Anatomy of a Conference

Bjoern-Elmar Macek * Christoph Scholz * Martin Atzmueller Gerd Stumme
Knowledge and Data Engineering Group, University of Kassel
Wilhelmshöher Allee 73, D-34121 Kassel, Germany
{macek,scholz,atzmueller,stumme}@cs.uni-kassel.de

ABSTRACT

This paper presents an anatomy of Hypertext 2011 – focusing on the dynamic and static behavior of the participants. We consider data collected by the CONFERATOR system at the conference, and provide statistics concerning participants, presenters, session chairs, different communities, and according roles. Additionally, we perform an in-depth analysis of these actors during the conference concerning their communication and track visiting behavior.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Science

General Terms

Human Factors, Measurements

Keywords

social network analysis, rfid, proximity, contact network, conference

1. INTRODUCTION

In business and science, conferences provide important interactions: They foster the exchange of knowledge and enable face-to-face contacts between their participants for personal networking, e. g., in order to start interesting discussions, to form and strengthen cooperations (and business relations), and to initiate new projects. Understanding the mechanisms in such contexts is important to increase the efficiency and effectiveness of individual networking. Therefore, the analysis of conferences provides an interesting research field. However, such an analysis is not easy if conventional tools like questionnaires are used, since then mostly *static* analyses of the behavior and processes can be performed, while the *dynamic* nature of conference interactions is not accounted for.

In this paper, we present an in-depth analysis of the static and dynamic nature of a conference (Hypertext 2011). We collected data using the CONFERATOR system:[1] It employs active RFID technology provided by the SocioPatterns consortium.[2] CONFERATOR is a personalized conference management system for organizing social contacts and the conference program. Using the system, RFID data capturing the contacts and locations of the conference participants were collected at Hypertext 2011. To this end, we used a new generation of resource-aware active RFID tags, called proximity tags. The technical innovation of these proximity tags is the ability to detect other proximity tags within a range of up to 1.5 meters.

One of the first experiments using this kind of RFID tags at conferences was performed by Cattuto and colleagues, cf. [1, 9, 18]. We extend their findings with a number of (un-)expected results for homophily and session attendance of the participants. To the best of our knowledge, this paper proposes the first comprehensive analysis of the track attendance of the participants, their communication behavior and an analysis concerning their submitted papers. By investigating different correllations between the selected features in the data we find insights into the anatomy of the Hypertext conference 2011. We also describe an analysis of the data along several dimensions: First, we provide an overview of the collected data, discuss the overall structure, and analyze general effects concerning different groups (e.g., presenters, chairs, track participants, etc.). Furthermore, we consider different communities, e.g., concerning the individual tracks and sessions, but also automatically mined communities. We show an analysis of different roles in these contexts by characterizing the different participating subjects and groups at the conference and by mining role profiles.

The rest of the paper is structured as follows: Section 2 discusses related work. After that, Section 3 introduces the RFID-Setup and explains the CONFERATOR system in more detail. Next, Section 4 describes the collected dataset. Section 5 starts the analysis: We discuss the community structure and the static and dynamic analysis of the behavior of conference participants. Furthermore, we analyze different roles and derive role profiles using pattern mining. Finally, Section 6 concludes the paper with a short summary.

2. RELATED WORK

Homophily and mixing patterns in social networks have been investigated, e.g., by McPherson et al. [13] from a sociological point of view. They observed, that it is far more likely for people to connect to each other if they have something in common. We extend those findings by showing that in some contexts people are more interested in talking to people with different fields of interest. Cattuto and colleagues presented several important results by analyzing social dynamics in various environments using RFID technol-

*Both authors contributed equally to this work.

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

HT'12, June 25–28, 2012, Milwaukee, Wisconsin, USA.

Figure 1: Proximity tag (left) and RFID reader (right)

ogy: In [9], the authors compare the social activity of conference attendees with their research seniority and their activity in social web platforms like Facebook, Twitter and others. They also extend their focus to schools [17] and hospitals [11]. They present aggregations of contact measures between different groups of users. In contrast to their work, we focus on correlating the conversation profiles and the participants' track attendance with features like the track communities, the organizational roles within the conference such as session chairs and speakers, and their submitted papers. The characterization of nodes in a social network is an interesting and challenging task. Several works like [10] and [16] present methods to cluster nodes of a social network into different roles. In this work, we focus on the method proposed in [16], because this method allows us to consider a given community structure.

Subgroup discovery [20, 7] aims at identifying exceptional patterns with respect to a given target property of interest according to a specific quality measure. We apply subgroup discovery for the characterization of different roles. Similar work has been done, for example, in characterizing spammers [6], and in identifying profiles for the maturity of tags in social bookmarking systems [3].

3. CONFERATOR – A SOCIAL CONFERENCE MANAGEMENT SYSTEM

In the following section we first outline the active RFID technology used in the CONFERATOR system. Next we introduce the CONFERATOR and its functionality.

3.1 RFID Setup

One of the key components of CONFERATOR is a new generation of so-called proximity tags (see Figure 1), developed by the SocioPatterns project. The most important feature of these tags is the possibility to detect other proximity tags within a range of up to 1.5 meters, which allows the identification of face-to-face contacts.

The RFID setup at a conference requires the installation of RFID-Readers at fixed positions in the conference area. The RFID readers (see Figure 1) receive the signals from the tags that are worn by the participants and forward them to a central server. This makes it possible to determine the location of each tag and therefore the location of a conference participant at room-level basis. For obtaining the location of participants there are several options [15], including a simple algorithm proposed in [14]: Here, the participant is assigned to the room whose RFID readers received most packages with the weakest signal strength. For more details on the proximity tags, we refer to Barrat et al. [8] and the OpenBeacon website.[3]

3.2 Conferator

The CONFERATOR-system [2] is a social and ubiquitous conference guiding system. CONFERATOR consists of two parts: The TalkRadar[4] of the University of Pittsburgh. TalkRadar is based on Pittsburgh's Conference Navigator [19]. and the PeerRadar.

[3] `http://www.openbeacon.org`

[4] Since June 2011, CONFERATOR is jointly developed with the Personalized Adaptive Web Systems Lab (`http://www2.sis.pitt.edu/~paws/`)

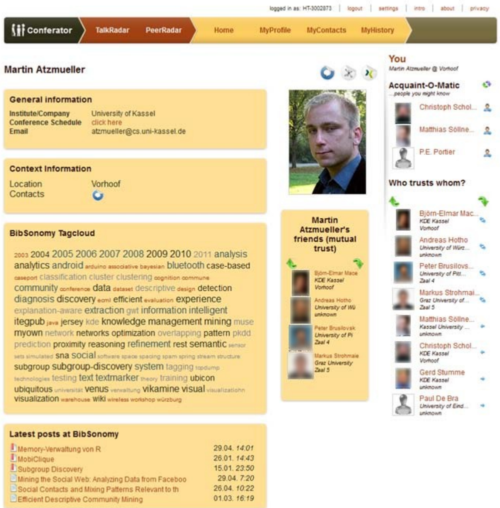

Figure 2: Screenshot of the CONFERATOR's PeerRadar showing a user profile page. The page shows information about latest BibSonomy posts, trust circles, context information (e.g. current position), social tags and general information (e.g. institute or email address).

