and data coherence, a second five minute snapshot was taken at exactly 20 minutes into each observation round (RFID_T2_all). Since a single signal from a badge worn by a person could be received by more than one static room tag, we estimated the chance that a person is in a particular room by calculating the fraction of the detected contacts between the person's badge and each room tag over a 30-second window. Therefore in addition to the RFID_all datasets (all detected contacts between badges and room tags) two more datasets were produced for T1 and T2: on the one hand only the cases where more than half of the contacts between a badge and any room tag over the past 30 seconds involved one particular room, so the person was considered to have a probability of more than 50% of being in that room at that time (RFID_50%), and on the other hand datasets where a person had a probability of more than 90% of being in a particular room (RFID_90%).
- *Interactivity*: For each time period and day, the number of distinct pairs engaging in face-to-face interaction within each room was computed, multiplied by two and divided by the total number of people present to obtain the interactivity ratio. As above, different datasets were created for all contacts (RFID_all), interactions with at least 50% probability (RFID_50%) and 90% probability (RFID_90%) for each of the two time periods T1 and T2.
- *Movement Flow*: For each time period and day, the number of distinct people who caused detection switching from one RFID reader to another was calculated and assigned as hourly movement flow to one of eight 'gates' positioned at midpoint between the locations of two readers (see figure 1). An additional requirement was introduced due to the problem of badges being picked up by different readers simultaneously, i.e. a switch was counted if the most likely reader changed and two consecutive packets were received by the new most likely reader (RFID_flow_all). The dataset without this requirement was analysed as well (RFID_flow_noisy).
- *Interaction Networks*: A network graph was created by assuming a tie between two nodes if the two associated badges have reported contact. Since a contact requires two badges facing each other within a distance of about 1-1.2 metres, there is the possibility that the graph excludes some contacts that actually took place. People could for instance engage in an interaction by talking from opposite sides of a larger room, or by standing side-by-side, so the graph may not account for all interactions. A related source of possible inaccuracy involves the duration of detected contacts. If

contacts are not detected reliably, it is impossible to determine whether two short interactions recorded close in time were part of the same longer meeting or whether they were in fact two separate interaction occasions. To account for this problem two different sets of networks were created, which model different scenarios based on applying different time thresholds. Therefore a contact was either counted as 'continuing' if two badges reported proximity within 30 seconds or less from the last time they reported during that contact occasion (RFID_30sec) or within 5 minutes from the last reporting time (RFID_5min). Contact occasions were then classified as 'passing' (shorter than 30 seconds), 'short' (30 seconds to 2 minutes), or 'long' (2 minutes or more) to evaluate strengths of ties. The sum of the overall duration of the reported contact time is assigned as value to the tie. Therefore, a series of eight basic valued undirected network graphs were created (passing, short, long and all contacts for each RFID_30sec and RFID_5min).

Based on these various datasets and metrics collected with three different methods (direct observations, surveys, RFID sensors), the following chapter will analyse results in order to answer the research question, whether and how the automated method of RFID sensing produces similar or divergent findings to traditional social science methods.

4. Social Life in a Research Environment

Different aspects of space usage and collective human behaviours in the workplace environment studied will be investigated in this chapter. Since all data contains a series of single events with four variables (day, time, location, behaviour), the datasets can be aggregated in different ways to analyse collective behaviours:

- **Group by day and time**, i.e. detailed information on location is disregarded; this investigates behaviours in all locations for each single time period and on each single day;
- **Group by day and location**, i.e. detailed information on times of the day is disregarded; this investigates behaviours and how they fluctuate over the course of the two weeks in each location;
- **Group by time and location**, i.e. detailed information on day is disregarded; this investigates behaviours and how they fluctuate over the course of a working day in each location;
- **Group by location**, i.e. detailed information on day and time of the day is disregarded; this investigates collective behaviours in each location.

The section will discuss the following aspects in detail: 1) Movement flows, i.e. the collective patterns of usage of circulation spaces; 2) Patterns of occupancy, i.e. the location and distribution of people in the building as a whole as well as across different rooms; 3) Patterns of Interactivity, i.e. how many of the people present were interacting on average and 4) Relationship patterns and social networks, i.e. the structure of different sets of interactions and relationships between people.

