Analyzing Group Interaction on Networks of Face-to-Face Proximity using Wearable Sensors

Martin Atzmueller
Tilburg University (CSAI),
University of Kassel (ITeG)
m.atzmuller@uvt.nl

Lisa Thiele
TU Braunschweig
Institute of Psychology
lisa.thiele@tu-braunschweig.de

Gerd Stumme
University of Kassel
ITeG Research Center (KDE)
stumme@cs.uni-kassel.de

Simone Kauffeld
TU Braunschweig
Institute of Psychology
s.kauffeld@tu-braunschweig.de

***Abstract*—The Internet of Things as well as smart devices enable the capture of multi-modal social interaction data at large-scale. For understanding the behavior of the involved actors, as well as to enable structural modeling, the analysis of the according social interaction networks is essential. In contrast to standard approaches that capture social network data using questionaires, this paper describes an analysis at larger scale using sensor data collected by Radio Frequency Identification (RFID) tags. We complement it by informant self-report data obtained using surveys. We focus on the social network of a students' freshman week, and investigate research questions concerning the communication behavior and structure, gender homophily, and inter-relations of sensor-based (RFID) and self-report social networks.**

I. Introduction

With the emergence of the Internet of Things (IoT) the collection of multi-modal social interaction data is enabled at unprecedented scale. For providing insights into human behavior, the analysis of social interaction structures, patterns, and their dynamics is an important task. This is enabled by ubiquitous and mobile devices, sensor networks, and in particular using wearable sensors, cf. [1]–[4].

In this paper, an adapted and substantially extended revision of [5], we present an analysis of social interactions on networks of face-to-face proximity complemented by self-report data in the context of a students' freshman week. This combination enlarges the focus on individual attributes and interrelationships due to social processes [6]. We observed a set of freshman (psychology) students during their first week at university. We collected two types of network data: Person-to-person (face-to-face) interaction using (1) self-report questionnaires and (2) active RFID tags with proximity sensing, cf. [2]. Such a combined content-based analysis of the different networks can provide more general and more reliable understanding and insight into the actual interactions and their subjective motivation. We focus on structural and dynamic behavioral aspects as well as gender homophily. Furthermore, we investigate the relation of social interaction networks of face-to-face (F2F) proximity and networks based on self-reports (SRN), extending the analysis in [7]. The applied RFID tags were developed by the Sociopatterns[1] consortium. They provide the unique feature of measuring face-to-face proximity interactions [2] that can be used for constructing a social interaction network (cf. Section IV for more details). We consider F2F proximity as a proxy for actual face-to-face communication between participants. On the students' freshman week, we can study group interaction and dynamics at large scale, since the event took place on five consecutive days. Our insights can provide important clues for the organization and impact of such an introductory week – one of its major goals is networking and establishing contacts between students, for facilitating the start of studies, students' collaboration and their benefits of one another.

Our analysis focuses on three research questions:

1) Can we observe distinctive structural and behavioral patterns in the face-to-face proximity network with respect to the freshman week?
2) Can we identify gender homophily effects on F2F?
3) How is the F2F interaction network structurally associated with the multiple SRNs?

Summarizing our results, we show that there are distinctive structural and behavioral patterns in the face-to-face proximity network corresponding to the activities of the freshman week. In particular, we analyze the evolution of the contact structure, as well as the individual connectivity and densification of the network according to the phases of the event. Furthermore, we show the influence of gender homophily on the face-to-face proximity activity. Finally, our results also show a correlation between F2F and SRN. These results indicate that there are relations between the different networks, also concerning different types of communication. However, F2F and SRN are still complementary and should both be considered for analyzing the contact network data.

The rest of the paper is structured as follows: Section II discusses related work. After that, Section III briefly summarizes basic definitions and concepts of graph and network theory used throughout the paper. Next, Section IV summarizes the applied method; we describe the data collection setup, the collected datasets, and discuss data validity. Then, Section V presents the analysis results and discusses these in detail. After that, Section VI discusses limitations of our study. Finally, Section VII concludes with a summary and outlines interesting directions for future work.

[1] http://www.sociopatterns.org

978-1- 5386-1208-8/18/$31.00

II. RELATED WORK

The analysis of human contact patterns and their underlying structure is an interesting and challenging task. An analysis, e. g., using proximity information collected by bluetooth devices as a proxy for human proximity is presented in [8]. However, given the range of interaction of bluetooth devices, the detected proximity does not necessarily correspond to face-to-face contacts [2].

The SocioPatterns collaboration developed an infrastructure that detects close-range and face-to-face proximity (1-1.5 meters) of individuals wearing proximity tags with a temporal resolution of 20 seconds [9]. This infrastructure has been deployed in various environments for studying the dynamics of human contacts, e. g., conferences [9]–[11], schools [12], [13], museums [14] and workplaces [15], [16]. For the latter, [15], [16] also analyze RFID and survey data. Similarly, [10] analyzes the interactions and dynamics of the behavior of participants at conferences; the connection between research interests, roles and academic jobs of conference attendees is further analyzed in [11]. The Sociopattern framework also provides the technical basis of our Conferator system [17], [18] – a social conference guidance system. Another approach for observing human face-to-face proximity and communication is the Sociometric Badge [19].[2] It records more details of the social interaction, but requires significantly larger devices.

The analysis of interaction and groups, and their evolution, respectively, are prominent topics in social sciences, e. g., [20]–[25]. The temporal evolution of contact networks is analyzed, for example, in [26]. Also, the evolution of social groups has been investigated in a community-based analysis [27] using bibliographic and call-detail records. Furthermore, the analysis of link relations and their prediction is investigated in, e. g., [28]–[30]. Overall, social interaction networks in online and offline contexts, important features, as well as methods for analysis are summarized in [31].

In contrast to the approaches summarized above, this paper focuses on networks of face-to-face proximity (F2F) at a students' freshman week, combining RFID-based networks of a newly composed group with networks obtained by self-reports (SRN). To the best of the authors' knowledge, this is the first time that such an analysis has been performed using real-world networks of face-to-face proximity of a newly composed group together with the corresponding questionaire data. Specifically, we analyze patterns and interaction dynamics in those networks, investigate gender homophily and assess the relations between F2F and SRN.

III. BACKGROUND

Below, we summarize basic definitions and concepts of graph and network theory used throughout the paper. For more details, we refer to standard literature, e. g., [32]–[34]. For instantiations, we used the *igraph* and *sna* software packages of the *R* platform for statistical computing.[3]

[2]http://hd.media.mit.edu/badges
[3]http://www.r-project.org

A. Networks & Graphs

An (undirected) *graph* $G = (V, E)$ is an ordered pair, consisting of a finite set V containing the *vertices* (or *nodes*), and a set E of *edges* (or *connections*) between the vertices, with $n := |V|$, $m := |E|$. In a *directed graph*, E denotes a subset of $V \times V$. For simplicity, we write $(u, v) \in E$ in both cases for an edge belonging to E. We represent a (social) *network* as a graph, and use the terms synonymously in the following. A *weighted* graph is a graph $G = (V, E)$ together with a function $\mathrm{w} : E \to \mathbb{R}^+$ that assigns a positive weight to each edge. For the *adjacency matrix* $A \in \mathbb{R}^{n\times n}$ with $n = |V|$ holds $A_{ij} = 1$ ($A_{ij} = w(i, j)$) iff $(i, j) \in E$ for $i, j \in V$, assuming a bijection from $1, \ldots, n$ to V.