TalkRadar allows conference participants to manage their conference schedule, PeerRadar is like an online business card, that supports the social interaction at a conference. In PeerRadar, for example, it is possible for conference participants to see their own contacts or to browse through other conference attendees' user profiles (see Figure 2). CONFERATOR has successfully been deployed at several events, e.g., the LWA 2010[5] and LWA 2011[6] conferences, the Hypertext 2011[7] conference, and a technology day of the VENUS[8] project. In this paper, we focus on data collected with the PeerRadar component of CONFERATOR at Hypertext 2011.

4. DATA SET

In the following section we first describe our dataset collected at the Hypertext 2011 conference in Eindhoven, before presenting some overview statistics of the collected data.

4.1 RFID Data

At the Hypertext 2011 conference, we asked each conference participant to wear an active RFID tag. All in all 75 of 95 participants took part in our experiment which started June 6, 2011 at 14:00 and ended June 9, 2011 at 14:00. In the four days of the conference we recorded 2620 face-to-face contacts between participants. As in [18], a face-to-face contact is recorded when the duration of the contact is at least 20 seconds. A contact ends when the two corresponding proximity tags do not detect each other for more than 60 seconds. Obviously the length of a contact plays an important role in defining a contact. In Figure 3, we see the distribution of the corresponding contact durations of all conference face-to-face contacts. Here, the x-axis represents the minimum duration of a contact in seconds, while the y-axis shows the probability of a contact having at least this duration. The axes are scaled logarithmically. As already observed, e. g., in [12] and [4], we see that

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

[6] `http://lwa2011.cs.uni-magdeburg.de/`

[7] `http://www.ht2011.org/`

[8] `www.iteg.uni-kassel.de/`

most of the contacts are less than one minute and that the durations show a long-tailed distribution. The average path length (APL) is also similar to the findings in [12] and [4].

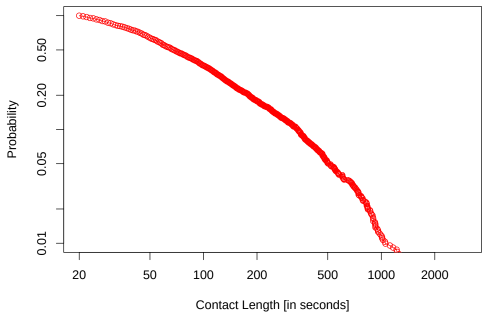

Figure 3: Cumulated contact length distribution of all face-to-face contacts between participants of the hypertext conference.

In the following, we introduce the notation for the contact graph $G_\Sigma(i)$. An edge $\{u, v\}$ is contained in $G_\Sigma(i)$, iff the sum of all contact durations between participants u and v is at least i seconds. In Table 1 we present some standard statistics of the contact graph $G_\Sigma(i)$. The diameter of the contact graph $G_\Sigma(i)$ shows similar values to those already presented results in [12][4].

The high average degree of the contact graph $G_\Sigma(20)$ indicates that those taking part in the experiment (at least briefly) came into contact with the majority ($\frac{41}{75} = 55\,\%$) of the other participants. For longer conversations this average degree decreases very quickly. Here, for example, in average each participant only has contact with approximately $\frac{10}{75} = 13\,\%$ of the other participants taking into account conversations longer than 10 minutes.

Table 1: General statistics for several contact graphs with different thresholds (in seconds). Here d is the diameter, APL the average path length and LCN the largest clique number in $G_\Sigma(i)$

Network	$\|V\|$	$\|E\|$	d	Avg.Deg.	APL	LCN
$G_\Sigma(20)$	68	698	4	41	1.76	14
$G_\Sigma(60)$	66	498	4	30	1.91	11
$G_\Sigma(300)$	60	246	5	16	2.36	8
$G_\Sigma(600)$	58	142	7	10	3.01	5
$G_\Sigma(900)$	53	98	8	7	7.39	4

In this paper, we focus on the different community structures, i.e. partitionings, induced by country of origin, academic status, affiliation with the Hypertext conference series, and affiliation with one of the four conference tracks. In Table 2, we present some statistics about the different community stuctures. We classify participants as highly affiliated with the Hypertext conference series if they presented a paper more than three times at Hypertext conferences in different years. The affiliation of a participant is low when he or she has never presented a paper or presented a paper at Hypertext 2011 for the first time. All other participants are classified with a medium affiliation. For every author and coauthor of a paper we define his or her track membership by the track the paper was submitted to. The session and track chairs are also assigned to their respective tracks. For attendees who could not be assigned to a track, this information is not available (*n/a*).

Table 2: Partitions of the set of participants into communities according to country, academic status, affiliation with HT and track. For each community, its number of participants is listed.

Country	
Australia	3
Austria	3
Belgium	2
Canada	2
Denmark	2
Finland	1
France	1
Germany	11
Ireland	2
Italy	5
Japan	6
Netherlands	9
Poland	1
Slovakia	1
Spain	3
United Kingdom	10
USA	10
n/a	3

Academic Status	
Professor	14
PhD-candidate	34
PhD	20
Other	7
Affiliation with HT	
high	12
medium	17
low	46
Track	
DynHyp	12
SocialMedia	19
StoryTelling	6
UbiquHyp	5
n/a	33

As already mentioned we placed several RFID readers at fixed positions in the conference area. To identify the track attendance of all participants we particularly fixed one RFID reader in each lecture room. Figure 4 gives an overview about how many track members attended their own and the other tracks, respectively.

In the Social Media pie chart we see for example, that 60% of all participants who visited the Social Media track are also members of the Social Media track. 5% are members of the Interaction, Narrative and Story Telling track, 11% are members of the Emerging Structures and Ubiquitous Hypermedia track and 24% are members of the Dynamic and Computed Hypermedia track. A more detailed analysis of the track attendance and the behaviour of the participants is described in Section 5.2.