4.1 Movement Flows

At the time of the study, the building was not in heavy use. There was not much teaching taking place and some academics were attending conferences. This is evident in rather low overall figures for movement flows. The busiest areas of the building were in the main corridor on the ground floor (called 'The Street') near the entrance and the canteen. The majority of the building was much quieter – average observed movement flow at any gate (as captured through 62 strategically located gates) was 27 people per hour and the average for corridors (apart from

the main circulation) was 22 people per hour.

Due to the limited number of RFID readers placed in the building, movement flows from sensor data could only be recorded for eight gates (two on the ground and six on the first floor, see figure 1).

For those eight gates, both the observations and the RFID sensors capture flows in a similar order of magnitude: an hourly flow of 21 people on average per gate is observed, while the sensors record 11 people (RFID_flow_all) and 16 people respectively (RFID_flow_noisy). It makes sense that the observation figures are slightly higher, since observations capture all building users and not just the participants of the study wearing badges.

However, only little overlap is found between observation and sensor data at first sight. If flows are correlated for all single events (i.e. for each day, time and gate separately) (N=640), the correlations are highly significant ($p < 0.0001$), but show a very low coefficient of $R^2 = 0.03$ (for both flow_all and flow_noisy). If data is grouped by day and time (N=78) to show movement flows through all gates, the coefficients rise slightly to $R^2 = 0.07$ with $p < 0.017$ (flow_all) and $R^2 = 0.09$ with $p < 0.007$ (flow_noisy), yet this still does not show much coherence between the two data collection methods. The scattergrams for the correlations show a rather wide-spread distribution of data points around the regression line (as seen in figures 2a and b).

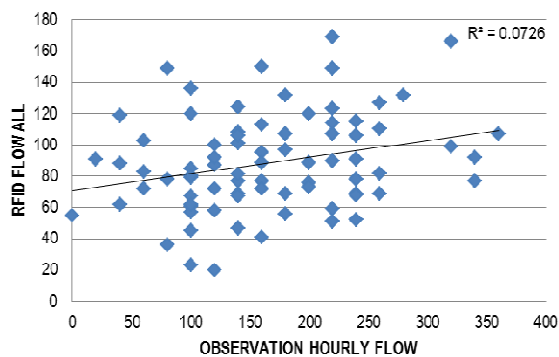


Figure 2a: Correlation between observed and sensor-captured movement flow, aggregated for all gates (flow_all).

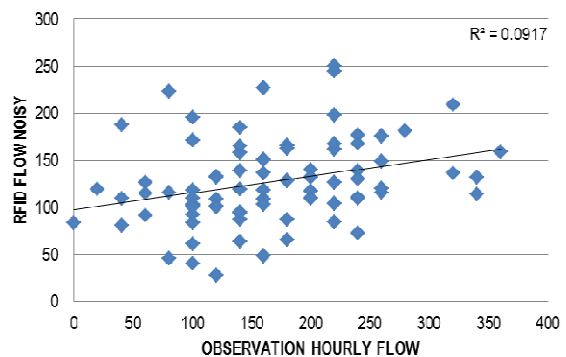


Figure 2b: Correlation between observed and sensor-captured movement flow, aggregated for all gates (flow_noisy).

If data is grouped by gate (N=8), correlation coefficients rise to $R^2 = 0.21$ (flow_all) and $R^2 = 0.20$ (flow_noisy), but the relationship is no longer significant ($p < 0.25$ and $p < 0.26$) due to the small sample size. These correlations are mainly disturbed by rather low sensor figures for the two gates B and F, which can be explained by the location of the gates and the most likely routes of the study participants, which do not lead through those gates a lot of the times (see figure 1). Observation figures are not affected by this, since all building users are counted instead of participants only. If gates B and F are excluded, the correlation coefficients rise to $R^2 = 0.59$ (flow_all) and $R^2 = 0.58$ (flow_noisy), but the relationships are just not significant at the 0.05 level ($p < 0.076$ and $p < 0.077$).

Overall, it can be noted that no relevant differences between the two RFID datasets flow_all and flow_noisy can be detected.

In summary it could be argued that movement flows as captured by observations and sensors do match to some degree and show some trends for correlations. As expected, aggregating data gives better results overall. This is for two main reasons: showing flows for all gates reduces the

discrepancy introduced by the biased distribution of study participants in particular locations of the building, while showing flows for all time periods diminishes the inaccuracy of the temporal sampling of observations. The data seems to suggest that there could potentially be quite a good match between observed and sensor-captured movement flows if data is grouped by gate, however, our data does not allow making this inference properly due to the small numbers of gates and the rather selective nature of the participant sample. We would hypothesize that a study with a more representative sample of building users and a better coverage of the building with more RFID readers and thus more gates would show a significant and high correlation between observed and sensor-captured movement flow.

