On the Evolution of Social Groups During Coffee Breaks

Martin Atzmueller
Knowledge and Data Engineering Group (KDE)
University of Kassel, Germany
atzmueller@cs.uni-kassel.de

Andreas Ernst
Center for Environmental Systems Research
University of Kassel, Germany
ernst@usf.uni-kassel.de

Friedrich Krebs
Center for Environmental Systems Research
University of Kassel, Germany
krebs@usf.uni-kassel.de

Christoph Scholz
Knowledge and Data Engineering Group
University of Kassel, Germany
scholz@cs.uni-kassel.de

Gerd Stumme
Knowledge and Data Engineering Group
University of Kassel, Germany
stumme@cs.uni-kassel.de

ABSTRACT

This paper focuses on the analysis of group evolution events in networks of face-to-face proximity. First, we analyze statistical properties of group evolution, e.g., individual activity and typical group sizes. Furthermore, we define a set of specific group evolution events. We analyze these using real-world data collected at the LWA 2010 conference using 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 of humans [8]. Until recently, empirical studies to this end were very costly and time-consuming, since the individual behaviors of a larger group of people had to be observed – for a longer time period in a not too small area. Now, the rise of social networking sites such as Second Life or Facebook significantly altered the situation, as it has become much easier to track the individual behavior of users. However, it has been argued that the behavior within these online platforms differs significantly from the offline behavior, e. g., that only a small share of friends in Facebook are really close connections, i. e., friends in the offline world [11, 14].

With the further development of sensor technology, however, it has become possible to track the behavior of individuals also in the offline world. 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 [5]. We utilize data of the Conferator [1] system [1] – a social conference guidance system for enhancing social interactions at conferences. Conferator applies active RFID proximity tags developed by the Sociopatterns collaboration.[2] In particular, these tags allow the collection of human face-to-face proximity. For our analysis, we utilize data that has been collected at an academic conference, i. e., LWA 2010.[3] For the event, participants of LWA 2010 were wearing the RFID tags for three days, at all times during the conference.

Based on these data, we have performed an analysis of the formation and breakup of groups. Our contribution can be summarized as follows: We provide a formal model of group evolution in networks of face-to-face proximity and present a definition of different group evolution events. We then consider social behavior of individuals and specifically analyze the evolution of social groups: First, we provide a statistical analysis of individual activity and typical group sizes during conference phases. 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.

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

The analysis of human contact patterns and their underlying structure is an interesting and challenging task in social network analysis. Eagle and Pentland [9], for example, presented an analysis using proximity information collected by bluetooth devices as a proxy for human proximity. However, given the range of interaction of bluetooth devices, the detected proximity does not necessarily correspond to face-to-face contacts [5].

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

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

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

[3] http://www.kde.cs.uni-kassel.de/conf/lwa10

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2579235.

temporal resolution of 20 seconds [7]. First analyses concerning group contact evolution have been reported in [4], where the temporal evolution of smaller group sizes (up to size four) were statistically analyzed. Another approach for observing human face-to-face communication is the Sociometric Badge.[4] It records more details of the interaction, but requires significantly larger devices.

The SocioPatterns framework provides also the technical basis of our Conferator system. In this context, Atzmueller et al. [2] 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 further analyzed in [12].

Groups and their evolution are prominent topics in social sciences, e. g., [8, 15]. Social group evolution has been investigated in a community-based analysis [13] using bibliographic and call-detail records. Backstrom et al. [3] analyze group formation and evolution in large online social networks, focussing on membership, growth, and change of a group. Furthermore, Brodka et al. [6] investigate group formation and group evolution discovery in social networks.

In contrast to the approaches summarized 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 (communities) defined by a dense graph-structure, but analyze subgroups given by 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.

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 RFID devices described above, which can sense and log the close-range face-to-face proximity of individuals wearing them. This allows us to map out time-resolved networks of face-to-face contacts among the conference attendees. In the following, we will refer to these 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 [7]. As in [7], 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. Due to limited space, we refer to [2] for more details on the LWA 2010 context.

