Formation and Temporal Evolution of Social Groups During Coffee Breaks

Martin Atzmueller[1], Andreas Ernst[2], Friedrich Krebs[2], Christoph Scholz[3], and Gerd Stumme[1]

[1] University of Kassel, Research Center for Information System Design
Knowledge and Data Engineering Group
Wilhelmshöher Allee 73, 34121 Kassel, Germany
{atzmueller, stumme}@cs.uni-kassel.de

[2] University of Kassel, Center for Environmental Systems Research
Wilhelmshöher Allee 47, 34117 Kassel, Germany
{ernst,krebs}@usf.uni-kassel.de

[3] Fraunhofer Institute for Wind Energy and Energy System Technology
Department Energy Informatics and Information Systems
Königstor 59, 34119 Kassel, Germany
christoph.scholz@iwes.fraunhofer.de

Abstract. Group formation and evolution are prominent topics in social contexts. This paper focuses on the analysis of group evolution events in networks of face-to-face proximity. We first analyze statistical properties of group evolution, e.g., individual activity and typical group sizes. After that, we define a set of specific group evolution events. These are analyzed in the context of an academic conference, where we provide different patterns according to phases of the conference. Specifically, we investigate group formation and evolution using real-world data collected at the LWA 2010 conference utilizing the Conferator system, and discuss patterns according to different phases of the conference.

1 Introduction

An important goal of social sciences is to reach a theoretical understanding of the process of group formation and evolution of humans [24]. Typically, in such contexts the analysis is enabled using empirical studies of human behavior. However, until recently, such studies were very costly and time-consuming, especially for larger groups: Here, the individual behaviors of a larger group of people had to be observed – for a longer time period in a not too small area. Indeed, now – with the rise of social networking sites such as Second Life or Facebook – the situation in data collection has changed significantly. With such systems in place, the situation is quite different, as it has become much easier to track the individual behavior of users. While there have been results indicating that online connections relate to offline connections in specific contexts, e. g., [30], it has also been argued that the behavior within these online platforms differs in many cases from the offline behavior and its inherent structures. Strong ties, for example, seem to correlate better than weak ties [30], but e. g., also only a small share of friends in Facebook are really close connections, i. e., friends in the offline world [32, 53].

Preprint of:
Martin Atzmueller, Andreas Ernst, Friedrich Krebs, Christoph Scholz and Gerd Stumme (2015). Formation and Temporal Evoluation of Social Groups During Coffee In: Postproceedings of the International workshops MUSE & SenseML 2014, Nancy, France, and MSM 2014, Seoul, Korea. Revised selected papers. Springer, Heidelberg, Germany (In Press).

With the further development of sensor technology, however, it has become possible to track the behavior of individuals also in the offline world. Using suitable sensors, collecting data from large(r) groups has become possible. In our work, we will make use of RFID technology to track not only the location of individuals, but also to observe their communication behavior [16].

We utilize data of the Conferator [4] system [5] – a social conference guidance system for enhancing social interactions at conferences. Conferator applies active RFID proximity tags developed by the Sociopatterns collaboration.[5] In particular, these tags allow the collection of human face-to-face proximity. For our analysis, we utilize data that has been collected at the academic conference LWA 2010.[6] For the event, the participants of the conference were wearing the RFID tags for three days, at all times during the conference time.

Based on these data, we have performed an analysis of the formation and breakup of groups. Our contribution can be summarized as follows:

1. We provide a formal model of group evolution in networks of face-to-face proximity and present a definition of different group evolution events.
2. We then consider social behavior of individuals and specifically analyze the evolution of social groups:
   (a) We provide a statistical analysis of individual activity and typical group sizes during conference phases.
   (b) Second, we investigate the temporal evolution of the proposed group evolution events throughout the conference and especially during the coffee breaks. As a result, we observe and discuss typical communication and activity patterns during these social events.
3. We analyze these patterns and characteristics and discuss quite clear-cut differences between conference sessions, coffee breaks, poster sessions, and free time.

The rest of this paper is structured as follows: Section 2 discusses related work. Section 3 describes the RFID hardware setting and gives a detailed overview on the collected real-world datasets. Section 4 describes the formalization of social groups and group transitions. After that, Section 5 presents the analysis. Finally, Section 6 summarizes our results and discusses future work.

2 Related Work

In this section, we discuss related work concerning the analysis of human contact behavior and the analysis of groups. We start with a detailed overview on the analysis of human contact behavior.

[4] `http://www.conferator.org`
[5] `http://www.sociopatterns.org`
[6] `http://www.kde.cs.uni-kassel.de/conf/lwa10`

2.1 Human Contact Behavior

The analysis of human contact patterns and their underlying structure is an interesting and challenging task in social network analysis. Eagle and Pentland [26], for example, presented an analysis using proximity information collected by bluetooth devices as a proxy for human proximity. However, given the interaction range of bluetooth devices, the detected proximity does not necessarily correspond to face-to-face contacts [16], as also confirmed by Atzmueller and Hilgenberg [8]. 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 [23]. This infrastructure was also deployed in other environments in order to study human contacts, such as healthcare environments [28, 35], schools [51], offices [21] and museums [29]. Here, first analyses concerning group contact evolution have been reported in [15], focusing on the temporal evolution of smaller groups (up to size four).

Another approach for observing human face-to-face communication is the Sociometric Badge.[7] It records more details of the interaction but requires significantly larger devices. Besides these two approaches, there is, up to our knowledge, no single empirical study in the social sciences that resulted in a history of all conversations of some event, where, for each face-to-face conversation, the names of the respective dialogue partners are stored together with exact time stamps for start and end of the conversation.

The SocioPatterns framework also provides the technical basis of our Conferator [3–5] system. In this context, Atzmueller et al. [6] analyze the interactions and dynamics of the behavior of participants at conferences; similarly, the connection between research interests, roles and academic jobs of conference attendees is analyzed in [34]. Furthermore, the predictability of links in face-to-face contact networks and additional factors also including online networks have been analyzed by Scholz et al. [45, 46].

2.2 Analysis of Groups

Groups and their evolution are prominent topics in social sciences, e. g., [24, 33, 56]. Wasserman and Faust [57] discuss social network analysis in depth, and provide an overview on the analysis of cohesive subgroups in graphs, both outlining methods for structural analysis, e. g., [27, 41] as well as for obtaining compositional descriptions, e. g., [1, 2]. Social group evolution has been investigated in a community-based analysis [42] using bibliographic and call-detail records. Backstrom et al. [14] analyze group formation and evolution in large online social networks, focussing on membership, growth, and change of a group. Furthermore, Brodka et al. [19, 20, 44] investigate group formation and group evolution discovery and prediction in social networks.