4.2 Patterns of Occupancy

Looking at the whole study area, on average 64 people were observed in snapshots at any one point in time (as compared to 73 people on the whole two observed floors). As expected, lower figures were obtained from the RFID sensors, since they captured study participants only. On average 27-29 people were recorded (depending on whether T1 or T2 is looked at). The RFID data with 50% and 90% probability of occupancy of participants shows even lower figures of 14-15 people occupying the study areas.

On the level of the whole study area (data grouped by day and time), occupancy data from RFID sensors and observations correlates to some degree. The RFID_90% datasets deliver the best results with $R^2=0.15$ and $p<0.001$ (T1) and $R^2=0.23$ and $p<0.0001$ (T2) as shown in figure 3a and b.

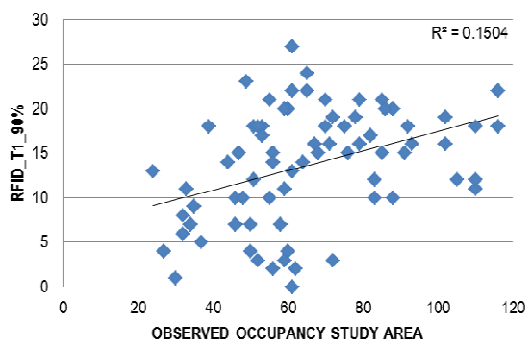


Figure 3a: Correlation between observed and sensor-captured occupancy at T1 for participants with 90% location probability.

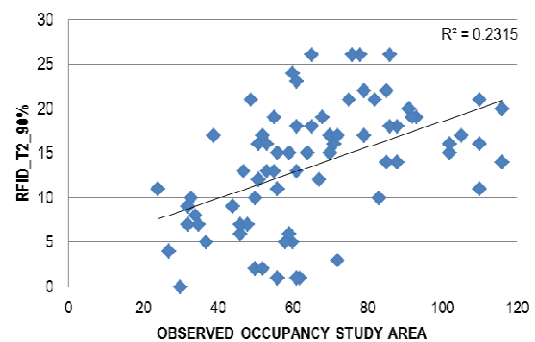


Figure 3b: Correlation between observed and sensor-captured occupancy at T2 for participants with 90% location probability.

The difference between the RFID datasets recorded with exactly the same methods and assumptions but at two slightly different times T1 and T2 is curious and invites further investigation. While we would expect deviations between single events (each time period on each observation day for each room), the overall numbers of people in the study area should remain relatively stable within a 5 minute time window. However, correlations show some mismatches, since $R^2=0.71$ (all), $R^2=0.76$ (RFID_50%) and $R^2=0.78$ (RFID_90%) for the correlations between sensor recorded occupancy at T1 and T2. This means there is a variance of 22%-29% although the two time stamps are just five minutes apart and the study area is sizeable.

Looking at occupancy data in more detail by grouping data reveals interesting insights. Grouping by time and room, no significant correlations show up at all. In contrast, grouping by day and room, moderate correlations appear. Again, the 90% location probability delivers best results with $R^2=0.25$ (T1) and $R^2=0.22$ (T2), both highly significant. If rooms with special functions (printers and social spaces, i.e. kitchens, common rooms, seminar rooms, cafeteria) are excluded and only offices are taken into account, those correlations rise to $R^2=0.32$ for both T1 and T2. Grouping by room, observed and sensor-captured occupancy correlates reasonably well

with $R^2=0.40$ (T1) and $R^2=0.32$ (T2) for the 90% probability datasets (both highly significant). Excluding nine social spaces (therefore $N=16$), increases the correlations to $R^2=0.63$ and $R^2=0.43$, while excluding all thirteen printing areas and social spaces (therefore $N=12$), results in correlations of $R^2=0.72$ and $R^2=0.53$ (all highly significant).