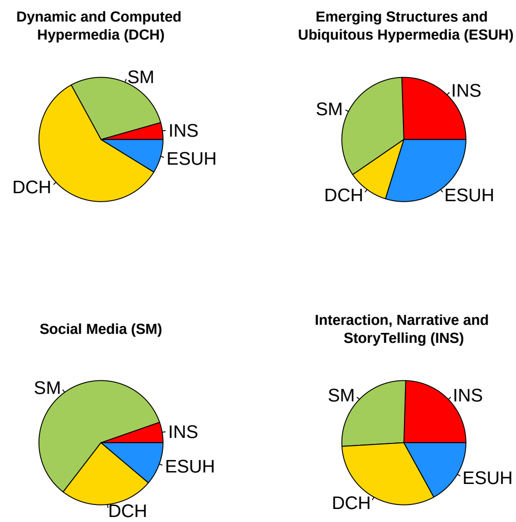

Figure 4: Overview of the track attendance for the different tracks. Each pie chart visualizes the distribution of track attendance by members of the different tracks.

5. ANALYSIS

In this section, we investigate the correlation between the given community structures and their contact patterns, followed by an in-depth analysis of the conversation behavior of participants and their visited tracks and sessions. We conclude the analysis by extracting several roles from contact graphs in order to reveal additional information on how the participants are embedded within the social network of this conference. For this purpose, we mine descriptive (subgroup) patterns characterizing prominent roles, and include a detailed time-based analysis.

5.1 Community Structure

In the following, we analyze the connection between the link structure of the contact graph and the four partitionings in communities listed in Figure 2. To analyze the compatibility of the link structure and a community structure, we use the alignment measure proposed in [16]. For this measure we recall from [16] the definitions of *complete node pairs* and *pure node pairs*. A *complete node pair* is a pair (u, v) of nodes where both nodes u and v are linked and belong to the same community. A *pure node pair* is a pair (u, v) where u and v are not linked and do not belong to the same community. As in [16], we define the parameters p and q as

$$\begin{aligned} p &= \frac{\text{\# complete node pairs}}{\text{\# total linked node pairs}} \\ q &= \frac{\text{\# pure node pairs}}{\text{\# total non-linked node pairs}} \end{aligned} \tag{1}$$

Here we note, that high values for p and q indicate that the community structure fits the link structure well. In our experiments we use the p- and q-values to analyze how the four different community partitionings induced by track, country, academic status and affiliation are aligned to the link structure of the hypertext contact graph. We focus in particular on the change of alignment when only longer contacts are considered. This means that we calculate and compare the p- and q-values for the contact graphs $G_\Sigma(60)$, $G_\Sigma(120)$, $G_\Sigma(180), \ldots$.

The results are shown in Figure 5 (p-value) and in Figure 6 (q-value). In these figures, for example, looking at contacts with contact lengths of more than 1 minute, we observe that the probability of being in contact within the same track-community is 39.3%. If there is no contact between two persons the probability of them being in different communities is 82.1%. In general, we see that the p-value fluctuation of the community structures, affiliation, country and academic status over the different time thresholds is rather low. Only the p value for country increases from 18.2% to 41.1% between time threshold 1 and 26.

Looking at the p-value for the community structure track we see an interesting development. The greater the length of a conversation the higher the probability of having a contact within the same track-community. Here, the increase is from 39.3% to 83%. A possible reason for this might be that some tracks are filtered out, because of the increasing time threshold. For example, when only participants of one track are available the p-value is clearly one. In this paper, we show that the probability to have a contact within the same community is dependent on the contact length. We validate our conclusion by calculating the p-values for the community structure track over different permutations of the participants' track attendance. Here, we repeat the experiment 100 times and average the p-value results. The result is shown in Figure 7. We see that the p-values of the real community structure increase much faster than the p-values of the random community structure. In Figure 6, we see that the q-value for all community structures track, country and affiliation and academic status is monotonically increasing. This is not surprising since the increase (from time threshold t to $t+1$) of the number of *total non-linked node pairs* must be at least the increase of the number of *pure node pairs*.

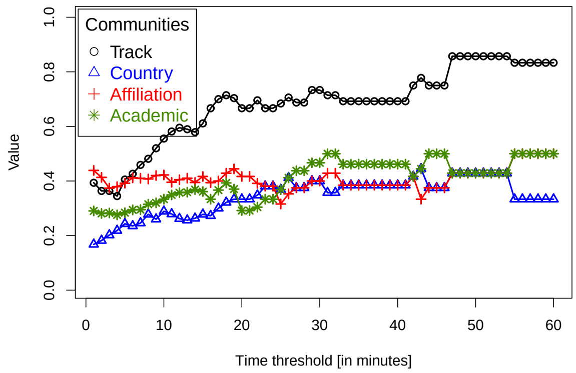

Figure 5: Overview of the p value results for the community structures track, country, affiliation and academic status.

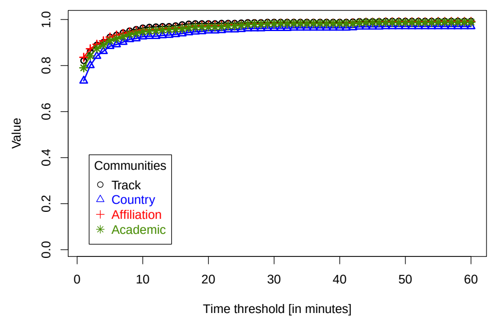

Figure 6: Overview of the q value results for the community structures track, country, affiliation and academic status.

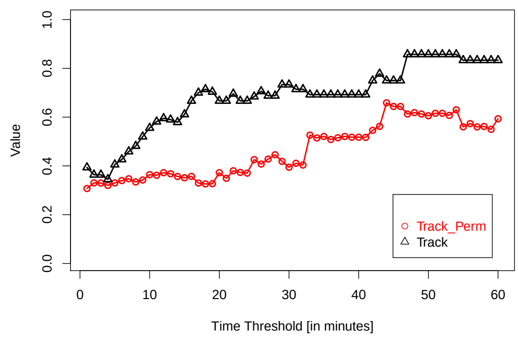

Figure 7: Overview of the p value results for the community structure track and the average p values of the community structure over 100 permutations of the participants' track-membership.

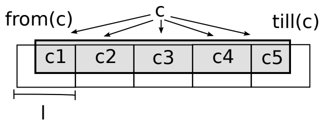

Figure 8: Example of a contact c sliced into five different parts with a maximum length of l seconds.