In contrast to the approaches above, this paper focuses on networks of face-to-face proximity at academic conferences: We extend the definitions for group formation and evolution in a fine-grained analysis and investigate the impact of different phases at a conference. Furthermore, we do not necessarily focus on groups defined by a dense graph-structure, but analyze respective groups that are connected by face-to-face contacts. 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.

[7] `http://hd.media.mit.edu/badges`

3 Face-To-Face Contact Data

In this section, we summarize the framework used for collecting face-to-face contact networks, before we briefly describe the Conferator system.

3.1 RFID Setup

At LWA 2010 we asked participants to wear the active SocioPatterns RFID devices (see above), which can sense and log the close-range face-to-face proximity of individuals wearing them. Using the UBICON framework [3, 4] for data collection, this allows us to map out time-resolved networks of face-to-face contacts among the conference attendees. In the following, we refer to the applied active RFID tags as *proximity tags*.

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 RF shield at the carrier frequency used for communication [23].

For estimating a face-to-face contact, we apply a similar threshold-based approach as in [23]: 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. With respect to the accuracy of the applied RFID tags, we refer to the results of Cattuto et al. [23] 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 [22]. 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.

The proximity tags also send out tracking signals at different power levels, that are received by antennas of RFID readers installed at fixed positions in the conference environment. These tracking signals are used to relay proximity information to a central server and also to provide approximate (room-level) positioning of conference participants, cf. [47,48]. This allows us to monitor encounters, e.g., the number of times a pair of participants is assigned to the same set of nearest readers. All the packets emitted by a proximity tag contain a unique numeric identifier of the tag, as well the identifiers of the detected nearby devices. For more information about the proximity sensing technology, we refer the reader to the website of SocioPatterns.[8] For more details on the context of the LWA 2010 conference, we refer to [6] for an in-depth presentation.

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

3.2 Conferator platform

The proximity tags described above provide the physical infrastructure for our social conference management system Conferator. Conferator [3–5] is a social and ubiquitous conference guidance system, aiming at supporting conference participants during conference planning, attendance and their post-conference activities. Conferator features the ability to manage social and face-to-face contacts during the conference and to support social networking. Among other features, it provides an overview on the current social events and interactions, a map for locating conference participants, a personalized schedule, and adaptive recommendation mechanisms for interesting contacts and talks at the conference.

Conferator has successfully been deployed at several events, e.g., the LWA 2010,[9] LWA 2011[10] and LWA 2012[11] conferences, the Hypertext 2011[12] conference, the INFORMATIK 2013[13] conference, the UIS 2015 workshop[14] and a technology day of the Venus[15] project. In this paper, we focus on the data obtained at LWA 2010 in Kassel.

4 Formal Model

Before we analyze the evolution of groups in face-to-face contact networks, it is necessary to give a definition of a temporal social network and a social group.

4.1 Modeling Social Groups

Let $F = ([t_1, t_2), [t_2, t_3) \ldots, [t_m, t_{m+1}))$ be a list of consecutive time windows. In this paper, all windows will have a duration of one minute. Similar to [20] we define a *temporal social network* TSN as a list of single social networks $(SN_1, \cdots, SN_m)$.

$$SN_i = (V_i, E_i), i = 1, 2, \ldots, m,$$

where V_i is the set of all participants who had at least one face-to-face contact with some other participant within the time window $[t_i, t_{i+1})$. Two participants $u, v \in V_i$ are connected by an edge $e := (u, v)$ in E_i if they had at least one face-to-face contact within the time window $[t_i, t_{i+1})$.

We define a *social group* G in the social network $SN = (V, E)$ as a subset of vertices $G \subseteq V$ where G is a connected component of SN with $|G| > 1$. We denote the set of all social groups of SN by $\mathcal{G}$, and the set of all social groups of SN_i by $\mathcal{G}_i$.

9 `http://www.kde.cs.uni-kassel.de/conf/lwa10/`
10 `http://lwa2011.cs.uni-magdeburg.de/`
11 `http://lwa2012.cs.tu-dortmund.de/`
12 `http://www.ht2011.org/`
13 `http://informatik2013.de/`
14 `http://enviroinfo.eu/ak-uis/uis-2015`
15 `http://www.iteg.uni-kassel.de/`

4.2 Modeling Group Transitions

As in [20] we differentiate between the group transitions *form*, *merge*, *grow*, *continue*, *shrink*, *split*, and *dissolve* between two consecutive time windows $[t_i, t_{i+1})$ and $[t_{i+1}, t_{i+2})$. However, below we provide more formal and stricter definitions that allow us to classify evolution events without exceptions.

- We say that a group G *forms* in SN_{i+1}, iff $G \in \mathcal{G}_{i+1}$ and $\forall g_i \in G : \nexists G^{'} \in \mathcal{G}_i : g_i \in G^{'}$.
- We say that groups $G_1, \ldots, G_m$ in SN_i *merge*, iff $m \geq 2$ and $\exists G \in \mathcal{G}_{i+1}$ such that $\bigcup_{i=1}^{m} G_i \subseteq G$.

We say that group G in $\mathcal{G}_i$

- *grows*, iff $\exists! G^{'} \in \mathcal{G}_{i+1} : G \subset G^{'}$,
- *continues*, iff $\exists G^{'} \in \mathcal{G}_{i+1} : G^{'} = G$,
- *shrinks*, iff $\exists! G^{'} \in \mathcal{G}_{i+1} : G \supset G^{'}$,
- *splits*, iff $\exists G_1, \ldots, G_m \in \mathcal{G}_{i+1}$, with $m \geq 2$ such that $\bigcup_{i=1}^{m} G_i \subseteq G$,
- *dissolves*, iff $\forall g_i \in G : \nexists G^{'} \in \mathcal{G}_{i+1} : g_i \in G^{'}$.

5 Analysis

In this section, we first describe the applied dataset. After that, we provide statistical analysis results on individual activity, before we investigate group formation and evolution in detail and provide illustrating examples.

5.1 Dataset

Each link in the applied LWA 2010 network indicates physical face-to-face proximity and can be weighted by the cumulated duration of all face-to-face proximity contacts between the linked persons.

