Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles

Martin Atzmueller[1], Stephan Doerfel[1], Andreas Hotho[2],
Folke Mitzlaff[1], and Gerd Stumme[1]

[1] University of Kassel, Knowledge and Data Engineering Group
Wilhelmshöher Allee 73, 34121 Kassel, Germany

[2] University of Würzburg, Data Mining and Information Retrieval Group
Am Hubland, 97074 Würzburg, Germany

{atzmueller, doerfel, mitzlaff, stumme}@cs.uni-kassel.de
hotho@informatik.uni-wuerzburg.de

Abstract. This paper focuses on the community analysis of conference participants using their face-to-face contacts, visited talks, and tracks in a social and ubiquitous conferencing scenario. We consider human face-to-face contacts and perform a dynamic analysis of the number of contacts and their lengths. On these dimensions, we specifically investigate user-interaction and community structure according to different special interest groups during a conference. Additionally, using the community information, we examine different roles and their characteristic elements.

The analysis is grounded using real-world conference data capturing community information about participants and their face-to-face contacts. The analysis results indicate, that the face-to-face contacts show inherent community structure grounded using the special interest groups. Furthermore, we provide individual and community-level properties, traces of different behavioral patterns, and characteristic (role) profiles.

1 Introduction

During a conference, social contacts form an essential part of the experience of the participants. Commonly, the term "networking" is used for describing the inherent processes in such interactions. Typically, there are different (implicit and explicit) communities present at a conference, defined according to interests or membership to certain tracks or special interest groups. In order to enhance the conferencing experience, e.g., for obtaining interesting recommendations for papers and contacts, knowledge discovery techniques can often be applied. During a conference, ubiquitous computing approaches, e.g., based on RFID-tokens, provide dynamic adaptation options.

In this paper, we focus on the analysis of social data and contact patterns of conference participants: We consider communities of participants and their visited talks and we analyze their face-to-face contacts in these contexts. We examine different explicit and implicit roles of the participants, validate the community structures, and analyze various structural properties of the contact graph.

Our contribution is three-fold: We present an in-depth analysis of the social relations and behavioral patterns at the conference, analyze different community structures, and identify characteristics of special roles and groups. We focus on profiles of the participants and their dynamics considering face-to-face contacts, and a set of different community structures. For these, we use grounded information considering special interest groups, as given by the participants during registration in comparison to the emerging communities at the conference. In this way, we investigate how the established community structures emerge in the dynamic conference context. Finally, we describe and analyze different roles and groups, e.g., organizers and different subcommunities at a conference, in order to identify distinguishing characteristics.

The rest of the paper is structured as follows: Section 2 discusses social applications for conferences, and issues of privacy and trust. After that, Section 3 considers related work. Next, Section 4 provides the grounding of our approach presenting an in-depth analysis and evaluation of real-world conference data. Finally, Section 5 concludes the paper with a summary and interesting directions for future research.

2 Social Conferencing

During a conference, participants encounter different steps and phases: Preparation (before the conference), during the actual conference, and activities after the conference. Appropriate talks and sessions of interest need to be selected; talks and discussions need to be memorized. Additionally, social contacts during a conference are often essential, e.g., for networking, and are often revisited after a conference, as are the visited talks. All of these steps are supported by the system CONFERATOR[3]: It is in joint development by the School of Information Sciences, University of Pittsburgh (conference management component, as a refinement of the Conference Navigator [1]) and the Knowledge and Data Engineering group at the university of Kassel (social and ubiquitous PEERRADAR component).

A first prototype of CONFERATOR [2], developed by the Knowledge and Data Engineering group, has been successfully deployed at the LWA 2010 conference at the University of Kassel in October 2010. The system is based on the UBICON framework[4]: The PEERRADAR is applied for managing social and ubiquitous/real contacts, supported by embedding social networks such as Facebook, XING, and LinkedIn, while the TALKRADAR enables the personalization of the conference program. Furthermore, active RFID proximity-tags, cf., [3], provide for detecting the location and contacts between conference participants.