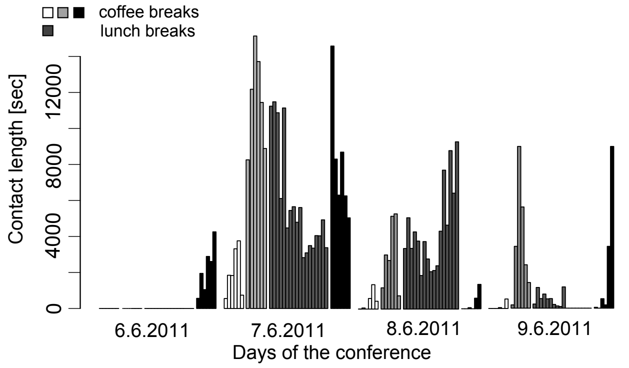

Figure 9: Time slices containing contact durations for the complete conference except for the sessions. The start times of the coffee breaks were as follows: 8:30, 10:30, 15:30 and 16:00. Their duration was always 30 minutes. The start and duration of lunch breaks varied. Except for the last day all started at 12:30 and took at least one hour. Each bar represents a 5 minute slice; adjacent bars belong to the same break.

5.2 Communication and Tracks

In this section, we analyze how the participants and different tracks connect with each other. Furthermore, we indirectly consider their current research topics using contacts and session attendances as proxies. We discriminate between several relevant groups of time intervals in the conference's schedule, namely the poster session, the sessions (where the speakers present their work), the coffee break, and the lunch breaks after the sessions. Since there are almost no conversations during the lectures, we also take the breaks and the poster session into account when analyzing the contacts.

We interpret the contact lengths as a measure for social activity. In order to capture the change of social dynamics over time we divide the contacts into intervals of a fixed length $l = 5$ minutes, as depicted in Figure 8.

5.2.1 Contacts on a Global Scale

In order to get a general overview of the social activity, we present a complete overview of all the breaks of the four day conference in Figure 9. Since the setup of the CONFERATOR system started in the middle of the first day, all previous time slices are empty. As expected, there were a lot of interactions between participants which decreased over time as the conference progressed. This can partly be explained by leaving participants who were returning their RFID tags. The short peaks at the last two coffee breaks are also an exception and might be explained by the conference attendees saying goodbye to each other.

5.2.2 Social Activity of Communities

Hypertext 2011 addressed a variety of research fields. It started in 1987 as a group of researchers and companies with the main focus on hypertext and the internet and first widened its interest to *Interaction, Narrative, and Storytelling* (INS). Afterwards it broadened its scope towards the social and semantic web and finally to ubiquitous topics. This is reflected in the the four tracks in 2011.

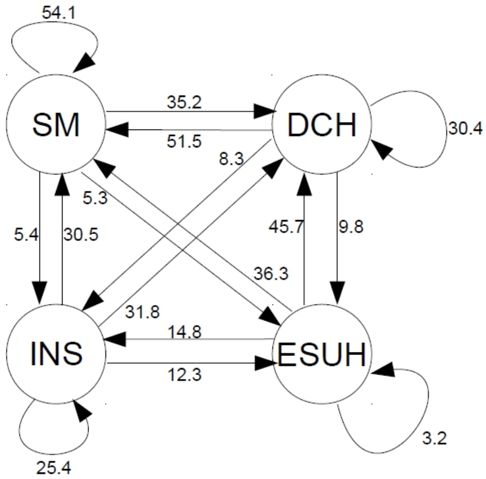

Figure 10: The contact length distribution (in percent) for conversations between all combinations of tracks. Here, for example a directed edge from track SM to track INS with weight 5.4 indicates, that the fraction of all cumulated contacts between the SM track and the INS track relative to all cumulated contacts of SM is 5.4 percent.

Since the benefits of greater creativity in a broader and more diverse environment only appear if conversations and exchange of ideas is going on beyond the tracks' bounds, we investigate the links between the tracks. A complete overview of the social activity for these communities is given in Figure 10. Obviously all tracks are linked very well - which indicates good opportunities for inspiring conversations. Nevertheless, there are some differences: While the older tracks are focused on talking with their own members and the biggest communities, namely *Dynamic and Computed Hypermedia* (DCH) and *Social Media* (SM), *Emerging Structures and Ubiquitous Hypermedia* (ESUH) as the youngest addition to this conference concentrated primarily on communication with the two larger tracks and less with their direct research colleagues.

Figure 12 shows the communication structure between professors, post docs and research assistants. It is noticeable that conversations between professors and research assistants are significantly shorter than conversations between members of other groups such as for instance professors and post docs. These two groups actually had the three longest conversations among them during the experiment.

5.2.3 Individual Social Activity

A closer look at the communication structure reveals that, as expected, participants can get alot of attention by holding a talk. What might be unexpected is the people who will be attracted. For our analysis we do not consider two keynote lectures for which the presenters did not wear RFID tags. Furthermore, the session directly before the poster session is also excluded, since we assume that the attention easily shifted away from the recent speakers of the last session. The final series of lectures is also removed due to the low number of participants. In Figure 13, we plot the distribution of contact lengths between all tracks, highlighting the two that just ended their parallel session. The average contact lengths per track member depicted in this figure reveal that the majority of participants talked to members of those tracks that just presented their

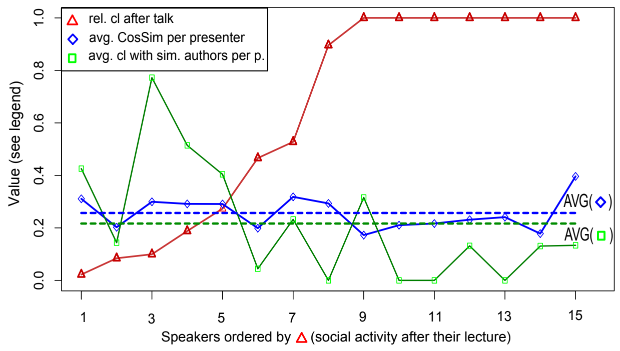

Figure 11: The triangles denote the normalized contact lengths (cl): dur_u, while the averaged similarities $CosSim^u$ of the paper of u to the other research results in the proceedings are represented by the diamonds. The percentage of the duration of contacts , that u had with the top 10 similar speakers based upon $CosSim$: $t_{\text{rel}}(u)$ are represented by the squares.

work. We examplarily plot the data for only one session. However, the same observation holds for five of the six considered coffee breaks. So being a member of a track that recently gave a lecture in a session seems to attrack conversation partners.