The *degree* $\deg(i)$ of a node i in a network is the number of connections it has to other nodes, i. e., $\deg(i) := |\{j \mid A_{ij} = 1\}|$. In weighted networks, we complement the degree of a node i by its strength $\mathrm{s}(i) = \sum_j A_{ij}$, i. e., the sum of the weights of the attached edges. The *density* d of a graph is the propertion of the set of all possible edges that are actually present: for undirected graphs, $d := \frac{2m}{n(n-1)}$, for directed graphs $d := \frac{m}{n(n-1)}$. The *clustering coefficient* (or transitivity) C_v [35] for a vertex $v \in V$ in a graph $G = (V, E)$ is defined as the fraction of possible links among v's neighbors which are contained in E. It quantifies how densely the neighborhood of a node is connected. The graph's clustering coefficient C is given by the mean of all node's clustering coefficients. A *path* $v_0 \to_G v_n$ of *length* n in a graph G is a sequence $v_0, \ldots, v_n$ of nodes with $n \geq 1$ and $(v_i, v_{i+1}) \in E$ for $i = 0, \ldots, n-1$. A *shortest path* between nodes u and v is a path $u \to_G v$ of minimal length. The *diameter* $dia(G)$ of a graph G is the largest shortest path distance between any pair of nodes contained in G. A *strongly connected component* of G is a subset $U \subseteq V$, such that $u \to_{G^*} v$ exists for every $u, v \in U$. A *weakly connected component* is defined accordingly, ignoring the direction of edges.

B. Centrality Measures

In network theory, the centrality of a node $v \in V$ in a network G is usually an indication of how important the vertex is [34]. The betweenness centrality *bet* measures the number of shortest paths of all node pairs that go through a specific node. $bet(v) = \sum_{s\neq v\neq t\in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$. Hereby, σ_{st} denotes the number of shortest paths between s and t and $\sigma_{st}(v)$ is the number of shortest paths between s and t passing through v. Thus, a vertex has a high *betweenness centrality* if it can be found on many shortest paths between other vertex pairs.

The closeness centrality *clos* considers the length of these shortest paths. Then, the shorter its shortest path length to all other reachable nodes, the higher a vertex ranks: $clos(v) = \frac{1}{\sum_{t\in V\setminus v} d_G(v,t)}$. $d_G(v, t)$ denotes hereby the geodesic distance (shortest path) between the vertices v and t.

The eigenvector centrality *eig* of a node is an important measure of its influence. Intuitively, a node is central, if it has many central neighbors. The eigenvector centrality $eig(v)$ of node v is defined as $eig(v) = \lambda \sum_{\{u,v\}\in E} eig(u)$, where $\lambda \in \mathbb{R}$ is a constant.

C. Comparing Graph Structures and Rankings

Besides standard statistical measures of correlation and ranking, e. g., [36], we apply the quadradic assignment procedure [37] (QAP) for comparing network structures. For comparing two graphs G_1 and G_2, it estimates the correlation of the respective adjacency matrices [37] and tests a given graph level statistic, e. g., the graph covariance, against a QAP null hypothesis. QAP compares the observed graph correlation of (G_1, G_2) to the distribution of the respective resulting correlation scores obtained on repeated random row and column permutations of the adjacency matrix of G_2.

Furthermore, we apply Cohen's Kappa [38] for comparing directed network structures denoting the agreement on their adjacency matrices. That is, we consider each directed link as a rating of an observer, depending on the respective direction, cf. [7]. We define a $n \times n$ confusion matrix $F \in \mathbb{R}^{n \times n}$; an entry F_{ij} defines the number of cases that the first observer assigned a particular case to category i and the second observer to category j. Then, F_{jj} measures the number of agreements for category j. The observed (proportional) agreement is given by $P_o := \frac{1}{N} \sum_{j=1}^{k} F_{jj}$, and the expected proportional agreement by $P_e := \frac{1}{N^2} \sum_{i=1}^{k} r_i c_i$, where $r_i := \sum_{j=1}^{k} F_{ij}$ and $c_j = \sum_{i=1}^{k} F_{ij}$ are the row and column totals for categories i and j, respectively. Then, the final Cohen's Kappa measure is the following: $\kappa = \frac{P_o - P_e}{1 - P_e}$.

IV. METHOD

Below, we describe the method applied for observing social interactions. We discuss context, setup, and describe the applied dataset. Finally, we discuss data collection and validity.

A. Context and Setup

We examined the first week of freshman students at a psychology degree program. This freshman week is organized as a special course (five days) before the regular courses start, with a total attendance time of about 25 hours. The course aims to provide the new students with relevant information about the university, the degree program, and its contents. Furthermore, professors and other lecturers, the department chairs, and important committees are introduced. In particular, this week offers a major opportunity to become acquainted with fellow students. For the freshman week that we analyze in the context of this paper, 75% of the time (i. e., about 19 hours over five days) the events took place in a separate facility, which was suitable for the intended data collection and technically equipped for this purpose.

The structure of the freshman week included organized plenary sessions and 'free sessions'. The first day consisted of a general introduction (plenary) and a special introductory (free) session helping students to get to know each other. In the following days, plenary sessions mixed with 'free sessions' took place. Figure 1 shows the contact activity during the freshman week. In particular, on the first day (Monday) the students got welcomed, spent time to get to know one another and got the relevant overall information about the studies. On Tuesday students were introduced to the various departments, which was continued on Wednesday. On Thursday, students were given information about possible post-graduate occupational areas. Finally, on Friday the students got information about the exams and chose the cources for their first semester. Further, they got much time to interact freely during an one hour lunch buffet. As shown in Figure 1, the time spent in the equipped accomodation varied over the five consecutive days.

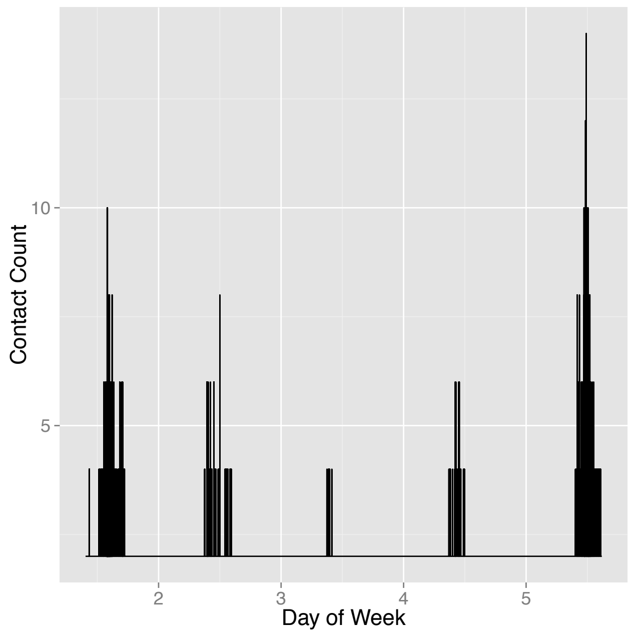

Fig. 1. Overview on the contact behavior: Contact count (per second) during the freshman week.

B. Dataset

The dataset contains data from 77 students (60 females and 17 males) attending the introductory freshman week. We asked each student to wear an active RFID tag while they were staying at the facility. In total, the dataset contains 16780 proximity contacts, i. e., all detected contacts during the indoor activities including breaks and intervening periods. Moreover, also close smoking areas, the garden and the outer entrance area could be reached as well.