In CONFERATOR, privacy is a crucial issue: A variety of user data is collected. Therefore, appropriate steps for their secure storage and access were implemented. CONFERATOR implements privacy measures using a refined trust system: It features several privacy levels (private, trusted, public) for organizing access, e.g., to location or contact information. In the analysis, we aim at providing implicit k-anonymity in the presentation and discussion: We provide results at the level of special interest groups, or (for more detailed results) at groups containing at least four participants.

---

[3] http://www.conferator.org
[4] http://www.ubicon.eu

3 Related work

Regarding the tracking and analysis of conference participants, there have been several approaches, using RFID-tokens or Bluetooth-enabled devices. Hui et al. [4] describe an application using Bluetooth-based modules for collecting mobility patterns of conference participants. Furthermore, Eagle and Pentland [5] present an approach for collecting proximity and location information using Bluetooth-enabled mobile phones.

One of the first experiments using RFID tags to track the position of persons on room basis was conducted by Meriac et al. (cf., [6]) in the Jewish Museum Berlin in 2007. Cattuto et al. [7] added proximity sensing in the Sociopatterns project[5]. Isella et al. [8] conducted further experiments on a variety of contact networks obtained via RFID technology. Alani and colleagues, e.g., [3], also added contact information from social online networks in the live social semantics experiments. In [9], the authors analyze social dynamics of conferences focussing on the social activity of conference participants in those experiments. They analyze, for example, their activity in social web platforms like Facebook, Twitter and other social media together with status and their research seniority. These experiments have also extended their focus from conferences to schools [10] and hospitals [11].

Our work uses the same technical basis (RFID tokens with proximity sensing) as the Sociopatterns project, which allows us to verify their very interesting results independently. Furthermore, in this paper we significantly extend the analysis: We are able to use further techniques in order to characterize different roles, communities and participant relations. Concerning this, we especially focus on the community and role dynamics which are enabled by given gold-standard community structures according to a set of special interest groups. The conference navigator by Brusilovsky [1] allows researchers attending a conference to organize the conference schedule and provides a lot of interaction capabilities. However, it is not connected to the real live activity of the user during the conference. In our application, we measured face-to-face conctacts, increased the precision of the localization component compared to previous RFID-based approaches, and linked together tag information and the schedule of a workshop week. Furthermore, we implemented a light-weight integration with BibSonomy and other social systems used by participants. This is the basis for new insights into the behavior of all participants.

Thus, in comparison to the approaches mentioned above, we are able to perform a much more comprehensive evaluation of the patterns acquired in a conference setting, since our data provides a stable ground truth for communities (the special interest groups). These provide a grounding not only for the verification of the structural properties of the contact patterns, but also for the roles and communities.

Identifying different "roles" of nodes and finding so called "key actors" has attracted a lot of attention, ranging from different measures of centrality (cf., [12]) to the exploration of topological graph properties [13] or structural neighborhood similarities [14]. We focus on a metric that measures, how much a node connects different communities, cf., [15]. It can be based on initially given community structures or on a probabilistic model.

---

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

4 Grounding

In this section, we present an analysis of the collected conferencing data. After introducing some preliminaries, we first discuss a grounding of the communities given through the assignment of participants to special interest groups. After that, we consider roles concerning academic status (Prof., PostDoc, PhD-Student and Student) and roles concerning the conference organization (organizers and regular participants).

4.1 Preliminaries

In the following section, we briefly introduce basic notions, terms and measures used throughout this paper. We presume familiarity with the concepts of *directed* and *undirected* Graphs $G = (V, E)$ for a finite set V of nodes with edges $(u, v) \in V \times V$ and $\{u, v\} \subseteq V$ respectively. In a *weighted* graph, each edge is associated with a corresponding edge weight, typically given by a mapping from E to $\mathbb{R}$. We freely also use the term *network* as a synonym for a graph. For more details, we refer to standard literature, e.g., [16,17].

In the context of social network analysis, a *community* within a graph is defined as a group of nodes such that group members are densely connected among each other but sparsely connected to nodes outside the community [18]. Community structure was observed in several online social networks [19,20] and is sometimes also called "virtual community" [21].