To summarise insights from the analysis of occupancy, some correlations between sensor-captured and observed levels of occupancy can be found, but differences between the two methods overall are still reasonably large. Accounting for locational inaccuracies by looking at the whole study area delivers only small improvements to the match. Reducing temporal inaccuracies, mainly by taking out time as a variable (rather than day) seems to result in better overlaps between sensor data and observation data. This means temporal imprecisions could be the main reason behind the mismatch between sensor and observed occupancy. Finally, disregarding specific locations while errors induced by temporal sampling are taken into account at the same time produces the best correlation results. This could be due to the fact that in social and shared spaces people may not be captured as reliably by the static RFID tags.

4.3 Patterns of Interactivity

On the two floors studied, overall 22.3 people were observed interacting at any one moment in time. This figure reduces to 19.8 people for the whole study area and 5.3 people for the 25 rooms fitted with static RFID tags. RFID sensors captured an average of 2.0 people (T1) and 2.7 people (T2) interacting at any one moment in time across the 25 rooms. Figures are slightly lower for the datasets with 50% probability (1.8 and 2.6) and 90% probability (1.6 and 2.3).

The overall ratio of interactivity obtained from observations was 31% for the building, 34% for the study area and 32% for the studied rooms. The comparable figure of interactivity from sensor data across all studied rooms was significantly lower, i.e. 12% (T1) and 18% (T2)². Therefore it seems that the sensors did not capture every interaction taking place. It is interesting again to distinguish by type of space, since the observed interactivity ratio for offices was 10% (as compared to sensor derived figures of 7% at T1 and 12% at T2); for social spaces it was 61% (versus 17% and 28%); and for printing areas it was 21% (versus 6% and 12%). This means that interactions in offices were more comparable between the two methods than interactions in shared spaces, where the RFID sensors may have missed interactions.

Correlating interactivity ratios as observed with those captured by sensors results in similar findings to the previous section on occupancy, for instance grouping data by room brings up correlation coefficients of $R^2=0.26$ at T1 ($p<0.011$) and $R^2=0.43$ at T2 ($p<0.001$).

In summary, observed interactivity ratios tend to be higher than sensor recorded ones. There is some degree of overlap between the two methods, particularly in types of analysis that diminish the temporal errors of the observation and diminish the recording errors of the sensors, for instance by focusing on specific spaces (i.e. offices) for the whole duration of the study.

4.4 Relationship Patterns and Social Networks

The social networks in the study have 61 nodes with varying numbers of ties: the survey based networks range from 76 ties (Influence) to 618 ties (Survey_all); the RFID based networks have 118 (RFID_30sec_long) to 424 ties (RFID_5min_all) and the distance networks are fully connected, i.e. have 3660 ties. Therefore all networks (apart from the distance networks) are rather sparse, as is common for social networks of interaction and collaboration.

In order to test how much the networks derived from different methods match each other

² The figures for RFID_all and RFID_50% are even lower than this. The 90% figures are used as previously, since they seem to be the most reliable.

structurally, correlations were obtained by correlating the value of a tie that connects each dyad (reported tie strength, duration of contact, or distance) across the different networks following the so called Quadratic Assignment Procedure (QAP) (Krackhardt 1987). Results from the QAP analysis are shown in table 1 below.

It can be seen that some of the survey networks correlate with each other. The same is true for most of the RFID contact networks, which correlate well with each other. The distance networks are highly correlated, too (as expected).

In particular, Survey_all seems to be a good overall illustration of the relationship structures participants reported in detail, e.g. who they see how often in a planned ($R^2=0.78$) and unplanned way ($R^2=0.78$), whom they email ($R^2=0.82$), who they are friends with ($R^2=0.52$) and who they socialise with outside of work contexts ($R^2=0.63$). Therefore, Survey_all could be seen as a representation of most important relationships. The RFID contact networks containing all conversations (RFID_5min_all and RFID_30sec_all) are most representative within the RFID set, however, passing conversations with a 30 second threshold for continuing contact and long conversations with a 5 minute threshold correlate quite well across the dataset of RFID derived contact networks, too.

Table 1: Correlation matrix with coefficients for QAP analysis of all networks; $R^2 < 0.5$ are shown in light grey; $0.5 < R^2 < 0.8$ are shown in bold dark grey; $R^2 > 0.8$ are shown in bold black; negative correlations are shown in green; only significant correlations are shown and significance levels are indicated by * ($p < 0.05$) and ** ($p < 0.01$).