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. 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.

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

3.2 Conferator platform

The proximity tags provide the physical infrastructure for our social conference management system Conferator. Conferator [1] is a social and ubiquitous conference guidance system, aiming at supporting conference participants during conference planning, attendance and their post-conference activities. It 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 recommendations for interesting contacts and talks.

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.

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 [6] we define a *temporal social network* TSN as a list of single social networks

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

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 $\mathcal{G}_i$.

As in [6] we differentiate between the following group transitions between two consecutive time windows $[t_i, t_{i+1})$ and $[t_{i+1}, t_{i+2})$, but provide more formal and stricter definitions that allow us to classify evolution events without exceptions:

- We say that 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 $\forall g_i \in G : \nexists G^{'} \in \mathcal{G}_{i+1} : g_i \in G^{'}$.

Table 1: 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

Table 2: General statistics for LWA 2010 dataset. 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

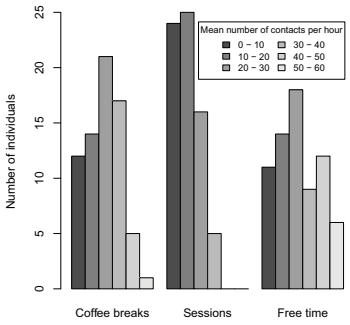


Figure 1: Histograms of contact activities during conference phases. Except for the lowest category the cells denote right-closed and left open intervals.

5. ANALYSIS

In this section, we first provide statistical analysis results on individual activity, before we investigate group formation and evolution in detail.

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 shows statistics on the individual group evolution events for different minimum group sizes for the LWA 2010 dataset. Table 2 provides a detailed overview on the dataset. As already observed in many other contexts [7, 10, 12] the distributions of all aggregated face-to-face contacts lengths between conference participants are heavy-tailed. The diameter, average degree and average path length of G are similar to the results presented in [2,10].

5.1 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).

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.

5.2 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.

5.2.1 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. 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, $F(1.61, 111.2) = 36.138, p < .01$, 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: $X^2(2) = 65, p < 0.01$).

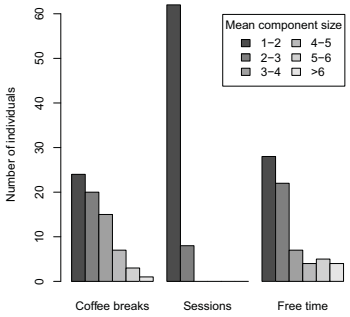


Figure 2: Histograms of conversation group sizes during conference phases. Except for the lowest and highest categories histogram cells are right-closed and left open intervals. Note that component size 1 is included in the statistics to cover the case of solitary standing conference participants.

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.

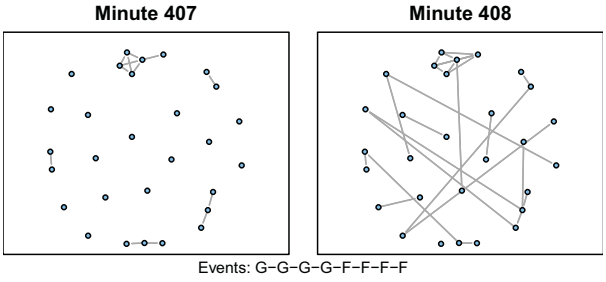
5.2.2 *Group Transitions.*

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`.

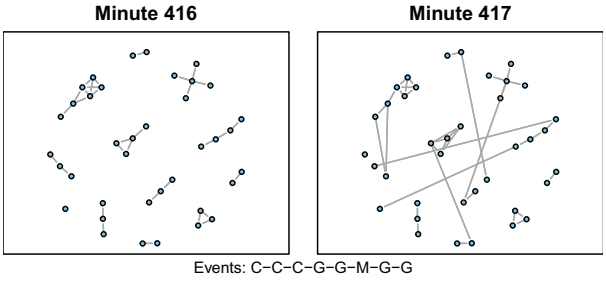
We start by illustrating some typical network configurations during the first coffee break of the conference (Minutes 416–446). In doing so, we will exemplify some of the typically occurring types of transitions.