For a given undirected graph $G = (V, E)$ and a community $C \subseteq V$ we use the following notation: $n := |V|$, $m := |E|$, $n_C := |C|$, $m_C := |\{\{u, v\} \in E : u, v \in C\}|$ – the number of *intra-edges* of C, and $\bar{m}_C := |\{\{u, v\} \in E : |\{u, v\} \cap C| = 1\}|$ – the number of *inter-edges* of C.

For formalizing and assessing community structure in networks, this work focuses on two prominent measures: The *modularity* [18] and the *segregation index* [22]. The modularity measure is based on a comparison of the number of edges within a community to the expected such number given a null-model (i.e., a randomized graph). Thus, the modularity of a (disjoint) community clustering is defined to be the fraction of the edges that fall within the given clusters minus the expected such fraction if edges were distributed at random. This can be formalized as follows: The modularity MOD of a graph clustering is given by

$$\text{MOD} = \frac{1}{2m} \sum_{i,j} \left(A_{i,j} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j) \,,$$

where A is the adjacency matrix, C_i and C_j are the clusters containing the nodes i and j respectively, k_i and k_j denote the degree of i and j. Further, $\delta(C_i, C_j)$ is the *Kronecker delta* symbol that equals 1 if $C_i = C_j$, and 0 otherwise.

A straightforward generalization of the above formula to a modularity measure wMOD in weighted networks [23] considers A_{ij} to be the weight of the edge between nodes i and nodes j, and replaces the degree k_i of a node i by its strength $\text{s}(i) = \sum_j A_{ij}$, i. e., the sum of the weights of the attached edges. For overlapping communities, the modularity can also be generalized as described in [24].

The *segregation index* SIDX [22] compares the number of expected inter-edges to the number of observed inter-edges, normalized by the expectation:

$$\text{SIDX}(C) = \frac{E(\bar{m}_C) - \bar{m}_C}{E(\bar{m}_C)} = 1 - \frac{\bar{m}_C n(n-1)}{2mn_C(n-n_C)} \ .$$

Thus, the segregation index explicitly includes the size n of the graph. By averaging the segregation over all clusters one obtains the segregation of a cluster model.

4.2 Available Data

For capturing social interactions, RFID proximity tags of the Sociopatterns project were handed out to the participants of the conference LWA 2010. 70 out of 100 participants volunteered to wear an RFID tag which (approximately) detected mutual face-to-face sightings among participants with a minimum proximity of about one meter. Each such sighting with a minimum length of 20 seconds was considered a contact which ended when the corresponding tags did not detect an according sighting for more than 60 seconds.

Using the recorded contact data, we generated undirected networks $\textsf{LWA}[\geq i]_*$, $\textsf{LWA}[\geq i]_\Sigma$, and $\textsf{LWA}[\geq i]_\#$. An edge $\{u, v\}$ is created, iff a contact with a duration of at least i *minutes* among participants u and v was detected ($i = 1, \ldots, 15$). For $i \geq 5$[minutes], for example, we can filter out "small talk" conversations. In $\textsf{LWA}[\geq i]_\#$ each edge is weighted with the *number* of according contacts, in $\textsf{LWA}[\geq i]_\Sigma$ it is weighted with the *sum* of all according contact durations, whereas $\textsf{LWA}[\geq i]_*$ is unweighted.

Table 1 contains some statistics for $\textsf{LWA}[\geq i]_*$, $i = 0, 5, 10$. The diameters and average path lengths coincide with those given in [8] for the Hypertext Conference 2009 (HT09). For $\textsf{LWA}[\geq 0]_*$, Figure 1 shows the degree and contact length distributions. The latter exhibits characteristics that are comparable with those given for HT09, whereas the degree distributions differ by exhibiting two peaks – one around 10 and one around 20 – in contrast to only one peak around 15 for HT09. We hypothesize that this deviation is due to a more pronounced influence of the conference organizers at LWA 2010 in relation to the total number of participants (approx. 15% of the participants in $\textsf{LWA}[\geq 0]_*$ were organizers). This hypothesis is supported by removing all organizers from $\textsf{LWA}[\geq 0]_*$ and recalculating the degree distribution, yielding a single peak in the interval $[15, 20)$.