Using the F2F proximity networks, we generated a set of undirected networks $F2F(i)$ using a minimal contact duration i in order to distinguish weaker and stronger ties. An edge $\{u, v\}$ is created, iff a face-to-face contact with a duration of at least i *seconds* among participants u and v was detected ($i \in \{0, 60, 180, 300\}$. While in general topical importance in conversation varies, e. g., [39], we utilize the methodological approach using increasing minimal contacts lengths for focusing on stronger ties, cf. [10], [12]: As presented, e. g., in [10], [11] longer contacts tend to correlate with more homophily-induced conversations. Furthermore, the 'filter' thresholds were selected acording to the thresholds used for detecting the end of a contact, and on the empirical fact that there

TABLE I

HIGH LEVEL STATISTICS FOR NETWORKS $F2F(i)$ OF FACE-TO-FACE PROXIMITY WITH A MINIMAL CONTACT (DURATION) THRESHOLD i, COLLECTED USING THE RFID DEVICES. STATISTICS ARE DETERMINED ACCORDING TO THE AGGREGATED CONTACT LENGTH AND MINIMAL INDIVIDUAL CONTACT THRESHOLDS i: NUMBER OF NODES AND EDGES, AVERAGE DEGREE, AVERAGE STRENGTH (WEIGHTED DEGREE), AVERAGE PATH LENGTH APL, DIAMETER dia, DENSITY d, CLUSTERING COEFFICIENT C, AVERAGE BETWEENNESS, EIGENVECTOR AND CLOSENESS CENTRALITIES, NUMBER AND SIZE OF THE LARGEST WEAKLY CONNECTED COMPONENT #CC AND $|\mathrm{CC}|_{\max}$, RESPECTIVELY.

Network	$\|V\|$	$\|E\|$	∅deg.	∅str.	APL	dia	d	C	∅bet.	∅eig.	∅clos.	#CC	$\|\mathrm{CC}\|_{\max}$
$F2F(0)$	77	1622	42.13	67286.55	1.45	3	0.55	0.65	48.89	0.14	$9.6 \cdot 10^{-05}$	1	77
$F2F(60)$	77	1176	30.55	60613.71	1.61	3	0.40	0.51	45.16	0.12	$3.8 \cdot 10^{-05}$	1	77
$F2F(180)$	76	592	15.58	44404.00	1.90	3	0.21	0.33	53.68	0.10	$1.1 \cdot 10^{-05}$	1	76
$F2F(300)$	75	374	9.97	33882.88	2.20	4	0.13	0.27	68.92	0.08	$5.8 \cdot 10^{-06}$	1	75

were breaks of about 300 seconds between longer items of the schedule of the freshman week. For each edge $\{u, v\}$, we determine a weight according to the *sum* of all according contact durations between u and v. Note that we chose to aggregate contacts over the whole event (in contrast, for example, to the procedure in [14]), due to the total short duration of the observed time during the event. Table I contains summary statistics for $F2F(i)$, $(i \in \{0, 60, 180, 300\}$.

C. Collection and Validity of RFID data

The RFID deployment at the freshman week utilized a variant of the MYGROUP [18] system for data collection. Participants volunteered to wear active RFID proximity tags, which can sense and log the close-range face-to-face proximity of individuals wearing them. During the freshman week, participants then integrated the RFID tags into their name tags. This setup allowed us to map out time-resolved networks of face-to-face contacts among the attendees.

A proximity tag sends out two types of radio packets: Proximity-sensing signals and tracking signals. Proximity radio packets are emitted at very low power and their exchange between two devices is used as a proxy for the close-range proximity of the individuals wearing them. Packet exchange is only possible when the devices are in close enough contact to each other (1-1.5 meters). The human body acts as an radio frequency shield at the carrier frequency used for communication [9]. As in [9], we record a face-to-face contact when the length of a contact is at least 20 seconds. A contact ends when the proximity tags do not detect each other for more than 60 seconds. The tracking signals are received by antennas of RFID readers installed at fixed positions in the facility environment. These tracking signals are then used to relay proximity information to a central server.

With respect to the accuracy of the applied RFID tags, we refer to the results of Cattuto et al. [9] who confirm (1) that if the tags are worn on the chest, then very few false positive contacts are observed, (2) face-to-face proximity can be observed with a probability of over 99% using the interval of 20 seconds for a minimal contact duration. This is in the range of human inter-annotator-agreement [40]. Compared to their experiments, our setup is even more conservative since we use a threshold of 60 seconds when determining the end of a contact. Furthermore, it is important to note that we focus on face-to-face proximity as a proxy for actual communication; due to the applied thresholds (see above), face-to-face proximity situations which include episodes that are, e. g., briefly side-by-side or over the shoulder, can typically also be captured.

D. Collection and Validity of Self-Report Data

At the very end of the week we captured all the self-reported communication network ties: Each student was handed out exhaustive name lists. Then, we asked the students to select those fellow students, with whom (1) they interacted much during the introductory course, (2) they would like to cooperate, (3) they would ask for advice (mentoring). Using this data we modeled according matrices denoting directed networks with links from respondents (rows) to persons to be selected (columns).

Although the RFID-data contain almost all interactions taking place at the equipped venue, the self-reports also contain the times that could not be captured (e.g., the evenings in a restaurant), cf. [7]. Of course, communication and social interaction self-reports underly certain biases, cf. [41]–[43]. When asking the students to mark their fellows on an exhaustive name list, whom they communicated much during the introductory week, we are faced with several cognitive filters. The understanding of "much", memory effects and the emotional weighting of contacts influence the informant data. However, these cognitive filters are very relevant [7], [42], [44], as they capture the perceived importance of a tie. Therefore, we explicitly interpret the survey data as the interaction network as perceived by the students.

V. ANALYSIS

In the following we present the analysis results focussing on the three research questions outlined above and discuss the results in detail. In particular, we investigate

1) structural and behavioral patterns of F2F,
2) aspects of gender homophily of F2F, and
3) structural associations between F2F and SRN.

A. Structural Patterns - Contact Network

We first consider structural aspects of the contact network. After that, we investigate homophily effects.

1) Distributional Contact and Degree Patterns: Table I shows that the contact network is well connected, with a rather low diameter ($d = 3$). As expected, the density of the network is reduced with an increasing minimal contact (duration) threshold i, while the diameter remains relatively stable.

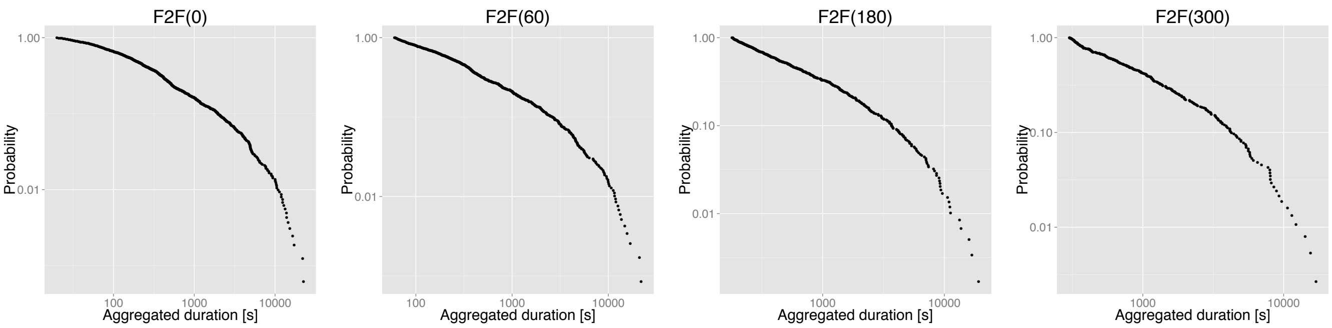

Fig. 2. Aggregated contact lengths for the networks $F2F(i), i \in \{0, 60, 180, 300\}$ seconds.