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.

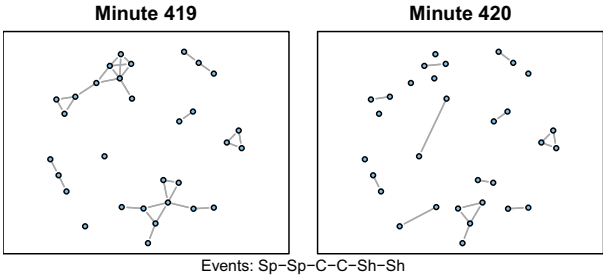
Between minute 407 and 408 a total of eight growing and forming events occur. People already leave the session rooms prior to the end of the session and start getting in contact.



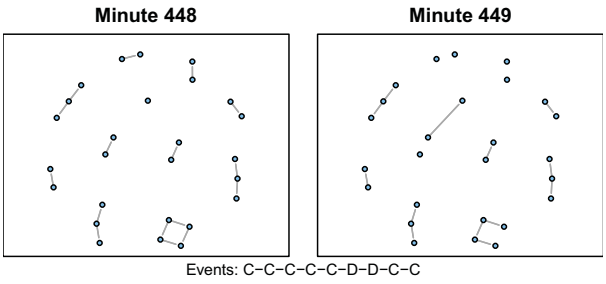
(a) Minute 407 → 408



(b) Minute 416 → 417



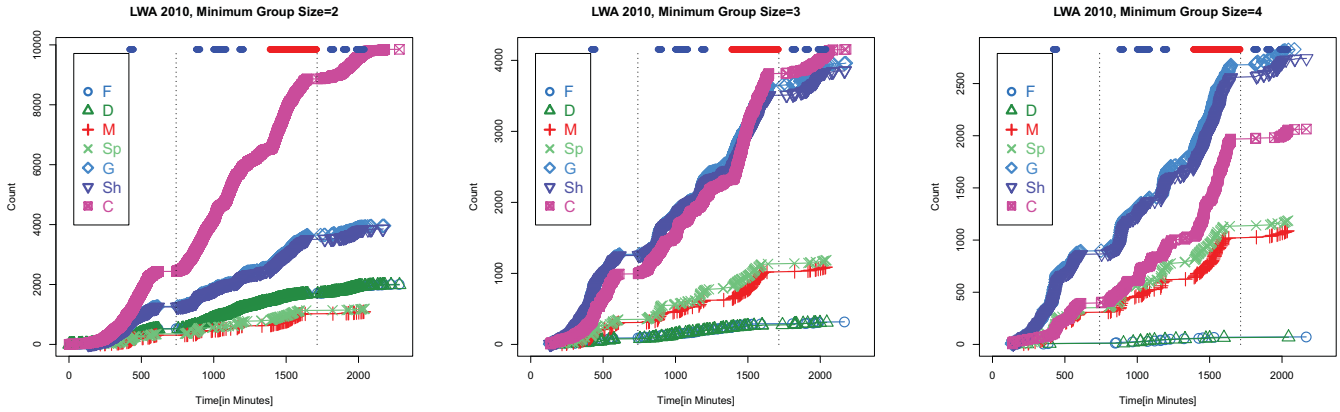
(c) Minute 419 → 420



(d) Minute 448 → 449

Figure 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

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



(a) (b) (c)

Figure 4: 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.

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 below): 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 4, 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.