	F2Fplanned	F2Funplanned	Email	Collab bin	Friends	Social	SocMedia	Influence	Survey_All	RFID_5min_passing	RFID_5min_short	RFID_5min_long	RFID_5min_all	RFID_30sec_passing	RFID_30sec_short	RFID_30sec_long	RFID_30sec_all	DIST_ang	DIST_axtopo	DIST_metric	DIST_topo	
F2Fplanned																						
F2Funplanned	.63**																					
Email	.69**	.64**																				
Collab bin																						
Friends																						
Social		.59*																				
SocMedia																						
Influence																						
Survey_All	.78**	.78**	.82**		.52*	.63*																
RFID_5min_passing																						
RFID_5min_short																						
RFID_5min_long								.50*														
RFID_5min_all								.46**	.61*	1.0**												
RFID_30sec_passing								.37*	.80*	.80*	.85**											
RFID_30sec_short								.50*		.92*	.95**	.75**										
RFID_30sec_long																						
RFID_30sec_all								.41*		.88**	.91**	.67**	.83**									
DIST_ang	.34**	.33**	.29**		.20*			.33**	.27**	.20*	.21*	.20**	.23**	.21**				.21**				
DIST_axtopo	.35**	.35**	.31**		.21*			.35**	.26**	.20*	.21**	.20**	.23**	.22**				.20**	.99**			
DIST_metric	.37**	.40**	.43**					.38*	.41**	.30**	.20*	.28**	.23**	.27**	.27**			.23**	.57**	.58**		
DIST_topo	.40**	.41**	.43**	.20*		.20*		.42**	.41**	.29**	.21*	.25**	.22**	.26**	.26**			.22**	.51**	.53**	.97**	

The only reported network structure correlating with sensor derived contact networks is Survey_all with five significant correlations and correlation coefficients ranging from $R^2=0.37$ (RFID_30sec_passing) to $R^2=0.50$ (RFID_5min_long). This means there is roughly a 50% overlap between those reported as most important people at work and what sensors capture regarding

real-time face-to-face encounters in the office environment. Whether the 50%-60% mismatch between the methods is grounded in recall bias or stems from contacts missed by sensors cannot be established. Whether the 30 second or 5 minute threshold for continuing conversations makes more sense in constructing the RFID networks is also not conclusive from the data. However, aggregating all contacts seems to be reasonable to match data with reported networks. Some of the seemingly obvious correlations, for instance between reported face-to-face encounter and sensor-based face-to-face encounter do not materialise, since they just fail the significance tests.³

The distance network with walking distances in metres seems to correlate best across all other networks, which is in line with findings from previous research in small-scale cellular office environments (Sailer and McCulloh 2012). Both reported and sensor derived network structures correlate with distance networks. This means that with increasing distance between pairs of people, interaction intensity and contact duration between them decrease. The fact that correlation coefficients tend to be higher for the reported networks than the RFID contact networks can only be speculated upon. It could be the case that the reality of longer distances creates repercussions in perceptions and is therefore more clearly mirrored in what people report rather than what the sensors capture.

5. Discussion and conclusions

Results presented in this paper suggest that manual and automated data gathering techniques mostly produce disparate findings. Network structures seem to show the strongest overlap between self-reported surveys and RFID sensors (around 40-50%), however, the aspirations of the study to gain in-depth insights into different types of interaction and collaboration networks and to allow investigating network multiplexity could not be realised as easily. Overlaps between observed and sensor-captured occupancy and interactivity ranged from 15%-72% (depending on data aggregation), while overlaps in movement patterns were rather small at 7-9%, yet showed good potential for a better match.

Reflecting on the differences between the methods used in this study, the diverging degrees of overlap in the various data sets may not come as a surprise, especially if comparing observation data with sensor-derived data. Four relevant issues make a comparison difficult: firstly, sample sizes differ, since observations look at all building users, while sensors only capture participants who volunteer to wear a badge. Secondly, temporal resolution and sampling differs, since observations are done as snapshots, while sensors are able to capture extended periods of time. Thirdly, locational resolution differs, since observations are very precise on where activities happen, while sensors can only give rather rough estimates and probabilities. Last but not least, inaccuracies occur for both methods, since observations may suffer from biases introduced by the human observers, while sensors may not capture all activities taking place.