It seems self-evident that the social attention is directed towards the speakers of the recent session, but this is only partially true. In the following, we examine the hypothesis that a speaker is socially more active in the break after the session in which he presented his work: We calculate the duration of all contacts in this interval for each speaker; $C^{u_1}_{u_2}[t_{\mathsf{from}}, t_{\mathsf{till}}]$ represents all contacts between users u_1 and u_2 from t_{from} to t_{till}. The sub- and superscripts of C and the denoted timestamps are interpreted as a filter for the contained contacts. Following this semantic, '*' will be used as a wildcard symbol. The sum of all contact lengths in seconds for a given set of conversations C is given by $\mathsf{dur}(C)$. We aim to keep the values comparable despite the different social nature of users u – some tend to talk more in general than others. Furtermore, coffee breaks are significantly shorter than lunch breaks. Therefore, we divide the durations by $\max_{\mathsf{dur}_u}$: This equals the maximum of the sum of all contact durations during each break of the same category for user u. Let t_{from} denote the start and t_{till} the end of the respective break, then

$$\mathsf{dur}_u = \frac{\mathsf{dur}(C^u_*[t_{\mathsf{from}}, t_{\mathsf{till}}])}{\max_{\mathsf{dur}_u}}$$

is a value in the interval [0, 1]. The higher the value, the more socially active was the user during this time. For $\mathsf{dur}_u = 1$, the break after the presentation was indeed the most active one.

As discussed above, we removed all speakers for our analysis that either did not wear an RFID tag or had their talk directly before the poster session, since the it has its own social dynamics. The values for all speakers are plotted in Figure 11. The speakers are ordered on the x-axis by increasing dur_u. It is easy to see, that seven (46%) of the observable speakers were most active after their lecture.

Then, a natural question is, whether there are any features that connect these seven speakers. An intuitive hypothesis claims, that presenters whose papers are related to the work of a large number of other presenters get more attention. However, in the data we cannot confirm this. In order to analyze, if increased social activity is related to the content of the presented work, we analyze the documents contained in the Hypertext proceedings: For every pair of speakers $u_1, u_2 \in S$, $CosSim(u_1, u_2)$ measures the cosine similarity of the stemmed bag of words representation of their papers with all stop words removed. In order to capture the overall relatedness of one paper to the others, we calculated the average value of all paper similarities with all other speakers' work for each presenter $u \in S$:

$$CosSim^u = \sum_{u' \in S, u \neq u'} \frac{CosSim(u, u')}{|S|}$$

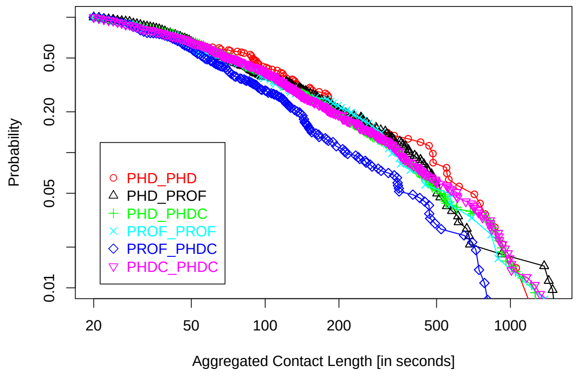

Figure 12: The cumulative contact length distribution for conversations between professors(PROF), post docs(PHD) and research assistants(PHDC). The two communities are seperated with an underscore within the legend.

The values were also plotted in Figure 11, marked with a diamond. Obviously, the hypothesis, that a higher $CosSim^u$, the more people might be interested in the work and also in speaking with the author does not hold. There is no direct correlation between both values. Nevertheless, it is worth mentioning that for five of the seven presenters who did not have an increased social activity ($\mathsf{dur}_u < 0,8$), the paper similarity measure is above average, while for six of out the seven presenters who were most active after their session, the value is below. This is the exact opposite of what might be expected. Since the differences between high and low values are too insignificant, we cannot draw strong conclusions. A reason for this might be that some information is lost by averaging the similarity values; the existence of speakers u' with very high values $CosSim(u, u')$ is not reflected.

Instead, we now focus on the contact lengths of speaker u to those 10 other speakers S_u whose work is most similar to their own. We plotted the following values and their average in the Figure 11:

$$t_{\text{rel}}(u) = \sum_{u' \in S_u} \frac{\mathsf{dur}(C^u_{u'}[*, *])}{\mathsf{dur}(C^u_*[*, *])}$$

We obtain a similar result as before, but observe a much stronger inverse correlation with the normalized contact lengths. This seems to justify the hypothesis, that speakers get a lot of attention mostly from those participants who did not present very similar research results. In the context of [13] and most of the assumptions in the state of the art of social network analysis, this result is surprising, since it is not only "similarity that breeds connection" but also differences.

Furthermore, not only the breaks after a session are of special interest, since the breaks before a session provide the possibility for session chairs and speakers to coordinate their presentation or clar-

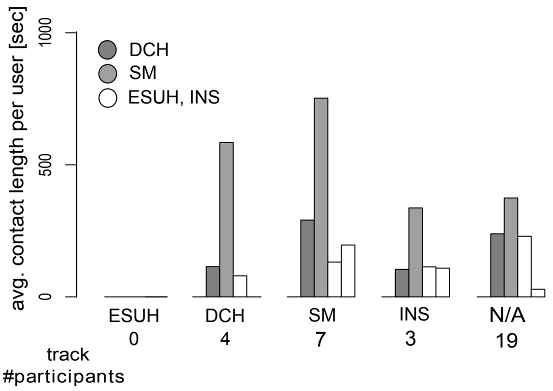

Figure 13: Normalized contact lengths for conversations between the participants of the different tracks after parallels session of the SM and the DCH track. Attendees without track assignment are denoted by category N/A.

ify final questions, e.g., the technical setup of the speaker's desk. Therefore, we tested the hypothesis that the structure of a conference organization may be reflected in the contact data. Despite the fact that there was only a small number of session chairs at the conference and some of them did not wear an RFID tag, there were no significant contacts between speakers and session chairs of the same track directly before the presentations.

5.2.4 *Session Attendance of Communities*

In the following, we examine the session attendance of the participants. We measure the attention and popularity of the given tracks by interpreting the session attendance as a decision process in which the members of the audience had to choose between two tracks to follow (see Figure 14).

The most obvious observation is that all tracks focused on their own community. Also a phenomenon that correlates with the observations above is that the new community ESUH played a special role at the Hypertext 2011 as it got a lot of attention from other tracks. This might reflect the mutual interest in one another and the beginning of an integration process of the communities. The big picture shows that SM was the most popular track and had even more attendance than the DCH track in 2011.

5.2.5 *Session Attendance of Individuals*

Based upon the hypothesis that people who focus on attending sessions of a favored track also have the most contacts to its members, we calculate the following two vectors for each user: One contains the number of presentations visited for each track and the other contains the length of contacts with its members. The cosine similarities between those two vectors are plotted in Figure 15.