Other statistics (e. g., strength distribution) also suggest evidence for structural similarities among HT09 and LWA 2010. Therefore, we conclude, that LWA 2010 was a typical technical conference setup, and results obtained at the LWA 2010 are expected to hold in other conference scenarios with similar size, too.

Furthermore, we extracted the "*visited talks*", i.e., the talks visited by each participant using the RFID information, resulting in 773 talk allocations for the conference participants.

4.3 Community Structure

LWA 2010 was a joint workshop week of four special interest groups of the German Computer Science Association (GI).

Table 1. High level statistics for different networks: Number of nodes and edges, Average degree, Average path length APL, diameter d, clustering coefficient C, number and size of the largest weakly connected component #CC and $|CC|_{max}$ respectively. Additionally the size of the contained special interest groups is given.

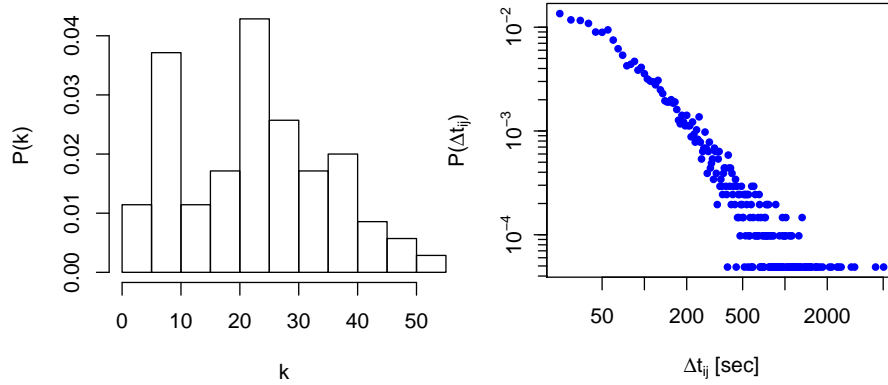| Network	$|V|$	$|E|$	Avg.Deg.	APL	d	Density	C	#CC	$|CC|_{max}$	$|KDML|$	$|WM|$	$|IR|$	$|ABIS|$
LWA[≥0]$_*$	70	812	23.20	1.72	4	0.34	0.55	1	70	37	16	10	7
LWA[≥5]$_*$	65	227	6.99	2.53	5	0.11	0.33	1	65	34	15	9	7
LWA[≥10]$_*$	56	109	3.89	3.09	7	0.07	0.31	3	50	31	12	7	6

Fig. 1. Degree distribution $P(k)$ (left) and distribution of the different contact durations (right) in LWA[≥0]$_*$.

- **ABIS** focuses on *personalization and user modeling*.
- **IR** is concerned with *information retrieval*.
- **KDML** focuses on all aspects of *knowledge discovery and machine learning*.
- **WM**, for 'Wissensmanagement', considers all aspects of *knowledge management*.

During the registration for LWA 2010, each participant declared his affiliation to exactly one special interest group: KDML (37), WM (16), ABIS (7), IR (10), for a total of 70 participants. Since these interest groups capture common research interests as well as personal acquaintance, the set of participants is naturally clustered accordingly.

As a first characteristic for the interest groups, we aggregated the visited talks groupwise per track. Although several sessions were joint sessions of two interest groups, Figure 2 clearly shows for each group a strong bias towards talks of the associated conference track.

Analysis of Community Dynamics For our analysis of community dynamics, we focused on the contact graph using different minimal contact durations: In this respect, we analyzed, if an according community structure may be observed in the respective contact graphs obtained during the conference. We considered the modularity measure and the segregation index for assessing the respective community allocations.

Table 2. Detail view on the distribution of the conference tracks of the talks visited by members of the different interest groups (numbers in percent). Members of the *ABIS* group, for example, visited ABIS talks in 58.8% of their visits, while they attended talks of the KDML track in 20.6%.