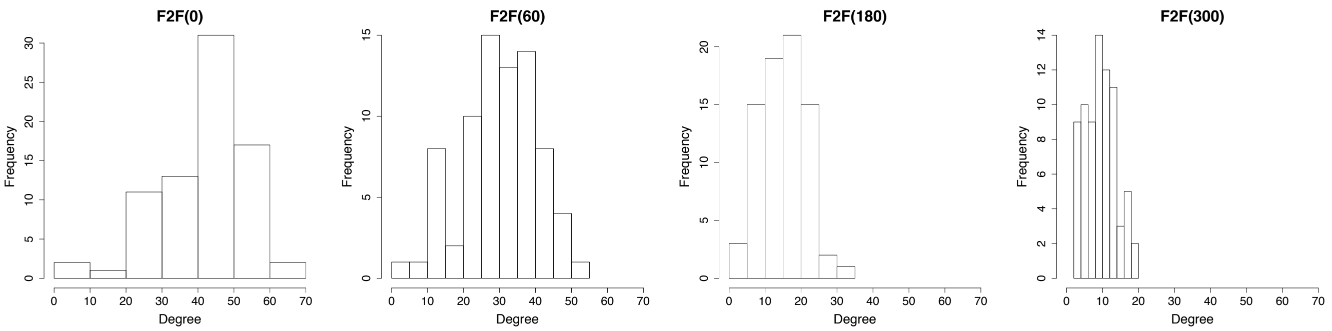

Fig. 3. Degree distributions for the networks $F2F(i), i \in \{0, 60, 180, 300\}$ seconds.

Furthermore, we observe an increasing average betweenness centrality for longer conversations, while the eigenvector centrality is slightly decreasing. Considering $F2F(300)$ the average degree is similar to the value of the self-report networks, see Table IV, which already indicates the impact of longer contacts. For increasing minimal conversation thresholds, the average degree is decreasing which can be explained by more focused conversations. This is also in line with the findings of [12], [45] on friendship network structures, group sizes and respective levels of intimacy.

Figures 2 and 3 showing the cumulative aggregated contact length and degree distributions depict these patterns in more detail. The contact length distribution follows a typical long-tailed distribution similar to those observed at conferences [10], [11] – with an increasing minimal contact threshold we observe a shift focusing on the longer conversations. Also, the degree distributions show that an increasing minimal contact threshold helps to select the more “active” participants with respect to a set of diverse contacts, indicated by medium to high degree nodes.

Concerning the degree distribution, we furthermore investigated the trend between the strength and the degree of a node, measured by the average strength $s(k)$ of nodes of degree k in comparison to the degree (k). As described in [26], typically a linear dependency of the average strength of nodes of a certain degree with the average link weight and the degree is expected. Then, a deviation of the trends of the lines shown in Figure 4 indicates some interesting trends: We observe increasing trends for the strength/degree lines that are more pronounced for the networks with lower thresholds, while the 300s network is almost converging to a horizontal trend line. This indicates the importance of certain “super-spreader” nodes with a large degree, cf. [26], that seem to be relatively important for shorter conversations: The increasing trends are stronger for smaller minimum contact thresholds, while they tend to decrease slightly in the network with the largest minimum contact

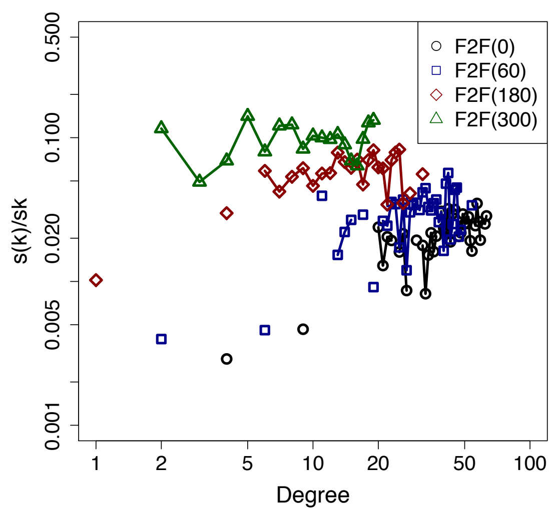

Fig. 4. Strength/degree proportion.

TABLE II

AGGREGATED CONTACT LENGTH STATISTICS AND NETWORK PROPERTIES FOR NETWORKS $F2F(i)$; CONTACTS FOR *female (f)* AND *male (m)* PARTICIPANTS: AVERAGE DEGREE, AVERAGE STRENGTH, AVERAGE BETWEENNESS, EIGENVECTOR AND CLOSENESS CENTRALITIES, RESPECTIVELY.

Network	∅deg.	∅str.	∅bet.	∅eig.	∅clos.
$F2F(0, f)$	43.12	71084.70	49.81	0.16	$9.6 \cdot 10^{-05}$
$F2F(0, m)$	38.65	53881.30	45.65	0.06	$9.5 \cdot 10^{-05}$
$F2F(60, f)$	31.43	64244.57	40.30	0.15	$3.8 \cdot 10^{-05}$
$F2F(60, m)$	27.41	47798.94	62.38	0.05	$3.8 \cdot 10^{-05}$
$F2F(180, f)$	16.15	46942.93	59.00	0.12	$1.1 \cdot 10^{-05}$
$F2F(180, m)$	13.40	34883.00	33.80	0.02	$1.1 \cdot 10^{-05}$
$F2F(300, f)$	10.05	35384.07	67.17	0.10	$0.6 \cdot 10^{-05}$
$F2F(300, m)$	9.67	27878.13	75.93	0.01	$0.6 \cdot 10^{-05}$

threshold (300s). This is also in line with the observation of the decreasing degree distributions for higher thresholds. A possible explanation for these findings concerns the spread of small talk (lower thresholds) in the social interaction network, while more meaningful (longer) conversations are more evenly distributed accross the nodes (higher thresholds). Such super-spreader nodes are rather important in the context of the freshman week for establishing initial connections between participants. A detailed investigation of their descriptive characteristics using SRN and additional information is therefore one major point of future work.

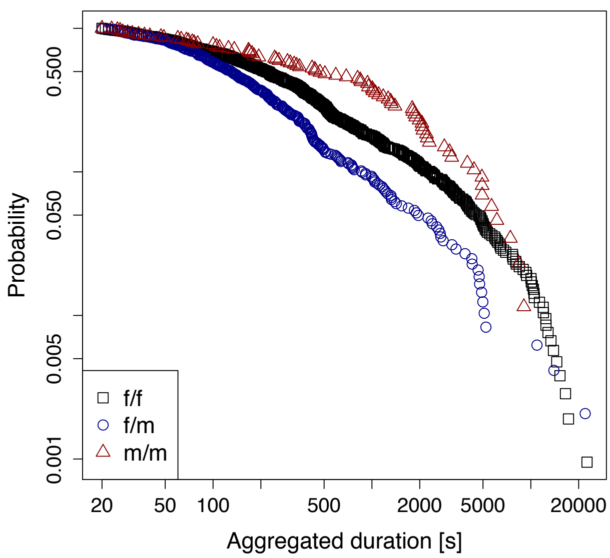

Fig. 5. Aggregated contact lengths for the contact networks between *female (f)* and *male (m)* participants.