Table 1. General statistics for LWA 2010 dataset. In addition to the number of nodes and edges, here d is the diameter, AACD the average aggregated contact-duration (in seconds) and *APL* the average path length.

$\|V\|$	$\|E\|$	*Avg. Deg.*	*APL*	*d (G)*	*AACD*
77	1004	26.07	1.7	3	797

Table 1 provides a detailed overview on the dataset. As already observed in many other contexts [23,29,34] the distributions of all aggregated face-to-face contacts lengths between conference participants are heavy-tailed. More than the half of all cumulated face-to-face contacts are less than 200 seconds and the average contact duration is less than one minute, but very long contacts are also observed. Overall, the diameter, the average degree and the average path length of G are similar to the results presented in [6, 29].

Table 2 shows statistics on the individual group evolution events for different minimum group sizes for the LWA 2010 dataset.

Table 2. Statistics on the individual group evolution events for different minimum group sizes.

	≥ 2	≥ 3	≥ 4	≥ 8
Forming	941	98	16	1
Dissolving	936	96	16	1
Merging	140	140	140	50
Splitting	146	146	146	53
Growing	839	839	461	94
Shrinking	835	835	463	83
Continuing	3951	1103	406	33

5.2 Social Behavior of Individuals

This analysis draws on assessing the quantity and the quality of contacts during the course of a conference and the respective heterogeneity of individual conference participants. We consider three different temporal phases during the conference, i. e., coffee (and lunch) breaks, conference sessions, poster session, and free time (i. e., the remaining time besides breaks and sessions).

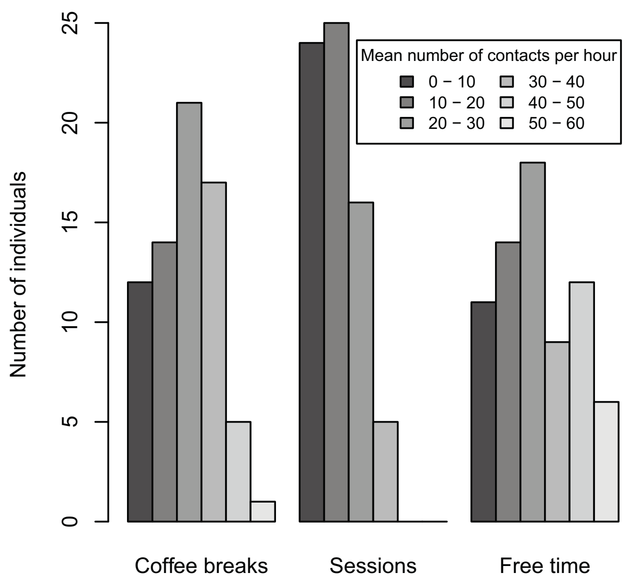


Fig. 1. Histograms of contact activities during conference phases. Except for the lowest category the cells denote right-closed and left open intervals.

The contact quantity provides an indicator of the networking activity of an individual while attending the conference. In a given phase of the conference, we measure contact activity by relating the number of minutes a participant attended to the number of minutes during which a contact with another participant was observed. The resulting indicator is quantified in terms of the mean number of contacts per hour of an individual participant during the respective conference phases. Figure 1 illustrates the results.

On average, individuals have 23 ($sd = 12.11$) contacts per hour during coffee breaks, 15 ($sd = 9.13$) during sessions, and 27 ($sd = 15.7$) during their free time. Differences in contacts per hour between conference phases are significant (repeated measures ANOVA with participants as within-factor and mean contacts per hour in different phases as dependent variables, $F(1.531, 105.634) = 32.216, p < .01$, Greenhouse-Geisser adjusted). Pairwise comparisons between phases using paired t tests show significant differences between session and coffee breaks ($T(74) = -6.64, p < .01$) and session and free time ($T(69) = -7.503, p < .01$). Differences between coffee breaks and free time were not significant ($T(69) = -2.009, p = .048$, adjusted alpha level = 0.017 (Bonferroni)). These overall and pairwise results were confirmed by the equivalent nonparametric test (Friedman test, $X^2(2) = 51.686, p < 0.01$).

Unsurprisingly, during coffee breaks or free times contact activity increases compared to session times. In both phases, a majority of the participants has more than 20 and up to 60 contacts per hour. In contrast, during session time the observed number of contacts decreases to 20 or less per hour for a big majority of the participants.

5.3 Evolution of Social Groups

In the following, we first investigate group statistics, focusing on group sizes during different conference phases. After that, we investigate group evolution events in detail.

Group Statistics. While the previous analysis focused merely on the quantity of contacts by an individual, the following investigation looks at a different property of the respective conversations. Thus, we determine the size of the conversation group an individual finds himself in during a given minute of the conference. Such conversation groups correspond to connected social network components as defined in Section 4.

Our assumption is that being member of a larger conversation group enables an individual on the one hand to spread his thoughts and ideas more widely and on the other hand allows him to perceive more diverse contributions from other individuals. Of course it has to be noted that face-to-face conversation or very small conversation groups can likewise yield high quality information exchange. However, in the context of this paper we use the size of the component an individual participant belongs to during a given phase of the conference as a proxy for the conversation quality of the respective individual. Figure 2 shows the respective results. On average, individuals find themselves in conversation groups of size 2.72 ($sd = 1.2$) during coffee breaks, 1.55 ($sd = .36$) during sessions, and 2.74 ($sd = 1.47$) during free time. The differences between conference phases are significant (repeated measures ANOVA with participants as within-factor and mean group size in different phases as dependent variables,