For all tracks the values span the full range from low to very high correlation. The core, however, has a significantly higher average than the small communities. This is not surprising, while INS like all of the older tracks is mainly focused on exchanging ideas with their colleagues they might already know from a Hypertext in previous years. They had only a small number of lectures compared to the rest. This leads to other tracks being visited more often than their own. For ESUH it is very similar. While the number of lectures is comparable to the tracks from the core giving them the opportunity to focus on their own presentations, they tend to socialize more with the core - maybe due to the integration process.

The core itself has far better opportunities to only listen to topics and also talk to members of their own tracks which is directly reflected in Figure 15.

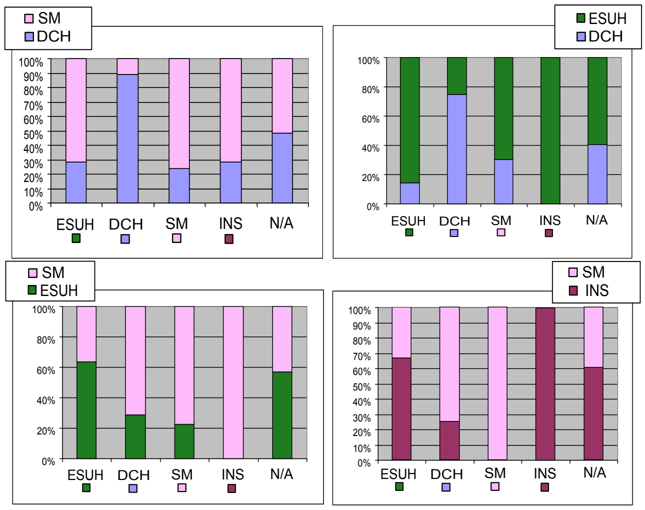

Figure 14: In five sessions, the four pairs of tracks were held in parallel (the combination of DCH/SM occurred twice). Each bar shows which percentage of the members of the respective track were spent in either of the two parallel sessions.

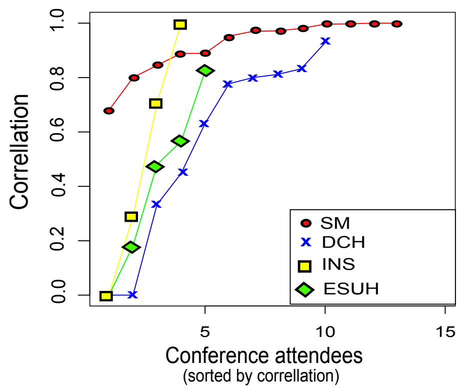

Figure 15: Correlation for the attendances of tracks and contacts to track members.

5.3 Roles

The characterization of nodes in a social network is a very challenging task. In this section we focus on exploring the connection between academic jobs and influential and authoritarian persons of the Hypertext conference. First, we discuss the concepts for determining roles. After that, we present a detailed time-based analysis of role patterns. Finally, we use subgroup discovery to find more interesting patterns.

5.3.1 *Determination of Roles*

For this purpose we use a technique proposed in [16], which devides all nodes (conference participants) into four roles: Ambassador, Bridge, Loner and Big Fish. Intuitively, Ambassadors are nodes with contacts to many diverse communities, whereas Big Fishes only have a lot of contacts within one or at least less communities. Bridges are similar to Ambassadors, but with less contacts. A Loner is a role with less connections to different communities and less contacts.

In the following, we define these four roles more formally. Here, we use the definition given in [16]. To identify one of the four roles [16] used the relative degree of a node and a community metric. Whereas the relative degree of a node is simply the degree of the node divided by the maximum degree of all nodes, it is much harder to calculate the community metric. Here [16] present the new community metric $rawComm$ that estimates the number of communities a node is connected to. The community metric $rawComm$ for a node u is defined as

$$rawComm(u) = \sum_{j \in N(u)} r_u(v), \tag{2}$$

where $N(i)$ is the neighborhood of node u. The function $r_u(v)$ is the community membership contribution from node v to node u. In [16] the function $r_u(v)$ is defined on unweighted graphs. We extend the definition of [16] for weighted graphs, taking into account the observation of section 5.1 that the probability of conversations to be in the same commnunity is dependent on the conversation length. We define the community membership contribution $r_u(v)$ from node v to node u as

$$r_u(v) = \frac{1}{1 + \sum_{k \in n_1} p_k + |n_2|(1-q)}, \tag{3}$$

where n_1 is a set of nodes in $N(u)$ that is linked to u. n_2 is a set of nodes in $N(u)$ that is not linked to u and p_k is the probability that a link of node k to v with weight w exists within the same community. The probability q is defined in equation 1.

Now we can define the four roles Ambassador, Bridge, Big Fish and Loner for node u. The role of node u is defined as

$$role(u) = \begin{cases} \text{Ambassador} & rdeg(u) \geq s, rawComm(u) \geq t \\ \text{Bridge} & rdeg(u) < s, rawComm(u) \geq t \\ \text{Big Fish} & rdeg(u) \geq s, rawComm(u) < t \\ \text{Loner} & rdeg(u) < s, rawComm(u) < t \end{cases},$$

, where $s, t \in [0, 1]$ are appropriate thresholds.

As described in Section 5.1, concerning all analyzed community structures, the track community fits best to the link structure of the social network. For this reason we decided to use the track community structure to analyze the function of all nodes in the network. One question that could arise here is why we do not simply count the number of communities a node is connected to. Unfortunately, as described below we do not know the tracks of all conference participants. For this reason we use the afore mentioned probabilistic model to determine the roles of the whole graph with the rawComm metric. We will compare our results to a similar analysis of another conference that was performed in [4].

5.3.2 Time-based Analysis

In our experiments we tested a lot of threshold parameters s and t. It turned out that the parameter setting $s, t = 0.4$ is a good choice to find an adequate number of Ambassadors and Bridges. In Figure 16 we see the results for the Ambassador analysis. As expected for conservations of two minutes or longer, most of the professors, session chairs and oldies function as Ambassadors. Here for conversations of one/two minute(s) or longer 75% of the professors are Ambassadors, 17% are Bridges and 8% are Loners.

As shown in Figure 17 for conversations longer than five minutes professors, oldies and session chairs become Bridges, and retain that status for conversations of greater lengths. A possible explanation of this is that for instance professors entering a conference venue generally know quite a number of people there. Thus they briefly greet and get into contact with many people and thereby function as Ambassadors. These conversations, however, will not take more than five minutes in most cases. Then, professors will possibly start having longer conversations with few people they know best. This is how they might lose their status as Ambassadors.

The observation that professors lose their status as Ambassadors is different to the observation in [4]. Here professors retain their status as Ambassadors over the whole time. An interesting observation is that similar to the result in [4] the number of Big Fishes is very small. In Figure 18 we see that the fraction of professors, oldies and session chairs who are Loners is significantly smaller than that of phd-candidates, phds and presenters.