2) Homophily Effects: In the following, we investigate gender-related differences in the contact networks. We focused on gender, because it is a very salient attribute in the context of examining psychology freshman in Germany. However, attributes such as age and ethnicity can influence the choice of interaction partners as well. Though, for our sample, we found age and ethnicity to be not varying very much. Table II shows network statistics for a set of aggregated contact networks contructed according to minimum contact thresholds as described above. The degree values of the larger group (females) are always larger than the co-group (males). This is consistent accross the networks induced by different minimal conversation thresholds, and also holds for the different strengths. This can already be regarded as a weak indicator of gender-related differences.

Figure 5 shows the aggregated cumulative contact lengths distributions between female and male participants. As shown in the figure, mixed-gender edges tend to correspond to shorter aggregated contacts compared to interactions between individuals of the same gender - intra-group communication is more frequent (red and black lines) compared to inter-group communication (blue line). Overall, the observations shown in Figure 5 confirm the trends of [12] in a new context: The results indicate, that the aggregated contacts are broadly distributed – there is no typical contact duration for a specific type of contact, and no characteristic time scale can be determined.

Table III shows further network statistics for *female* - *female*, *female* - *male*, and *male* - *male* contacts. As shown in the table, the intra-group networks are much more dense than the inter-group network, while the betweenness centrality in the inter-group network is the highest. A closer look reveals, that the betweenness values are rather unevenly distributed. There are many nodes with a betweenness value of zero (exclusively females), while there are also many with extreme values (e.g., betweenness > 100), a group that is dominated by male participants. This can be explained by the bi-partite graph structure and the smaller share of male participants which then act as important bridges in the graph.

In order to ground the statistical analysis further, we investigated the empirical contact distributions following the approach proposed in [12]: We compare the empirically determined contact ratios to a null model constructed by graphs such that the probability of an edge connecting two nodes is independent of the genders of the nodes, in order to assess the probability that the empirically observed contact networks is generated from such a network structure. For obtaining the null-model values and according confidence intervals we basically apply the approach presented in [46] for generating a set of random networks by rewiring the original one. Figure 6 shows the empirical values of the fraction of edges between *female* - *female* (ff), *female* - *male* (fm), and *male* - *male* (mm) participants, in comparison to the null-model as discussed above. The null model plot for each network covers the 95% confidence interval.

As can be observed from Figure 6, we can reject the null hypothesis of gender independence at the 5% confidence level: The communication edges are not gender-independent, since none of the empirical values fit with the 95% confidence interval of the null models. In line with this observation, the ratio of edges for *female* participants is always above the values obtained from the null model, as well as the edge ratio of *males*. Conforming to the analysis results discussed

TABLE III

AGGREGATED CONTACT LENGTH STATISTICS AND NETWORK PROPERTIES FOR $F2F(0)$: ALL CONTACTS, AND THOSE BETWEEN *female (f)* AND *male (m)* PARTICIPANTS: NUMBER OF NODES AND EDGES, AVERAGE DEGREE, AVERAGE STRENGTH, AVERAGE PATH LENGTH APL, DIAMETER d, DENSITY, CLUSTERING COEFFICIENT C, AVERAGE BETWEENNESS, EIGENVECTOR AND CLOSENESS CENTRALITIES, NUMBER AND SIZE OF THE LARGEST WEAKLY CONNECTED COMPONENT #CC AND $|\mathrm{CC}|_{\max}$.

Network	$\|V\|$	$\|E\|$	∅deg.	∅str.	APL	d	density	C	∅bet.	∅eig.	∅clos.	#CC	$\|\mathrm{CC}\|_{\max}$
$F2F(all)$	77	1622	42.13	67286.55	1.45	3	0.55	0.65	48.89	0.14	$9.6 \cdot 10^{-05}$	1	77
$F2F(f/f)$	60	1052	35.07	63548.33	1.41	2	0.59	0.66	38.42	0.15	$12.0 \cdot 10^{-05}$	1	60
$F2F(f/m)$	77	483	12.55	11744.99	2.03	4	0.17	0	69.56	0.06	$5.7 \cdot 10^{-05}$	1	77
$F2F(m/m)$	17	87	10.23	27282.35	1.38	3	0.64	0.79	14.4	0.45	$28.2 \cdot 10^{-05}$	1	17

above, the ratio of edges for mixed-gender interactions is below the values of the null model. This points to same-gender preferences, similar to results of [12] in the contexts of schools. Furthermore, we observe an impact of stronger ties, since the ratios for single gender contacts increase for increasing minimal conversation thresholds, while the ratios for mixed gender contacts decrease.

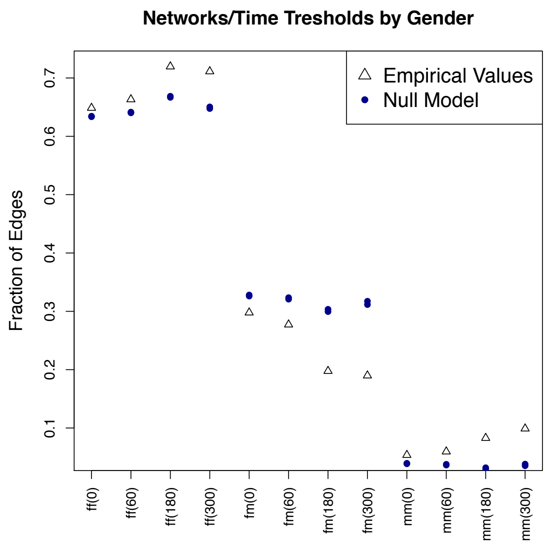

Fig. 6. Comparison of empirical edge ratio vs. null-model results (95% confidence interval) of the contacts of *female* and *male* participants.

B. Communication Behavior and Dynamics

For assessing the dynamic behavior and evolution of contacts, we performed a time-based analysis for which we partitioned the contact networks into time-slices of consecutive minutes and computed the according statistics for the cumulated contact count, the average degree and the density of the respective networks. The analysis results are shown in Figure 7. The figure shows the cumulated contact count, average degree and density, where we focused on the networks $F2F(0)$, $F2F(180)$, and $F2F(300)$ including all contacts vs. the longer minimal contact thresholds. Furthermore, the figure includes a normalized version using the respective maximum value of the individual measures as a reference factor. As can be observed in the figure, the distributions follow the overall contact activity, see Figure 1.

However, there are certain effects that can also be distinguished: As the first day consisted of an introduction and a special introductory session helping students to know each other well, the increase in the contacts and degree is expected. However, on the second and especially the last day there is a significant increase of contacts and degree as well. Furthermore, considering longer contacts we see that the density grows continuously and is again supported by the last day with many interactions, which indicates that the week indeed had a rather positive effect for strengthening interaction and contacts between participants.

Specifically, if we consider the normalized analysis results of degree and density of the respective networks, we observe that there is always a strong increase during the first and last day; however, for longer conversations the effect on the first day is not that strong compared to the last day – the behavior on the first day is mainly concerned with the weaker links – which confirms the expected behavior during the introductory sessions. For all measures, the growth of stronger links is more pronounced especially for the days later in the week, in particular the last day. Furthermore, we can consider the individual structures of the different days, i. e., concerning structured and free sessions: In general, in free sessions we observe significantly more contacts than during the structured sessions, as expected. However, for the last day we also observe a higher fraction of contacts (and also stronger ties as discussed above), while the time available for free interaction of the participants was the same as on the first day. Thus, in comparison to the first day, we observe more frequent and also more intense contact behavior. This shows the overall effect of the freshman week for supporting the groups' networking and contact behavior.