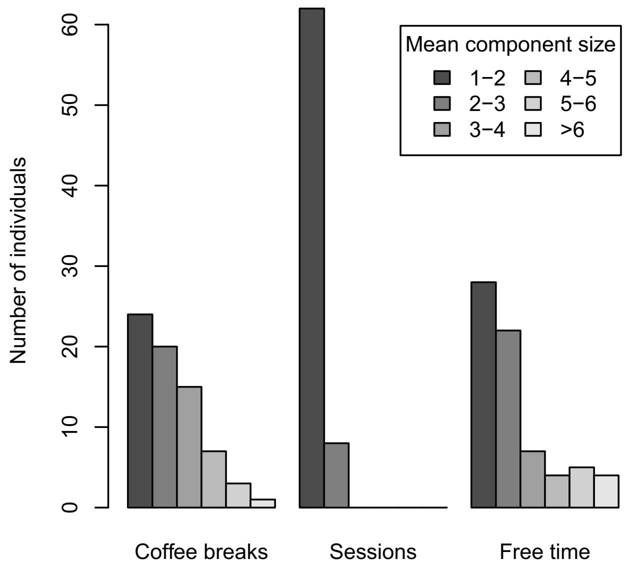
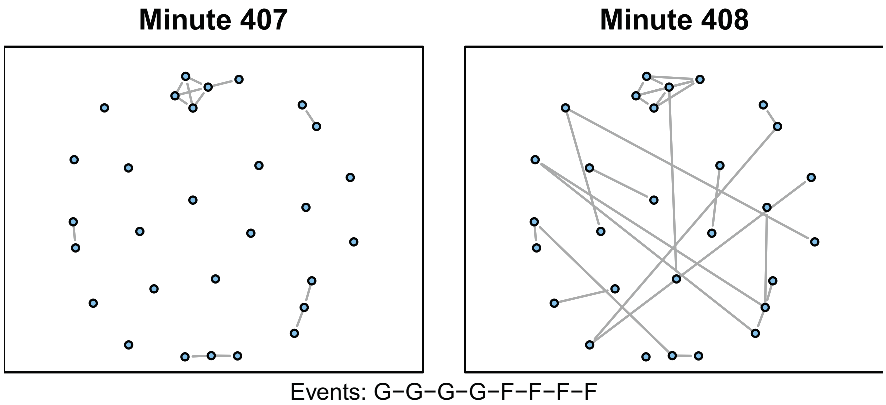


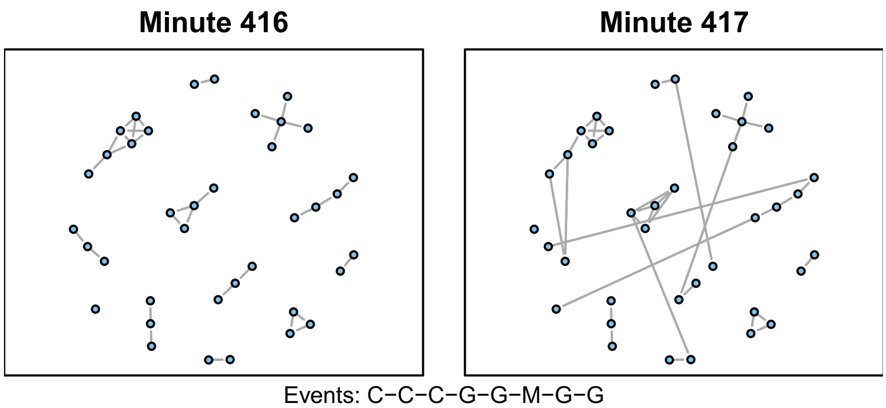
Fig. 2. Histograms of conversation group sizes during conference phases. Except for the lowest and highest categories histogram cells are right-closed and left open intervals. Note that component size 1 is included in the statistics to cover the case of solitary standing conference participants.

$F(1.61, 111.2) = 36.138, p < .01$, Greenhouse-Geisser adjusted). Pairwise comparisons between phases using paired t tests show significant differences between session time and coffee breaks ($T(74) = -8.81, p < .01$) and session time and free time ($T(69) = -7.43, p < .01$). Differences in group size between coffee breaks and free time were not significant ($T(69) = -0.88, p = .93$). These overall and pairwise results were confirmed by the nonparametric equivalent test (Friedman test: $X^2(2) = 65, p < 0.01$).

Clearly, during session times for the vast majority of individuals contacts are restricted to face-to-face (component size 2) or do not occur at all (component size 1). In sharp contrast, during coffee breaks or free times only one third of the participants remain in such small (conversation) groups while the others are found in larger groups up to size 6 and more. On the extreme end, around 10 participants are in average over all coffee breaks of the conference members of conversation groups of sizes exceeding 4. Similar circumstances are found during free time. Interestingly, despite significantly different activity patterns (see Figure 1 above) conversations groups tend to be smaller during free times compared to coffee breaks. However, this difference is statistically not significant.



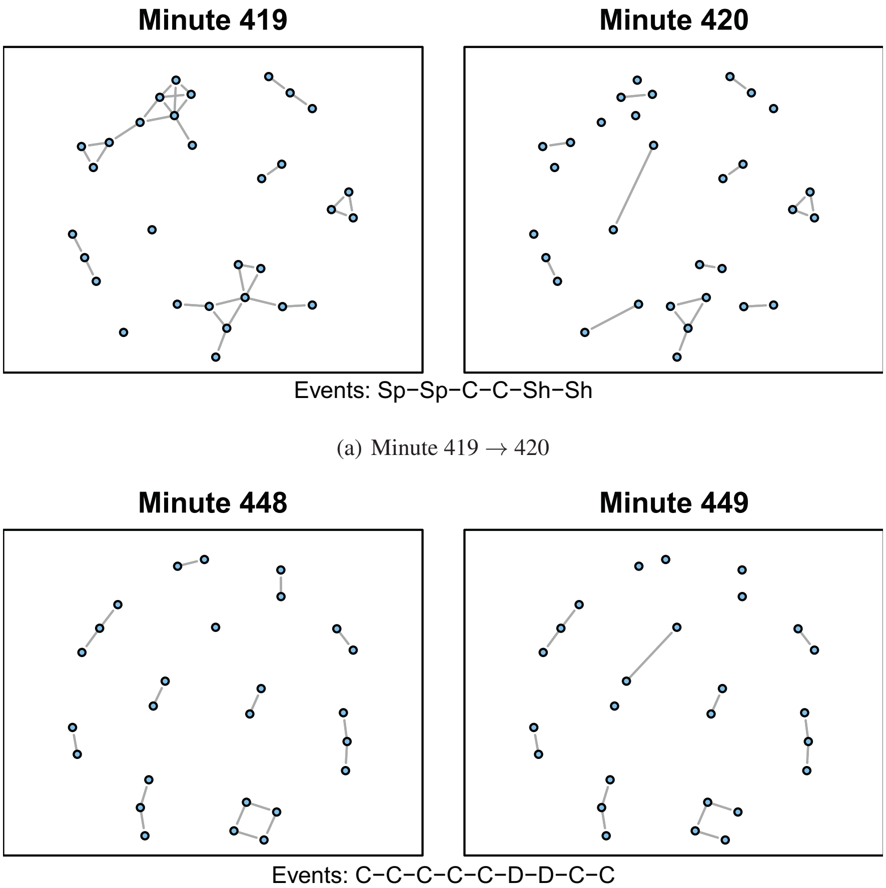
(a) Minute 407 → 408



(b) Minute 416 → 417

Fig. 3. Examples of group transitions at LWA 2010. The different transitions are depicted by the following annotations: C=Continuing, D=Dissolving, F=Forming, Sh=Shrinking, G=Growing, M=Merging Sp=Splitting