The differences to [4] concerning the results of the role analysis might be explained by the different kind of conference: The respective conference (LWA, of the german computer science society GI), is only held in Germany, and is regularly visited by a rather stable community. This offers the opportunity for more familiar relationships between researchers, which potentially results in longer conversations in general.

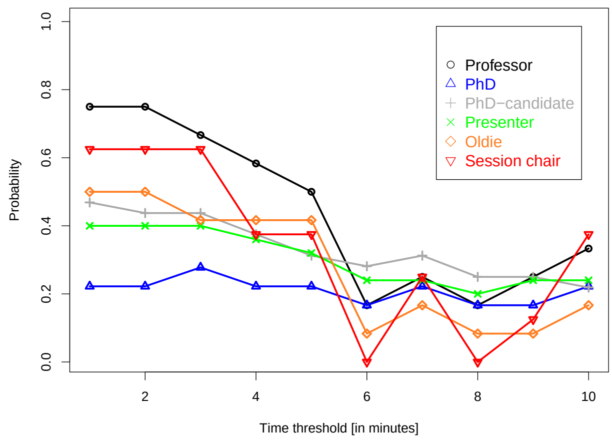

Figure 16: Fraction of professors, session chairs, etc. that belongs to the Ambassador role.

5.3.3 Mining Role Patterns

To characterize the different roles of the participants, we applied subgroup discovery techniques for mining role patterns. Subgroup discovery (cf. [5, 7]) aims at identifying interesting patterns with respect to a given target property according to a specific interestingness measure. Pattern mining using subgroup discovery is especially suited for identifying local patterns in the data, that is, *nuggets* that hold for specific subsets.

In our context, the target properties of interest are given by the different roles of participants in the contact graph. We aim at describing a subgroup (set of participants) with a specific role as closely as possible using a set of descriptive features, e.g., their country of origin, title, role as session chair, invited speaker, or presenter of a conference paper. We computed the roles according to different minimal conversation lengths (60, 180, 300 seconds). For subgroup discovery, we applied then the according role distributions. In the following, we discuss several examplary results. For an overview of the distribution of roles in the different episodes we refer to Table 3.

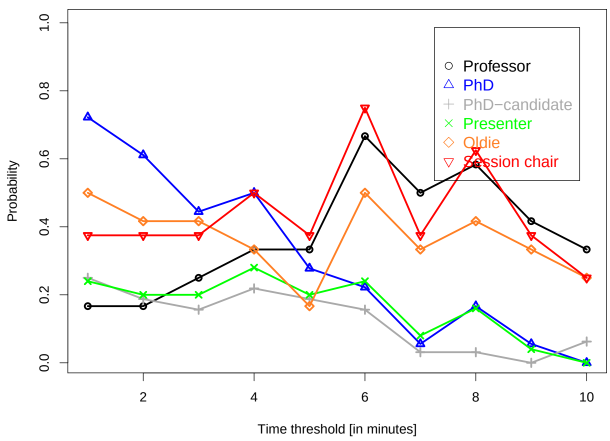

Figure 17: Fraction of professors, session chairs, etc. that belongs to the Bridge role.

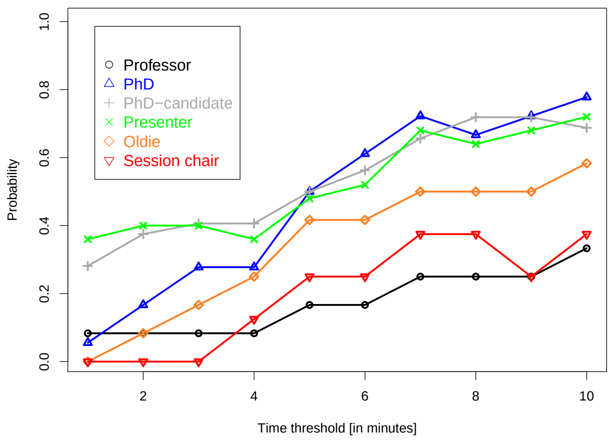

Figure 18: Fraction of professors, session chairs, ... that belongs to the Loner role.

Concerning the minimal conversation length of 60 seconds (Table 4), it is easy to see that the session chairs serve as Ambassadors during the confernce (the remaining session chairs are Bridges). Furthermore a *strong* affiliation to Hypertext plays an important role for being an Ambassador for the conference. The feature *Affiliation* denotes the familiarity with Hypertext, such that authors of at most one Hypertext paper published in 2011 get a *low* affiliation score, authors who published one or two papers before Hypertext 2011 get a *medium* affiliation score, and authors with at least 3 papers before Hypertext 2011 get a *strong* affiliation score. Considering the 60 seconds threshold, it is also evident that the participants from the Netherlands (including in particular the organizers) are typical bridges, as expected. This is especially visible in subgroup #3 of Table 4 with a target share of 100%.

Table 3: Overview on role shares (absolute/relative frequency) of the 66 conference participants wearing RFID-tags considering their (conversation) contact graphs: The table shows the statistics for three minimal conversation length thresholds (60, 180, 300 seconds).

sec	#Ambassador (%)	#Bridge (%)	#Loner (%)
60	29 (0.44)	26 (0.40)	11 (0.17)
180	28 (0.42)	18 (0.27)	20 (0.30)
300	21 (0.32)	16 (0.24)	28 (0.42)

Table 4: Subgroup results for a minimal conversation length of 60 seconds. The table shows the target variable, the lift (relative target increase w.r.t. the default), the share of the target in the subgroup, the size of the subgroup, and the subgroup pattern.

Min. Contact Length: 60 sec					
#	Target	Lift	Share	Size	Pattern
1	Ambassador	1.42	0.63	8	SessionChair=true
2	Ambassador	1.14	0.50	12	Affiliation=strong
3	Bridge	2.54	1.0	6	Country=Netherlands AND Presenter=No
4	Bridge	2.18	0.86	7	Country=Netherlands
5	Bridge	0.95	0.37	8	SessionChair=true

Considering the minimal conversation length of 180 seconds (Table 5) the overall picture changes a little. While the session chairs are stable in their roles, it seems, that the strengths of the Ambassador and Bridge associations is decreased.

Table 5: Subgroup results for a minimal conversation length of 180 seconds. The table shows the target variable, the lift (relative target increase w.r.t. the default), the share of the target in the subgroup, the size of the subgroup, and the pattern.