C. Structural Associations between F2F and SRN

As described above, we collected self-report data using questionaires including information about interactions, cooperation and mentoring relations. Table IV shows some statistics regarding the networks as undirected for an easier comparison to the F2F network. For the self-report interaction network (*SRN.int*), for example, we observe that the number of connections (483), diameter (4), the average degree (9.29) and the density (0.17) are rather different compared to the F2F, cf. Table I. We measure rather large deviations concerning the average degree and the density. As expected, this indicates that F2F covers more interactions during the observed time.

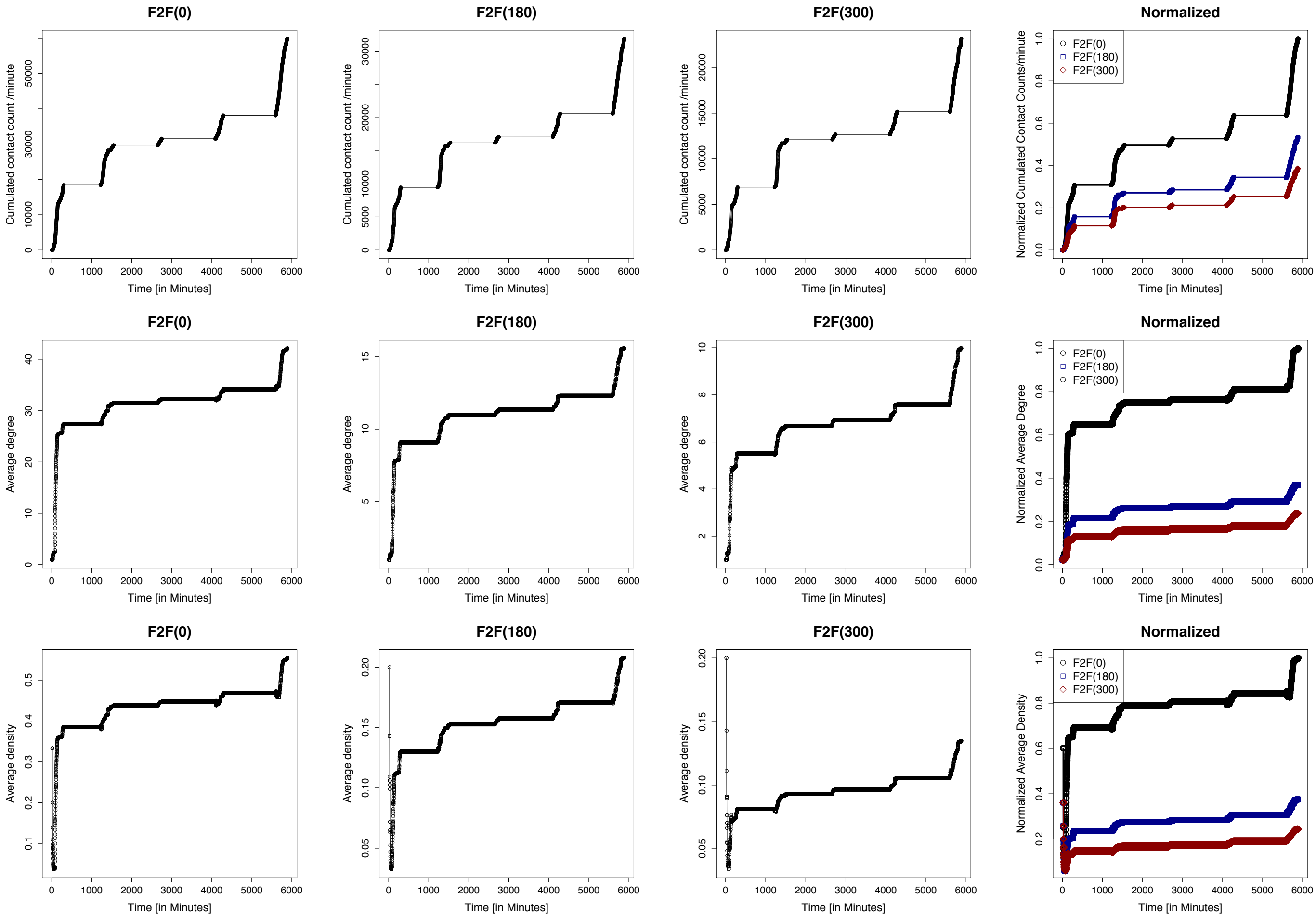

Fig. 7. Time-based analysis: Cumulated contact count/minute, average degree and density for the networks $F2F(0)$, $F2F(180)$, $F2F(300)$, as well as normalized versions; contact count (top), average degree (middle) and density (bottom).

These findings are also true for the cooperation and mentoring networks, while both surprisingly show a diameter that is actually equal to the diameter of F2F.

TABLE IV
OVERALL (UNDIRECTED) NETWORK PARAMETERS OF THE SELF-REPORT NETWORKS. #NODES (V), #EDGES (E), AVG. DEGREE, DIAMETER (D), DENSITY.

Network	$\|V\|$	$\|E\|$	∅deg.	d	density
Interaction	77	483	12.54	4	0.17
Cooperation	77	433	11.25	2	0.15
Mentoring	77	434	11.27	2	0.15

For assessing the networks on an individual level, we postprocessed the F2F such that it becomes comparable with the (directed) SRN.int. We aggregated contact duration and frequency over the whole week for each tag pair. Then, we defined a contact to be meaningful, when its frequency or duration scored above a certain threshold τ_F (frequency, F2F.freq) and τ_D (duration, F2F.dur). We selected the observed average duration of 13:19 (mm:ss) and average frequency 5.17 of all contacts as initial thresholds, cf. [7].

We then tested the matching of the self-report and the postprocessed F2F data by computing Cohen's Kappa [38]: This measure is typically used to assess the average agreement of two observers, e. g., with respect to their ratings of behavior [47]. Accordingly, we treat the self-report and the face-to-face contact data sources as two independent observers, obtaining 3239 comparable ratings. Furthermore, we applied the QAP test for measuring network correlation common in social network analysis. Cohen's Kappa yielded a value of $\kappa = .507$ (SRN.int vs. F2F.dur; 84.8% agreement) and $\kappa = .485$ (SRN.int vs. F2F.freq; 83.5% agreement), suggesting a fair congruence [7], [48]. This is also reflected by a QAP-Test: SRN.int vs. F2F.dur (0.32), SRN.int vs. F2F.freq (0.55), significant at $p < 0.001$. When we varied the thresholds τ_F and τ_D we observed only very small variations in the Kappa values which indicates that the thresholds were chosen rather well for the postprocessed F2F interaction network, also confirmed by the QAP-test values.

For the self-report cooperation and mentoring networks interestingly the matching between F2F and SRN was better for higher duration thresholds, i. e., stronger links (also confirmed by higher duration thresholds). For the cooperation network the best matching was obtained for a threshold $\tau_D^C \geq 20$ minutes, while the best threshold for the mentoring network was even higher, with $\tau_D^M \geq 23.3$ minutes. Again, this indicates the impact of longer conversations and the relation to SRN. This enables the option of deriving SRN information from F2F data given suitable thresholds. When we focus on

TABLE V
DEGREE CORRELATION BETWEEN THE AGGREGATED F2F NETWORK WITH DIFFERENT MINIMAL CONTACT THRESHOLDS AND SELF-REPORT INTERACTION NETWORKS ESTIMATED USING SPEARMAN'S RANK ORDER CORRELATION COEFFICIENTS: *P < .05, **P < .01, TWO-TAILED.