Group Transitions. In the following, we study the transition of the groups over time. Time is measured in minutes (excluding the nights). Minute 0 is 8:03 AM on Day 1 when the first signal of an RFID tag arrived, and Minute 2282 is the last signal recorded at 06:01 PM on Day 3. Day 1 ends in minute 740 with the last signal of the day on 08:23 PM; and Day 2 starts in Minute 741 at 08:14 AM with the first signal of the day. Day 2 ends in Minute 1714 with the last signal (concluding also the poster session) at 12:28 AM, and Day 3 starts in Minute 1715 at 08:34 AM. For detailed information, the conference schedule is available at `http://www.kde.cs.uni-kassel.de/conf/lwa10/program.html`.



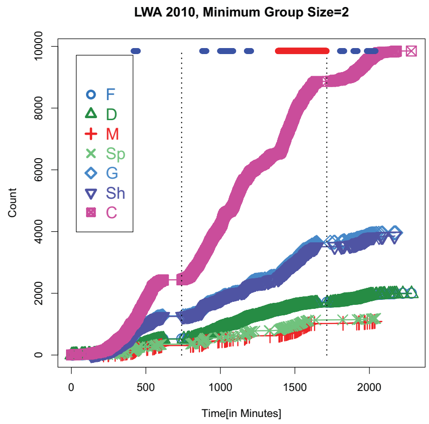
(a) Minute 419 → 420

(b) Minute 448 → 449

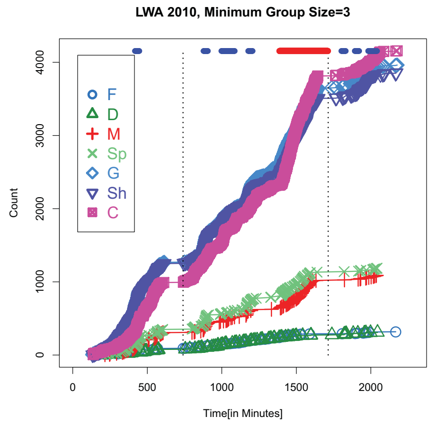
Fig. 4. Examples of group transitions at LWA 2010. The different transitions are depicted by the following annotations: C=Continuing, D=Dissolving, F=Forming, Sh=Shrinking, G=Growing, M=Merging Sp=Splitting

We start by illustrating some typical network configurations during the first coffee break of the conference (Minutes 416–446). In doing so, we will exemplify some of the typically occurring types of transitions. At the end of a session we expect conversation groups to build up while people leave the session rooms. The figures below show the contact networks during the final minutes of the session (minutes 407 and 408) and during the official beginning of the coffee break. In the footer line of the diagrams the group evolution events identified during the transition from t to t+1 are displayed.

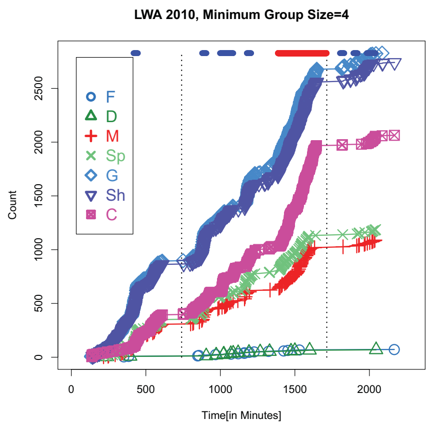
Between minute 407 and 408 a total of eight growing and forming events occur. People already leave the session rooms prior to the end of the session and start getting



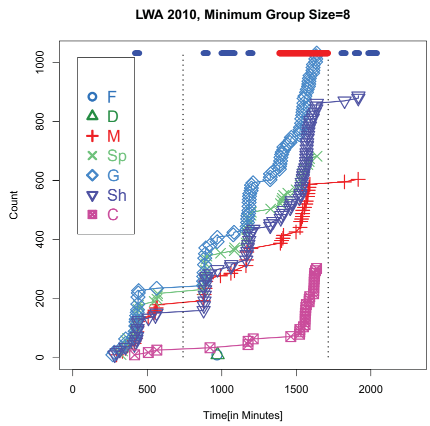
(a)



(b)



(c)



(d)

Fig. 5. Aggregated weighted occurrences of the transition types during the conference. C=Continuing, D=Dissolving, F=Forming, Sh=Shrinking, G=Growing, M=Merging Sp=Splitting. At the top of the figure, we mark the different coffee breaks (shown in blue); the red bar on Day 2 indicates the poster session.

in contact. Consistently, the diagram for minute 416 illustrates that at the beginning of the break numerous groups of different sizes are established. Towards minute 417 these groups either persist, or they grow or merge respectively. Compared to the other minutes of the coffee break during these two time spans the maximum frequency of *growing* events is found. Likewise for the first time span the maximum number of forming events during the coffee break is observed.

The circumstances during the end of the coffee break and beginning of the following sessions are well illustrated by the characteristics of the transition from minute 419 to 420 and 448 to 449, see Figures 5(a)-5(d): The first diagram shows a case of splitting and shrinking of larger groups. The second diagram illustrates that once conversation groups have shrunk most of the remaining small groups persist and only two groups dissolve. This situation marks the maximum number of *continuing* events found during the course of the regarded coffee break. The time span from minute 448 to 449 exhibits the maximum number of *splitting* events found for the considered coffee break.