Min. Contact Length: 180 sec					
#	Target	Lift	Share	Size	Pattern
1	Ambassador	1.47	0.63	8	SessionChair=true
2	Ambassador	0.98	0.42	12	Affiliation=strong
3	Bridge	1.05	0.29	7	Country=Netherlands
4	Bridge	1.83	0.50	6	SessionChair=true AND Affiliation=strong
5	Bridge	1.53	0.42	12	Affiliation=strong
6	Bridge	1.38	0.37	8	SessionChair=true

The 300 seconds minimal conversation length (which usually excludes smalltalk) continues the trend regarding the organizers, cf. Table 6. For the session chairs, their bridge role stabilizes. This is especially interesting concerning the session chairs who are not track chairs. The role of Ambassador is only more pronounced for those session chairs that also have a strong affiliation with the Hypertext conference.

6. CONCLUSIONS

In this paper, we described the anatomy of a conference – focusing on the dynamic and static behavior of the participants at Hypertext 2011. For the analysis, we applied data collected by the Conferator system. We presented basic overview statistics concerning the participants, the presenters, the session chairs, and different communities. Additionally, we performed an in-depth analysis of these actors during the conference concerning their communication behavior, their session and track attendence, and the influence of the according communities. We also analyzed the roles of the conference participants in a time-based analysis and a pattern mining approach for the characterization of roles.

Table 6: Subgroup results for a minimal conversation length of 300 seconds. The table shows the target variable, the lift (relative target increase w.r.t. the default), the share of the target in the subgroup, the size of the subgroup, and the pattern.

Min. Contact Length: 300 sec					
#	Target	Lift	Share	Size	Pattern
1	Ambassador	1.57	0.50	6	SessionChair=true AND Affiliation=strong
2	Ambassador	1.31	0.42	12	Affiliation=strong
3	Ambassador	1.18	0.37	8	SessionChair=true
4	Bridge	2.48	0.60	5	SessionChair=true AND TrackChair=false
5	Bridge	1.55	0.37	8	SessionChair=true

In summary, we found that longer conversations are more probable, if the dialogue partners are both members of the same track. In contrast to intuition, an analysis of the presenters showed, that these were more involved in talks with participants presenting rather dissimilar work based on the content of their papers. Finally, using a combined approach of applying role mining and subgroup discovery, we found that the strenght of the affiliation is one of the strongest features in patterns (as would be expected) that determines the ability to connect between different communities. Overall, our analyses span a wide range and should enable the reader to obtain a good impression of conference interactions – most specifically for the Hypertext conference.

7. ACKNOWLEDGEMENTS

This work has been performed in the VENUS research cluster at the interdisciplinary Research Center for Information System Design (ITeG) at the University of Kassel. VENUS is supported by the government of Hesse as part of the program for excellence in research and development (LOEWE). CONFERATOR applies active RFID technology which was developed within the SocioPatterns project, whose generous support we kindly acknowledge.

8. REFERENCES

[1] H. Alani, M. Szomszor, C. Cattuto, W. V. den Broeck, G. Correndo, and A. Barrat. Live Social Semantics. In *Intl. Semantic Web Conference*, pages 698–714, 2009.

[2] 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*, 3:1–6, 2011.

[3] M. Atzmueller, D. Benz, A. Hotho, and G. Stumme. Towards Mining Semantic Maturity in Social Bookmarking Systems. In *Proc. Workshop Social Data on the Web, 10th Intl. Semantic Web Conference*, 2011.

[4] M. Atzmueller, S. Doerfel, A. Hotho, F. Mitzlaff, and G. Stumme. Face-to-Face Contacts during a Conference: Communities, Roles, and Key Players. In *Proc. Workshop on Mining Ubiquitous and Social Environments (MUSE 2011) at ECML/PKDD 2011*, 2011.

[5] M. Atzmueller and F. Lemmerich. Fast Subgroup Discovery for Continuous Target Concepts. In *Proc. 18th Intl. Symposium on Methodologies for Intelligent Systems (ISMIS 2009)*, volume 5722 of *LNCS*, pages 1–15, 2009.

[6] 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*, 2009.

[7] M. Atzmueller, F. Puppe, and H.-P. Buscher. Exploiting Background Knowledge for Knowledge-Intensive Subgroup Discovery. In *Proc. 19th Intl. Joint Conference on Artificial Intelligence (IJCAI-05)*, pages 647–652, 2005.

[8] 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.

[9] A. Barrat, C. Cattuto, M. Szomszor, W. V. den Broeck, and H. Alani. Social Dynamics in Conferences: Analyses of Data from the Live Social Semantics Application. In *Proceedings Intl. Semantic Web Conference*, volume 6497 of *Lecture Notes in Computer Science*, pages 17–33, 2010.

[10] B.-H. Chou and E. Suzuki. Discovering Community-Oriented Roles of Nodes in a Social Network. In *DaWak*, pages 52–64, 2010.

[11] L. Isella, M. Romano, A. Barrat, C. Cattuto, V. Colizza, W. V. den Broeck, F. Gesualdo, E. Pandolfi, L. Rava, C. Rizzo, and A. E. Tozzi. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. *CoRR*, abs/1104.2515, 2011.

[12] 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*, abs/1006.1260, 2010.

[13] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annu. Rev. Sociol.*, 27(1):415–444, 2001.

[14] M. Meriac, A. Fiedler, A. Hohendorf, J. Reinhardt, M. Starostik, and J. Mohnke. Localization Techniques for a Mobile Museum Information System. In *Proceedings of WCI (Wireless Communication and Information)*, 2007.

[15] C. Scholz, S. Doerfel, M. Atzmueller, A. Hotho, and G. Stumme. Resource-Aware On-Line RFID Localization Using Proximity Data. In *Proc. ECML/PKDD 2011*, pages 129–144, 2011.

[16] J. Scripps, P.-N. Tan, and A.-H. Esfahanian. Exploration of Link Structure and Community-Based Node Roles in Network Analysis. In *ICDM*, pages 649–654, 2007.

[17] J. Stehle, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. V. den Broeck, C. Regis, B. Lina, and P. Vanhems. High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. *CoRR*, abs/1109.1015, 2011.

[18] M. Szomszor, C. Cattuto, W. V. den Broeck, A. Barrat, and H. Alani. Semantics, Sensors, and the Social Web: The Live Social Semantics Experiments. In *Proc. ESWC*, pages 196–210, 2010.

[19] C. Wongchokprasitti, P. Brusilovsky, and D. Para. Conference Navigator 2.0: Community-Based Recommendation for Academic Conferences. In *Proc. SRS*, 2010.

[20] S. Wrobel. An Algorithm for Multi-Relational Discovery of Subgroups. In *Proc. 1st Europ. Symp. Principles of Data Mining and Knowledge Discovery (PKDD-97)*, pages 78–87, 1997.