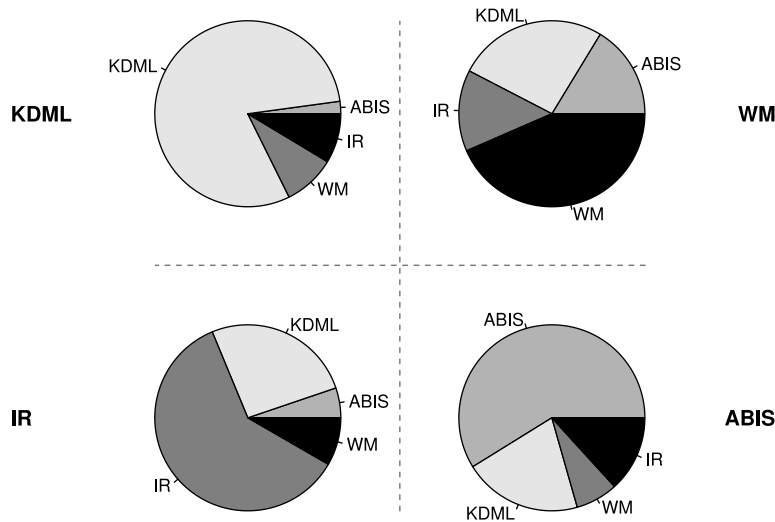|	ABIS	IR	KDML	WM
ABIS	**58.8**	13.2	20.6	7.4
IR	5.2	**60.4**	26.0	8.3
KDML	2.2	8.6	**80.2**	9.1
WM	16.3	14.1	26.1	**43.5**

Fig. 2. Distribution of the conference tracks of the talks visited by members of the different interest groups. The top-left figure, for example, shows the distribution of tracks visited by the KDML special interest group.

Figure 3 shows the obtained weighted and unweighted modularity scores for the contact graphs $\mathsf{LWA}[\geq i]_\Sigma$ and $\mathsf{LWA}[\geq i]_\#$ with $i = 1, \ldots, 15$, considering the interest groups as communities. The results suggest consistent community interaction according to the modularity in the corresponding networks. For analyzing the impact of repeated or longer conversations, we calculated the weighted modularity score on the same networks, using either the number of conversations or the aggregated conversation time between two participants as edge weights. Figure 3 shows that the obtained modularity scores are nearly constant across the different networks. This is a first hint for assuming that peers tend to talk more frequently and longer within their associated interest groups. As we will see below, this hypothesis is confirmed by the segregation index analysis.

To rule out statistical effects induced by structural properties of the contact graphs, we created a null model by repeatedly shuffling the group membership of all participants and averaging the resulting unweighted modularity scores. As Figure 3 shows, the shuffled group allocation shows no community structure in terms of modularity as expected.
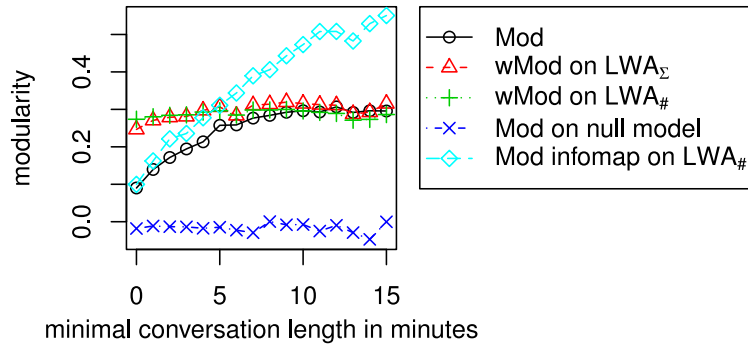
Fig. 3. Modularity scores for varying minimum conversation lengths, comparing the communities induced by the special interest groups to a null model, and to mined communities using Infomap.
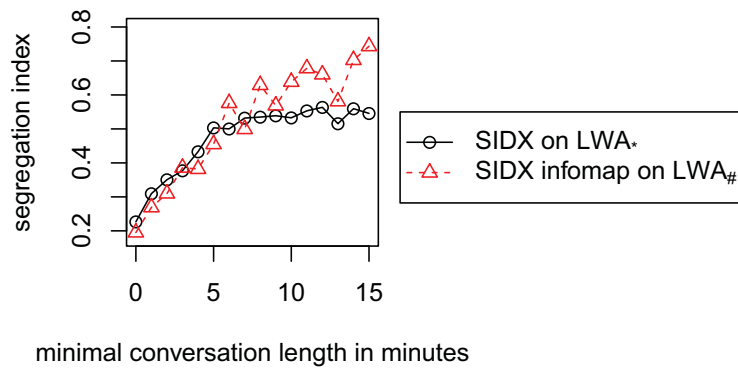

Fig. 4. Segregation index scores for varying minimum conversation lengths, comparing communities induced by the special interest groups to mined communities using Infomap.