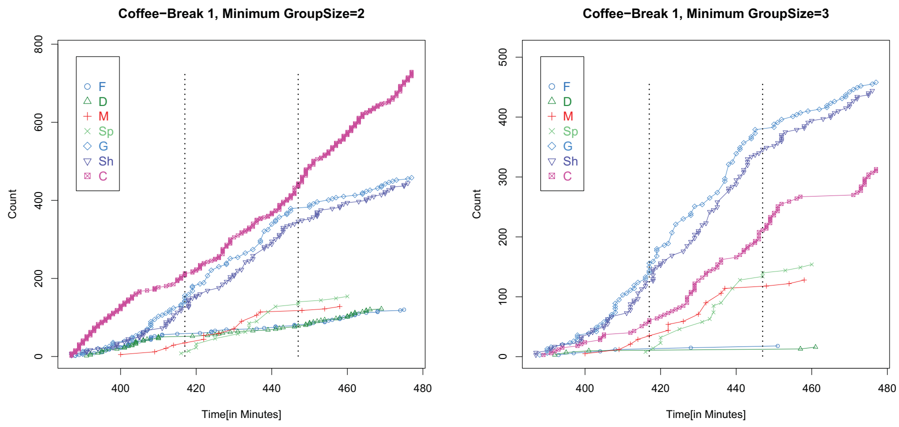
After these illustrating examples, we turn to a quantitative analysis of the group transitions:

- For our study we used different minimum group sizes. A minimum group size of $n \in \mathbb{N}$ means that we consider all groups with size greater or equal n.
- In Figure 5, we plotted, for each transition type, the weighted sum of all its transitions between minute 0 and t.
- For each transition of one of the types *continuing*, *dissolving*, *splitting* and *shrinking*, we add $|G_t|$ to the sum. For each transition of one of the types *forming*, *growing* and *merging*, we add $|G_{t+1}|$ to the sum.
- At the top of the figure, we mark the different coffee breaks (shown in blue); the red bar on Day 2 indicates the poster session.

We observe that for a minimum group size of 2 the number of *continuings* is the most dominating value. The number of *continuings* decreases rapidly when we consider groups with size greater than 3 only. This means that the *continuing*-event mostly appears in groups of size 2 or 3. In addition, we note that the group transition types *forming* and *dissolving* are observed mostly for groups of size 2. To our surprise it is very unlikely that a group of size greater than 3 will *form* or *dissolve*. Considering groups of size greater than 3 the group transitions *growing* and *shrinking* become the most dominating events. For larger groups, we observe a strong increase of *continuings* during the conference poster session.

It is interesting to see that the inverse transitions (i.e. *growing* vs. *shrinking*, *forming* vs. *dissolving* and *merging* vs. *splitting*) have almost identical curves. This is a first indicator for the hypothesis that growth and decay of communication groups are symmetric. As expected, they differ during communicative phases (coffee breaks etc.) such that the weighted sum of the increasing transition type grows earlier during this phase, while the sum of the corresponding decreasing type grows more at the end of the phase.

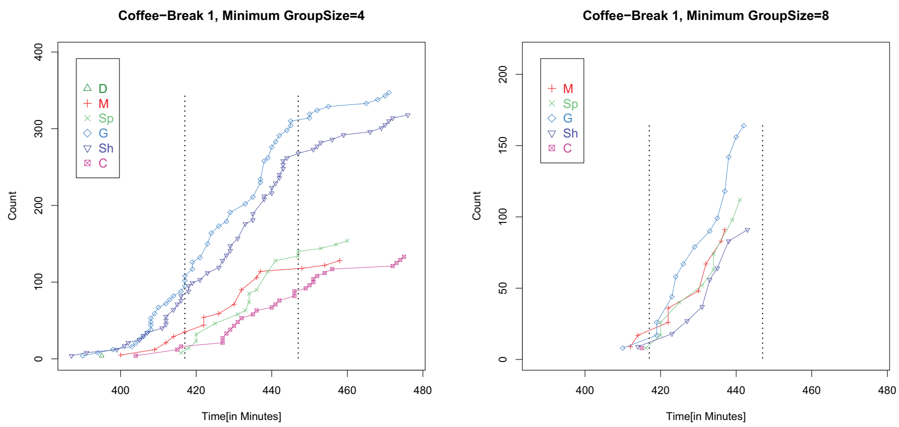
For some further illustrating examples, Figures 6-7 show a close-up of the global curves around the first coffee break, which started in Minute 416 and ended in Minute 446, including thirty minutes prior and after the break. Also, while the results of Figures 6 and 7 are quite similar to those of Figure 5, we also observe the clear trend that the most activity takes place during the coffee breaks. For example, for a minimum group size of 8 the coffee break can be detected very well (see Figure 7(b)): Here all the group transitions take place during the coffee break. This observation does also hold for all other coffee breaks.



(a)

(b)

Fig. 6. Close-up of the curves in Figure 5 around coffee break 1 for minimal groups sizes $GroupSize = 2, 3$. For better readability, all curves start at level 0 at the left end of the diagram. The different transition types are depicted by the following annotations: C=Continuing, D=Dissolving, F=Forming, Sh=Shrinking, G=Growing, M=Merging Sp=Splitting



(a)

(b)

Fig. 7. Close-up of the curves in Figure 5 around coffee break 1 for minimal groups sizes $GroupSize = 4, 8$. For better readability, all curves start at level 0 at the left end of the diagram. The different transition types are depicted by the following annotations: C=Continuing, D=Dissolving, F=Forming, Sh=Shrinking, G=Growing, M=Merging Sp=Splitting

6 Conclusions

We have used RFID technology to investigate the structure and dynamics of real-life face-to-face social contacts. We presented a formal model of detecting group dynamics in the data providing strict definitions that allow us to classify evolution events without exceptions. As an example, we took the interactions of participants of one conference and analyzed their individual activities, as well as the characteristic and quite clear-cut differences between conference sessions, coffee breaks, poster sessions, and free time. While the data have great face validity, it will certainly be useful to validate the data provided by the RFID technology with experimental means in future research to know more about possible technical artefacts.

Furthermore, we also aim to investigate the generality of the observed phenomena by extending the analysis focusing on a set of conferences, e. g., [34, 49]. Then, also subgroup and community detection methods aiming to describe such groups can provide further insights and data-driven explanations, e. g., [7, 9, 12]. Further fundamental issues concern the analysis of dynamics of groups and their evolution [25, 31]. Also, analytical methods can then potentially be used for grounding the evaluation of such structures, e. g., [36, 38]. We aim to investigate such approaches in more detail also concerning multi-layer networks, e. g., [54]. In addition, we will investigate how to embed findings on structure and dynamics into predictive methods: This includes, e. g., link prediction in such contexts [46,49,52], community analytics [7,55], and according pattern detection and modeling methods, e. g., [10, 11, 43]. Of course, this also extends to the semantics of user interactions [17, 39, 40, 50], their evaluation [37] and their explanation, e. g., [13, 18].