Further limitations of the study include its size. The study was restricted by the amount of readers installed in the environment and the rather small number of participants. Of the approximately 240 people working on the ground and first floors of the building, only 61 were recruited for the study. Additionally, the settings of the sensors could have been experimented with in order to calibrate what is captured and what is missed. For instance defining a contact by the close proximity range of 1-1.2 metres may have been a major cause of mismatching data. Future studies should therefore consider testing different settings and verifying them systematically with observations.

³ The following two correlations come closest to being significant: 1) F2F unplanned and RFID_5min_all (R2=0.28, p<0.080) and 2) F2F unplanned and RFID_30sec_short (R2=0.31, p<0.099)

A major strength of this study lies in the length of the observations, creating a unique dataset in the field of Space Syntax, where typically direct observations are conducted over the course of 2-3 days only. It is also the first known approach to systematically compare and test findings derived from manual and automated methods. Thus, important insights were generated by this study.

One of the main contributions includes highlighting the value of each of the two methods. Since overlap in findings is only partial, this means RFID technology uncovers a whole range of new phenomena of social life in complex buildings, previously not looked at in the context of Space Syntax research. This study argues that both methods – automated and manual – add distinctive value to researching phenomena of human behaviour in complex buildings and as such cannot be substituted for one another that easily. The results of this study thus suggest that replacing manual data gathering methods with automated ones will not necessarily yield the same findings and should therefore be approached with caution.

Another contribution of this paper lies in underlining the need to carefully consider temporal sampling in direct observations. Time seemed to be a major factor in creating inconsistent data. Even in the consistent setup of the RFID sensors, randomly defining two different time slots for the RFID generated snapshots resulted in mismatches of 20-30%, which seems considerable. It is therefore recommended to collect observation data for more than two days and group data by day and location for the analysis in order to decrease the temporal bias induced by direct observations.

By analysing which insights can be obtained by which data collection method and comparing automated with traditional methods, this paper has laid a foundation for the future development of research methods which investigate aspects of usage and human behaviours in complex buildings. This is much needed to further our understanding of how buildings work.

Acknowledgements

The authors would like to thank Daniele Quercia, Ilias Leontiadis and Christos Efstratiou (Computer Laboratory, University of Cambridge) for their help in running the RFID experiment. Parts of the research were funded by an EPSRC Post-Break Award to the lead author (EP/J501396/1 & EP/K503459/1).

References

- Batty, Michael. 2012. "Smart cities, big data." *Environment and Planning B: Planning and Design* 39(2):191-193.
- Batty, Michael, K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali. 2012. "Smart cities of the future." *The European Physical Journal Special Topics* 214 (1):481-518.
- Bernard, H. Russell. 2000. *Social research methods: qualitative and quantitative approaches*. Thousand Oaks: SAGE Publications.
- Bernard, H. Russell, and Peter D. Killworth. 1993. "Sampling in Time Allocation Research." *Ethnology* 32(2):207-215. doi: 10.2307/3773773.
- UCINET 6 for Windows. *Software for Social Network Analysis* 6.387. Analytic Technologies, Harvard.

- boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data. Provocations for a cultural, technological and scholarly phenomenon." *Information, Communication & Society* 15 (5):662-679.
- Bradburn, Norman M., Seymour Sudman, Ed Blair, and Carol Stocking. 1978. "Question Threat and Response Bias." *Public Opinion Quarterly* 42(2):221-234.
- Cattuto, Ciro, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. 2010. "Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks." *PLoS ONE* 5(7):e11596.
- Choudhary, Ruchi, Sonit Bafna, Yeonsook Heo, Ann Hendrich, and Marilyn Chow. 2010. "A predictive model for computing the influence of space layouts on nurses' movement in hospital units." *Journal of Building Performance Simulation* 3(3):171-184.
- Grajewski, Tad. 1992. *Space Syntax Observation Manual* (2001 unpublished revised edition: L. Vaughan). London: UCL Bartlett and Space Syntax Ltd.
- Heo, Yeonsook, Ruchi Choudhary, Sonit Bafna, Ann Hendrich, and Marylyn P. Chow. 2009. "A modeling approach for estimating the impact of spatial configuration on nurses' movement." In *7th International Space Syntax Symposium*, edited by Daniel Koch, Lars Marcus and Jasper Steen. Stockholm, Sweden: Royal Institute of Technology KTH.
- Hillier, Bill, and Julienne Hanson. 1984. *The social logic of space*. Cambridge: Cambridge University Press.
- Hossmann, Theus, Christos Efstratiou, and Cecilia Mascolo. 2012. "Collecting big datasets of human activity one checkin at a time." In *Proceedings of the 4th ACM international workshop on Hot topics in planet-scale measurement*. Low Wood Bay, Lake District, UK: ACM.
- Segmen Reference Manual* 0.65. Bartlett School of Graduate Studies, London.
- Isella, Lorenzo, Mariateresa Romano, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Wouter Van den Broeck, Francesco Gesualdo, Elisabetta Pandolfi, Lucilla Ravà, Caterina Rizzo, and Alberto Eugenio Tozzi. 2011. "Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors." *PLoS ONE* 6 (2):e17144.
- Isella, Lorenzo, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. 2011. "What's in a crowd? Analysis of face-to-face behavioral networks." *Journal of Theoretical Biology* 271 (1):166-180.
- Kawulich, Barbara. 2005. *Participant Observation as a Data Collection Method*. Forum Qualitative Social Research (FQS) 6 (2): Art. 43, <http://www.qualitative-research.net/index.php/fqs/article/view/466/997>.
- Krackhardt, David. 1987. "QAP partialling as a test of spuriousness." *Social Networks* 9 (2):171-186.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science." *Science* 323 (5915):721-723.
- Lopez de Vallejo, Irene. 2009. *Measuring Spatial and Temporal Features of Physical Interaction Dynamics in the Workplace*. Doctoral Dissertation, Bartlett Faculty of the Built

Environment, University College London, London.

- Olguin, Daniel O., Benjamin Waber, Kim Taemie, Akshay Mohan, Koji Ara, and Alex Pentland. 2009. "Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior." *Systems, Man, and Cybernetics, Part B: Cybernetics*, IEEE Transactions on 39 (1):43-55.
- Panisson, André, Alain Barrat, Ciro Cattuto, Wouter Van den Broeck, Giancarlo Ruffo, and Rossano Schifanella. 2012. "On the dynamics of human proximity for data diffusion in ad-hoc networks." *Ad Hoc Networks* 10 (8):1532-1543.
- Penn, Alan, Jake Desyllas, and Laura Vaughan. 1999. "The space of innovation: interaction and communication in the work environment." *Environment and Planning B: Planning and Design* 26 (2):193-218.
- Quantum GIS Geographic Information System 1.8*. Open Source Geospatial Foundation Project.
- Reiss, Albert J. 1971. "Systematic Observations of Natural Social Phenomena." In *Sociological Methodology*, edited by Herbert Costner, 3-33. San Francisco: Jossey-Bass.
- Sailer, Kerstin, and Ian A. McCulloh. 2012. "Social Networks and Spatial Configuration - How Office Layouts Drive Social Interaction." *Social Networks* 34 (1):47-58. doi: <http://dx.doi.org/10.1016/j.socnet.2011.05.005>.
- Sailer, Kerstin, and Alan Penn. 2007. "The performance of space – Exploring social and spatial phenomena of interaction patterns in an organisation." In International Architecture + Phenomenology Conference. Haifa, Israel.- 2009. Spatiality and transpatiality in workplace environments. In *7th International Space Syntax Symposium*, edited by Daniel Koch, Lars Marcus and Jasper Steen. Stockholm, Sweden: Royal Institute of Technology KTH.
- Stehlé, Juliette, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems. 2011. "High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School." *PLoS ONE* 6 (8):e23176.
- Van de Mortel, Thea F. 2008. "Faking it: social desirability response bias in self-report research." *Australian Journal of Advanced Nursing* 25 (4):40-48.
- Van den Broeck, Wouter, Marco Quaggiotto, Lorenzo Isella, Alain Barrat, and Ciro Cattuto. 2012. "The Making of Sixty-Nine Days of Close Encounters at the Science Gallery." *Leonardo* 45 (3):285-285.
- Whyte, William F. 1943. *Street Corner Society: The Social Structure of an Italian Slum*. Chicago: University of Chicago Press.
- Wu, Lynn, Benjamin Waber, Sinan Aral, Erik Brynjolfsson, and Alex Pentland. 2008. *Mining face-to-face interaction networks using sociometric badges: Predicting productivity in an IT configuration task*. SSRN (<http://ssrn.com/abstract=1130251>)