Threshold	Interaction	Cooperation	Mentoring
0	.267*	.292*	.202
180	.419**	.393**	.325**
300	.374	.399**	.370**

the centrality measures, especially on the degree centrality we also observe correlations between F2F and SRN, see Table V focusing on the aggregated contact networks according to different minimal contact thresholds. These confirm our observations for the matching between the networks discussed above. We observe the trend that the larger the (face-to-face) interaction, the higher the chance to be selected for cooperation or mentoring.

VI. LIMITATIONS

Like any empirical investigation, this study has some limitations. First, providing subjects with wearable sensors might bias the social situation and influence the subjects' behavior. However, the subjects' situation of being new at university and being provided with loads of information over a whole week, is very demanding and should let them forget the wearables soon. After the introductory week, the subjects described their behavior as typical and not being influenced by data collection.

Second, asking the participants to select the fellows whom they communicated "much" with during the introductory week is not independent from individual interpretation. Other social network approaches use targeted questions for weighting contacts, such as asking for length or exact frequencies. However, given the situation of 77 people who predominantly never met before and the information flood they were exposed anyway, we decided to use a less demanding network approach. That is, we presumed participants to report the contacts they individually consider important and interpret the self-reports explicitly as this. In the light of implementing RFID technology, we received data about frequency and duration of contacts, so we complement this with the self-reports.

Third, our sample consisted only of a group of psychology students, primarily consisting of young adults with an uneven gender distribution. While the gender and age distribution was representative of the examined field of study, our findings do not necessarily be generalizable to students groups of other courses or more diverse groups. However, we were able to capture data of a whole and newly composed group over time. Moreover, all students filled out the surveys and wore RFID tags. Thus, we achieved a return quote of 100%.

VII. CONCLUSIONS

This paper presented an analysis of social interaction on networks of face-to-face proximity using wearable sensors. The interaction data was collected in the context of a newly composed group, complemented by self-report data of the participants. Both datasets were used for modeling networks capturing physical as well as perceived interactions in order to understand physical and cognitively perceived network structures and the behavior of actors therein.

We analyzed data of a students' freshman week and showed that there are distinctive structural patterns in the F2F data corresponding to the activities of the freshman week. This concerns the evolution of contacts, the individual connectivity in the network, and the densification of the network according to the phases of the event. Furthermore, we showed the effects of gender homophily on the contact activity. Finally, our results also indicate existing structural associations between the face-to-face proximity network and various self-report networks.

In the context of introductory courses, this points out the importance of stronger ties (long conversations) between the students at the very beginning of their studies for fostering an easier start, better cooperativeness and support between the students. Our results especially show the positive effect of the freshman week for supporting the connectivity between students; the analysis also indicates the benefit of such a course of five days with respect to the interaction and contact patterns in contrast to shorter introductory courses. Also, sensor-based and self-report complement each other with respect to various aspects; therefore, we recommend complementary analysis.

For future work, we aim at exploring more of the connections between the complementary data sources. Including online data from social networks and social media for provides an extended social context, e. g., [49], in order to investigate, for example, if the online interactions can be predicted given the offline (sensor) data and vice versa. This is especially relevant in the context of mobile and smart devices, for investigating the connection between the physical and the virtual world, and for comprehensive modeling. Also regarding cognitive processes, the analysis of perceived interactions can then also yield important insights into subjective motivation and interaction strategies.

Furthermore, we aim to investigate theory-based approaches in that context, formalizing models using basic principles from (social) theory and prior knowledge. For example, we plan to examine the paradigms of induced homophily, e. g., if people that interact much are likely to become more similar over time, which can be modeled as hypotheses [50] in order to detect anomalies, and for further investigating longitudinal processes. For that, subgroup discovery and exceptional model mining, e. g., [51], provide interesting approaches for future research, especially when combining compositional and structural analysis, i. e., on attributed graphs, [52], [53]. Here, also the integration of prior knowledge, i. e., semantic knowledge formalized in knowledge graphs [54] is a further interesting direction to consider in future work.

REFERENCES

[1] N. Eagle and A. S. Pentland, "Reality Mining: Sensing Complex Social Systems," *Pers. Ubiquit. Comput.*, vol. 10, no. 4, pp. 255–268, 2006.

[2] A. Barrat, C. Cattuto, V. Colizza, J.-F. Pinton, W. V. den Broeck, and A. Vespignani, "High Resolution Dynamical Mapping of Social Interactions with Active RFID," *PLoS ONE*, vol. 5, no. 7, 2008.

[3] M. Atzmueller and K. Hilgenberg, "Towards Capturing Social Interactions with SDCF: An Extensible Framework for Mobile Sensing and Ubiquitous Data Collection," in *Proc. MSM (HT 2013)*. ACM, 2013.

[4] M. Atzmueller, "Social Behavior in Mobile Social Networks: Characterizing Links, Roles and Communities," in *Mobile Social Networking*. Berlin: Springer, 2014, pp. 65–78.

[5] M. Atzmueller, L. Thiele, G. Stumme, and S. Kauffeld, "Analyzing Group Interaction and Dynamics on Socio-Behavioral Networks of Face-to-Face Proximity," in *Proc. UbiComp Adjunct*. ACM, 2016.

[6] D. J. Brass, "A Social Network Perspective on Industrial/Organizational Psychology," in *The Oxford Handbook of Industrial and Organizational Psychology*, S. W. J. Koslowski, Ed. New York: Oxford University Press, 2012.

[7] L. Thiele, M. Atzmueller, S. Kauffeld, and G. Stumme, "Subjective versus Objective Captured Social Networks: Comparing Standard Self-Report Questionnaire Data with Observational RFID Technology Data," in *Proc. Measuring Behavior*, Wageningen, The Netherlands, 2014.

[8] N. Eagle, A. Pentland, and D. Lazer, "From the Cover: Inferring Friendship Network Structure by using Mobile Phone Data," *PNAS*, vol. 106, pp. 15 274–15 278, 2009.

[9] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, "Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks," *PLoS ONE*, vol. 5, no. 7, 2010.

[10] M. Atzmueller, S. Doerfel, A. Hotho, F. Mitzlaff, and G. Stumme, "Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles," in *Modeling and Mining Ubiquitous Social Media*, ser. LNAI. Berlin: Springer, 2012, vol. 7472.

[11] B.-E. Macek, C. Scholz, M. Atzmueller, and G. Stumme, "Anatomy of a Conference," in *Proc. Hypertext*. ACM, 2012, pp. 245–254.

[12] J. Stehle, F. Charbonnier, T. Picard, C. Cattuto, and A. Barrat, "Gender Homophily from Spatial Behavior in a Primary School: A Sociometric Study," *Social Networks*, vol. 35, no. 4, pp. 604–613, 2013.

[13] M. C. Pachucki, E. J. Ozer, A. Barrat, and C. Cattuto, "Mental Health and Social Networks in Early Adolescence: A Dynamic Study of Objectively-Measured Social Interaction Behaviors," *Social Science & Medicine*, 2014.

[14] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. V. den Broeck, "What's in a Crowd? Analysis of Face-to-Face Behavioral Networks," *CoRR*, vol. 1006.1260, 2010.

[15] K. Sailer, R. Pachilova, and C. Brown, "Human versus Machine - Testing Validity and Insights of Manual and Automated Data Gathering Methods in Complex Buildings," in *Proc. 9th Intl. Space Syntax Symp.* Seoul, Korea: Sejong Univ. Press, 2013.