Also, at the moment, we have focussed on macro phenomena like the overall group dynamics. But the technology we use also allows for combining off-line data about individuals (like e.g. their academic role of their scientific interests) with their communication behavior at meetings. The individual history of encounters and personal acquaintances certainly plays a further role. Moreover, architectural and constructional properties of the venue can influence the formation of groups, e.g. the localization of the buffet of the conference dinner, and so forth. Further directions here also include location based group and mobility patterns. By combining such additional knowledge with the observed real-time dynamics, we might get closer to a theory of real world face-to-face group dynamics. Such dynamics, in turn, might be taken as a proxy for the spread of information between people, or for in-depth discussions – depending on the kind of groups we observe.

Acknowledgements

This work has been supported by the VENUS research cluster at the Research Center for Information System Design (ITeG) at Kassel University. We thank SocioPatterns for providing privileged access to the SocioPatterns sensing platform used for collecting the contact data.

References

1. M. Atzmueller. *Knowledge-Intensive Subgroup Mining – Techniques for Automatic and Interactive Discovery*, volume 307 of *DISKI*. IOS Press, March 2007.
2. M. Atzmueller. Subgroup Discovery – Advanced Review. *WIREs: Data Mining and Knowledge Discovery*, 5(1):35–49, 2015.
3. M. Atzmueller, M. Becker, S. Doerfel, M. Kibanov, A. Hotho, B.-E. Macek, F. Mitzlaff, J. Mueller, C. Scholz, and G. Stumme. Ubicon: Observing social and physical activities. In *Proc. 4th IEEE Intl. Conf. on Cyber, Physical and Social Computing (CPSCom 2012).*, 2012.
4. 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*, 20(1):53–77, 2014.
5. M. Atzmueller, D. Benz, S. Doerfel, A. Hotho, R. Jäschke, B. E. Macek, F. Mitzlaff, C. Scholz, and G. Stumme. Enhancing Social Interactions at Conferences. *it+ti*, 53(3):101–107, 2011.
6. 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*, volume 7472 of *LNAI*. Springer, Heidelberg, Germany, 2012.
7. M. Atzmueller, S. Doerfel, and F. Mitzlaff. Description-Oriented Community Detection using Exhaustive Subgroup Discovery. *Information Sciences*, 2015.
8. M. Atzmueller and K. Hilgenberg. Towards Capturing Social Interactions with SDCF: An Extensible Framework for Mobile Sensing and Ubiquitous Data Collection. In *Proc. 4th International Workshop on Modeling Social Media (MSM 2013), Hypertext 2013*, New York, NY, USA, 2013. ACM Press.
9. M. Atzmueller and F. Lemmerich. VIKAMINE - Open-Source Subgroup Discovery, Pattern Mining, and Analytics. In *Proc. ECML/PKDD*, Heidelberg, Germany, 2012. Springer.
10. M. Atzmueller and F. Lemmerich. Exploratory Pattern Mining on Social Media using Geo-References and Social Tagging Information. *IJWS*, 2(1/2), 2013.
11. M. Atzmueller, F. Lemmerich, B. Krause, and A. Hotho. Who are the Spammers? Understandable Local Patterns for Concept Description. In *Proc. 7th Conference on Computer Methods and Systems*, Krakow, Poland, 2009. Oprogramowanie Nauko-Techniczne.
12. M. Atzmueller and F. Puppe. A Case-Based Approach for Characterization and Analysis of Subgroup Patterns. *Journal of Applied Intelligence*, 28(3):210–221, 2008.
13. M. Atzmueller and T. Roth-Berghofer. The Mining and Analysis Continuum of Explaining Uncovered. In *Proc. 30th SGAI International Conference on Artificial Intelligence (AI-2010)*, 2010.
14. L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group Formation in Large Social Networks: Membership, Growth, and Evolution. In *Proc. KDD*, pages 44–54, New York, NY, USA, 2006. ACM.
15. A. Barrat and C. Cattuto. *Temporal Networks*, chapter Temporal Networks of Face-to-Face Human Interactions. Understanding Complex Systems. Springer, Heidelberg, Germany, 2013.
16. 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. *CoRR*, abs/0811.4170, 2008.
17. U. Bojars, J. G. Breslin, V. Peristeras, G. Tummarello, and S. Decker. Interlinking the social web with semantics. *Intelligent Systems, IEEE*, 23(3):29–40, 2008.
18. S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca. Network analysis in the social sciences. *science*, 323(5916):892–895, 2009.