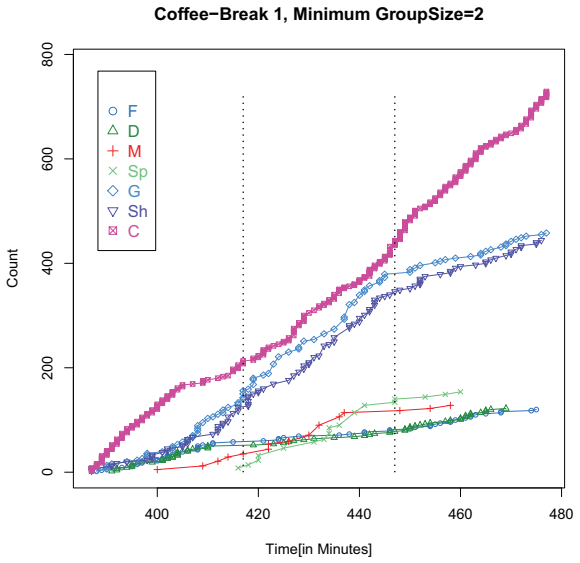
We observe that for a minimum group size of 2 the number of *continuings* is the most dominating value. The number of *continuings* decreases rapidly when we consider groups with size greater than 3 only. This means that the *continuing*-event mostly appears in groups of size 2 or 3. In addition, we 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. As an example, Figure 5, shows 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. While the results of Figure 5 are quite similar to those of Figure 4, 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 5(b)): Here all the group transitions take place during the coffee break. This observation does also hold for all other coffee breaks.

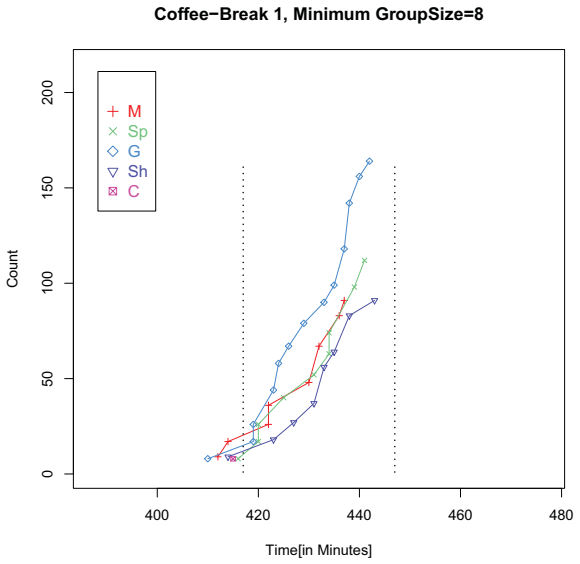
6. CONCLUSIONS

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. 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 analyzing a set of conferences.

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. Also, the



(a)



(b)

Figure 5: Close-up of the curves in Figure 4 around coffee break 1. For better readability, all curves start at level 0 at the left end of the diagram.

individual history of encounters and personal acquaintances certainly plays a 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. 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 interdisciplinary 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.

7. REFERENCES

[1] 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.

[2] 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 Verlag, Heidelberg, Germany, 2012.

[3] 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.

[4] A. Barrat and C. Cattuto. *Temporal Networks*, chapter Temporal Networks of Face-to-Face Human Interactions. Understanding Complex Systems. Springer Verlag, Heidelberg, Germany, 2013.

[5] 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.

[6] P. Bródka, S. Saganowski, and P. Kazienko. GED: The Method for Group Evolution Discovery in Social Networks. *SNAM*, 3(1):1–14, 2011.

[7] 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.

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

[9] N. Eagle, A. Pentland, and D. Lazer. From the Cover: Inferring Friendship Network Structure by using Mobile Phone Data. *PNAS*, 106:15274–15278, 2009.

[10] 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.

[11] 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.

[12] 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.

[13] G. Palla, A.-L. Barabasi, and T. Vicsek. Quantifying Social Group Evolution. *Nature*, 446(7136):664–667, April 2007.

[14] 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.

[15] J. C. Turner. Towards a Cognitive Redefinition of the Social Group. *Cah Psychol Cogn*, 1(2):93–118, 1981.