[16] C. Brown, C. Efstratiou, I. Leontiadis, D. Quercia, and C. Mascolo, "Tracking Serendipitous Interactions: How Individual Cultures Shape the Office," in *Proc. CSCW*. ACM, 2014, pp. 1072–1081.

[17] M. Atzmueller, D. Benz, S. Doerfel, A. Hotho, R. JÃschke, B. E. Macek, F. Mitzlaff, C. Scholz, and G. Stumme, "Enhancing Social Interactions at Conferences," *it - Information Technology*, vol. 53, no. 3, pp. 101–107, 2011.

[18] M. Atzmueller, M. Becker, M. Kibanov, C. Scholz, S. Doerfel, A. Hotho, B.-E. Macek, F. Mitzlaff, J. Mueller, and G. Stumme, "Ubicon and its Applications for Ubiquitous Social Computing," *New Review of Hypermedia and Multimedia*, vol. 20, no. 1, pp. 53–77, 2014.

[19] L. Wu, B. N. Waber, S. Aral, E. Brynjolfsson, and A. Pentland, "Mining Face-to-Face Interaction Networks using Sociometric Badges: Predicting Productivity in an IT Configuration Task," *SSRN 1130251*, 2008.

[20] J. C. Turner, "Towards a Cognitive Redefinition of the Social Group," *Cah Psychol Cogn*, vol. 1, no. 2, pp. 93–118, 1981.

[21] J. Coleman, *Foundations of Social Theory*. Cambridge, Mass.: Belknap Press of Harvard Univ. Press, 2000.

[22] M. Kibanov, M. Atzmueller, C. Scholz, and G. Stumme, "Temporal Evolution of Contacts and Communities in Networks of Face-to-Face Human Interactions," *Science China*, vol. 57, March 2014.

[23] C. Scholz, M. Atzmueller, and G. Stumme, "Predictability of Evolving Contacts and Triadic Closure in Human Face-to-Face Proximity Networks," *SNAM*, vol. 4, no. 217, 2014.

[24] M. Atzmueller, A. Ernst, F. Krebs, C. Scholz, and G. Stumme, "On the Evolution of Social Groups During Coffee Breaks," in *Proc. WWW 2014 (Companion)*. IW3C2 / ACM, 2014.

[25] M. Atzmueller, "Data Mining on Social Interaction Networks," *Journal of Data Mining and Digital Humanities*, vol. 1, June 2014.

[26] A. Barrat and C. Cattuto, *Temporal Networks*, ser. Understanding Complex Systems. Springer, 2013, ch. Temporal Networks of Face-to-Face Human Interactions.

[27] G. Palla, A.-L. Barabasi, and T. Vicsek, "Quantifying Social Group Evolution," *Nature*, vol. 446, no. 7136, pp. 664–667, April 2007.

[28] D. Liben-Nowell and J. M. Kleinberg, "The Link Prediction Problem for Social Networks," in *Proc. CIKM*. ACM, 2003, pp. 556–559.

[29] C. Scholz, M. Atzmueller, and G. Stumme, "On the Predictability of Recurring Links in Networks of Face-to-Face Proximity," in *Proc. WWW 2014 (Companion)*. IW3C2 / ACM, 2014.

[30] Christoph Scholz and Martin Atzmueller and Alain Barrat and Ciro Cattuto and Gerd Stumme, "New Insights and Methods For Predicting Face-To-Face Contacts," in *Proc. ICWSM*. AAAI, 2013.

[31] M. Atzmueller, "Analyzing and Grounding Social Interaction in Online and Offline Networks," in *Proc. ECML/PKDD*, ser. LNCS, vol. 8726. Berlin: Springer, 2014, pp. 485–488.

[32] R. Diestel, *Graph Theory*. Berlin: Springer, 2006.

[33] M. Gaertler, "Clustering," in *Network Analysis*, ser. LNCS, U. Brandes and T. Erlebach, Eds., vol. 3418. Springer, 2004, pp. 178–215.

[34] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, 1st ed. Cambridge University Press, 1994, no. 8.

[35] D. J. Watts and S. H. Strogatz, "Collective Dynamics of /'small-world/' Networks," *Nature*, vol. 393, no. 6684, pp. 440–442, june 1998.

[36] L. Sachs, *Applied Statistics*. Berlin: Springer, 1982.

[37] D. Krackhardt, "QAP Partialling as a Test of Spuriousness," *Social Networks*, vol. 9, pp. 171–186, 1987.

[38] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[39] P. Bearman and P. Parigi, "Cloning Headless Frogs and Other Important Matters: Conversation Topics and Network Structure," *Social Forces*, vol. 83, no. 2, pp. 535–557, 2004.

[40] C. Cattuto, Oral communication, December 1st, 2014.

[41] H. R. Bernard, P. Killworth, D. Kronenfeld, and L. Sailer, "The Problem of Informant Accuracy: The Validity of Retrospective Data," *Annual Review of Anthropology*, vol. 13, no. 1, pp. 495–517, 1984.

[42] L. C. Freeman, A. K. Romney, and S. C. Freeman, "Cognitive Structure and Informant Accuracy," *American anthropologist*, vol. 89, no. 2, pp. 310–325, 1987.

[43] P. V. Marsden, "Network Data and Measurement," *Annual Review of Sociology*, vol. 16, no. 1, pp. 435–463, 1990.

[44] W. D. Richards, "Data, Models, and Assumptions in Network Analysis," *Organizational communication: Traditional themes and new directions*, pp. 109–128, 1985.

[45] J. M. Vigil, "Asymmetries in the Friendship Preferences and Social Styles of Men and Women," *Human Nature*, vol. 18, no. 2, pp. 143–161, 2007.

[46] S. Maslov, K. Sneppen, and A. Zaliznyak, "Detection of Topological Patterns in Complex Networks: Correlation Profile of the Internet," *Phys. A: Statistical and Theoretical Physics*, vol. 333, pp. 529–540, 2004.

[47] R. Bakeman and V. Quera, *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge, UK: CUP, 2011.

[48] J. L. Fleiss, *Statistical Methods for Rates and Proportions*, 2nd ed. New York, NY, USA: Wiley, 1981.

[49] M. Atzmueller, "Mining Social Media," *WIREs: Data Mining and Knowledge Discovery*, vol. 2, pp. 411–419, 2012.

[50] M. Atzmueller, A. Schmidt, B. Kloepper, and D. Arnu, "HypGraphs: An Approach for Analysis and Assessment of Graph-Based and Sequential Hypotheses," in *New Frontiers in Mining Complex Patterns. Postproceedings NFMCP 2016*, ser. LNAI. Berlin: springer, 2017.

[51] M. Atzmueller, "Subgroup Discovery – Advanced Review," *WIREs: Data Mining and Knowledge Discovery*, vol. 5, no. 1, pp. 35–49, 2015.

[52] M. Atzmueller, S. Doerfel, and F. Mitzlaff, "Description-Oriented Community Detection using Exhaustive Subgroup Discovery," *Information Sciences*, vol. 329, pp. 965–984, 2016.

[53] M. Atzmueller, "Detecting Community Patterns Capturing Exceptional Link Trails," in *Proc. IEEE/ACM ASONAM*. IEEE Press, 2016.

[54] P. Ristoski and H. Paulheim, "Semantic Web in Data Mining and Knowledge Discovery: A Comprehensive Survey," *Web Semantics*, vol. 36, pp. 1–22, 2016.