19. P. Bródka, P. Kazienko, and B. Kołoszczyk. Predicting group evolution in the social network. In K. Aberer, A. Flache, W. Jager, L. Liu, J. Tang, and C. Guéret, editors, *Social Informatics*, volume 7710 of *Lecture Notes in Computer Science*, pages 54–67. Springer, 2012.
20. P. Bródka, S. Saganowski, and P. Kazienko. GED: The Method for Group Evolution Discovery in Social Networks. *SNAM*, 3(1):1–14, 2011.
21. C. Brown, C. Efstratiou, I. Leontiadis, D. Quercia, and C. Mascolo. Tracking Serendipitous Interactions: How Individual Cultures Shape the Office. In *Proc. CSCW*, pages 1072–1081, New York, NY, USA, 2014. ACM.
22. C. Cattuto. Oral communication, December 1st, 2014.
23. 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*, 5(7):e11596, 07 2010.
24. J. Coleman. *Foundations of Social Theory*. Belknap Press of Harvard Univ. Press, Cambridge, Mass., 2000.
25. G. Diakidis, D. Karna, D. Fasarakis-Hilliard, D. Vogiatzis, and G. Paliouras. Predicting the evolution of communities in social networks. In *Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*, WIMS '15, pages 1:1–1:6, New York, NY, USA, 2015. ACM.
26. N. Eagle, A. Pentland, and D. Lazer. From the Cover: Inferring Friendship Network Structure by using Mobile Phone Data. *PNAS*, 106:15274–15278, 2009.
27. S. Fortunato and C. Castellano. *Encyclopedia of Complexity and System Science*, chapter Community Structure in Graphs. Springer, 2007.
28. L. Isella, M. Romano, A. Barrat, C. Cattuto, V. Colizza, W. Van den Broeck, F. Gesualdo, E. Pandolfi, L. Ravà, C. Rizzo, and A. Tozzi. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS ONE*, 6:e17144, 2011.
29. L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. V. D. Broeck. What's in a Crowd? Analysis of Face-to-Face Behavioral Networks. *Journal of Theoretical Biology*, 271:166–180, 2011.
30. M. Kibanov, M. Atzmueller, J. Illig, C. Scholz, A. Barrat, C. Cattuto, and G. Stumme. Is Web Content a Good Proxy for Real-Life Interaction? A Case Study Considering Online and Offline Interactions of Computer Scientists. In *Proc. ASONAM*, Boston, MA, USA, 2015. IEEE Press.
31. 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*, 57, 2014.
32. J. Kim and J.-E. R. Lee. The Facebook Paths to Happiness: Effects of the Number of Facebook Friends and Self-Presentation on Subjective Well-Being. *Cyberpsychol Behav Soc Netw*, 14(6):359 – 364, 2011.
33. R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In P. S. Yu, J. Han, and C. Faloutsos, editors, *Link Mining: Models, Algorithms, and Applications*, pages 337–357. Springer, Heidelberg, Germany, 2010.
34. B.-E. Macek, C. Scholz, M. Atzmueller, and G. Stumme. Anatomy of a Conference. In *Proc. ACM Hypertext*, pages 245–254, New York, NY, USA, 2012.
35. A. Machens, F. Gesualdo, C. Rizzo, A. E. Tozzi, A. Barrat, and C. Cattuto. An Infectious Disease Model on Empirical Networks of Human Contact: Bridging the Gap between Dynamic Network Data and Contact Matrices. *BMC Infectious Diseases*, 13(185), 2013.
36. F. Mitzlaff, M. Atzmueller, D. Benz, A. Hotho, and G. Stumme. Community Assessment using Evidence Networks. In *Proceedings of the International Workshop on Mining Ubiquitous and Social Environments (MUSE2010)*, Barcelona, Spain, 2010.

37. F. Mitzlaff, M. Atzmueller, D. Benz, A. Hotho, and G. Stumme. Community Assessment using Evidence Networks. In *Analysis of Social Media and Ubiquitous Data*, volume 6904 of *LNAI*, 2011.
38. F. Mitzlaff, M. Atzmueller, D. Benz, A. Hotho, and G. Stumme. User-Relatedness and Community Structure in Social Interaction Networks. *CoRR/abs*, 1309.3888, 2013.
39. F. Mitzlaff, M. Atzmueller, A. Hotho, and G. Stumme. The Social Distributional Hypothesis. *Journal of Social Network Analysis and Mining*, 4(216), 2014.
40. F. Mitzlaff, M. Atzmueller, G. Stumme, and A. Hotho. Semantics of User Interaction in Social Media. In *Complex Networks IV*, volume 476 of *Studies in Computational Intelligence*. Springer, Heidelberg, Germany, 2013.
41. M. E. J. Newman. Detecting Community Structure in Networks. *Europ Physical J*, 38, 2004.
42. G. Palla, A.-L. Barabasi, and T. Vicsek. Quantifying Social Group Evolution. *Nature*, 446(7136):664–667, April 2007.
43. D. Parra, C. Trattner, D. Gómez, M. Hurtado, X. Wen, and Y.-R. Lin. Twitter in academic events: A study of temporal usage, communication, sentimental and topical patterns in 16 computer science conferences. *Computer Communications*, 2015.
44. S. Saganowski, B. Gliwa, P. Bródka, A. Zygmunt, P. Kazienko, and J. Kozlak. Predicting community evolution in social networks. *Entropy*, 17, 2015.
45. C. Scholz, M. Atzmueller, A. Barrat, C. Cattuto, and G. Stumme. New Insights and Methods For Predicting Face-To-Face Contacts. In *Proc. 7th Intl. AAAI Conference on Weblogs and Social Media*, Palo Alto, CA, USA, 2013. AAAI Press.
46. C. Scholz, M. Atzmueller, and G. Stumme. On the Predictability of Human Contacts: Influence Factors and the Strength of Stronger Ties. In *Proc. Fourth ASE/IEEE Intl. Conf. on Social Computing*, Boston, MA, USA, 2012. IEEE Computer Society.
47. C. Scholz, M. Atzmueller, and G. Stumme. Unsupervised and Hybrid Approaches for On-Line RFID Localization with Mixed Context Knowledge. In *Proc. 21st Intl. Symposium on Methodologies for Intelligent Systems*, Heidelberg, Germany, 2014. Springer.
48. C. Scholz, S. Doerfel, M. Atzmueller, A. Hotho, and G. Stumme. Resource-Aware On-Line RFID Localization Using Proximity Data. In *Proc. ECML/PKDD*, pages 129–144, 2011.
49. C. Scholz, J. Illig, M. Atzmueller, and G. Stumme. On the Predictability of Talk Attendance at Academic Conferences. In *Proc. 25th ACM Conference on Hypertext and Social Media*, New York, NY, USA, 2014. ACM Press.
50. S. Staab. Emergent semantics. *IEEE Intelligent Systems*, (1):78–86, 2002.
51. J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina, and P. Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE*, 6(8):e23176, 08 2011.
52. M. Steurer and C. Trattner. Predicting interactions in online social networks: An experiment in second life. In *Proceedings of the 4th International Workshop on Modeling Social Media*, MSM '13, pages 5:1–5:8, New York, NY, USA, 2013. ACM.
53. K. Subrahmanyam, S. M. Reich, N. Waechter, and G. Espinoza. Online and Offline Social Networks: Use of Social Networking Sites by Emerging Adults. *J Appl Dev Psychol*, 29(6):420 – 433, 2008.
54. L. Tang, H. Liu, J. Zhang, and Z. Nazeri. Community evolution in dynamic multi-mode networks. In *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 677–685, New York, NY, USA, 2008. ACM.
55. L. Tang, X. Wang, and H. Liu. Community detection via heterogeneous interaction analysis. *Data Min. Knowl. Discov.*, 25(1):1–33, July 2012.
56. J. C. Turner. Towards a Cognitive Redefinition of the Social Group. *Cah Psychol Cogn*, 1(2):93–118, 1981.
57. S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Number 8 in Structural analysis in the social sciences. Cambridge University Press, 1 edition, 1994.