Furthermore, we computed the segregation index for the communities on the unweighted graphs in order to exclude effects due to different network sizes. These could be introduced when filtering edges and nodes when increasing the minimum conversation length, cf., Table 1. As shown in Figure 4, the segregation index indicates a strongly ascending trend with increasing minimal conversation length concerning the corresponding networks and community allocations. This conforms to the intuition that more relevant (i. e., longer) conversations are biased towards dialog partners with common interests, as captured by the interest group membership.

Automatic Community Detection Additionally, the standard community detection algorithm Infomap [25], which is shown to perform well [26], was chosen for reference and applied to the same contact graphs. Figure 3 also shows the unweighted modularity scores for the obtained (disjoint) communities. At a minimum conversation length of 5 minutes, it almost coincides with the modularity scores of the interest group community

structure, and it nearly doubles both weighted and unweighted modularity scores in $\mathsf{LWA}[\geq 15]_{\Sigma}$ and $\mathsf{LWA}[\geq 15]_{\#}$.

Inspection of the obtained communities suggests that the applied algorithm yields communities in $\mathsf{LWA}[\geq 0]_{*}$ which are similar to the given interest groups but also more specialized (i. e., sub communities) in $\mathsf{LWA}[\geq i]_{*}$, $i \geq 5$. Figure 5 shows for reference in $\mathsf{LWA}[\geq 0]_{*}$, that ABIS and IR are nearly perfectly captured by Infomap but KDML is split mainly across two communities, one of which shared with WM.

It is important to note in this context, that we do not aim at evaluating any community detection algorithm: We rather exemplify the application of such algorithms and approximate an upper bound of the modularity score in the contact graphs.
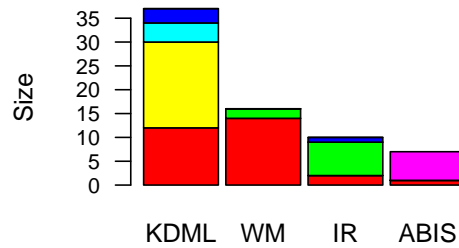

Fig. 5. Distribution of the interest groups across the six communities mined by the Infomap algorithm on $\mathsf{LWA}[\geq 0]_{*}$. Each color corresponds to a single (non-overlapping) community.

Similar findings are obtained by applying the MOSES algorithm [27] for detecting overlapping communities. Figure 6 shows exemplary community detection results for minimum conversation lengths ($min = 0, 1, 5, 10$) using the MOSES algorithm.

The community results in Figure 6 suggest, that the communities tend to focus more on the special interest groups with an increasing minimum conversation length threshold: The communities start with a mixture of different interest groups ($\mathsf{LWA}[\geq 0]_{*}$), but concentrate more and more on special sub-communities. These findings further confirm the initial results of Infomap and demonstrate the trend even more clearly, that more specialized sub-communities of the special interest groups are mined. This is especially significant for higher minimal conversation length thresholds, e. g., $\mathsf{LWA}[\geq 10]_{*}$, see Figure 6. In summary, this community-oriented analysis suggests, that participants actually tend to interact more frequently with members of their own special interest group for longer conversations.

Community Densities For analyzing social interactions across different interest groups we consider the density in the corresponding induced sub graphs.

Table 3 shows the density for each pair of interest groups $V_i, V_j \subseteq V$ in the complete contact graph $G = (V, E)$, the fraction of all actually realized edges in the set of possible edges between V_i and V_j.

Within the interest groups, the density values are strictly above the global density (cf., Table 1), but strictly below across different groups. This suggests that participants actually tend to interact more frequently with members of their own interest group.
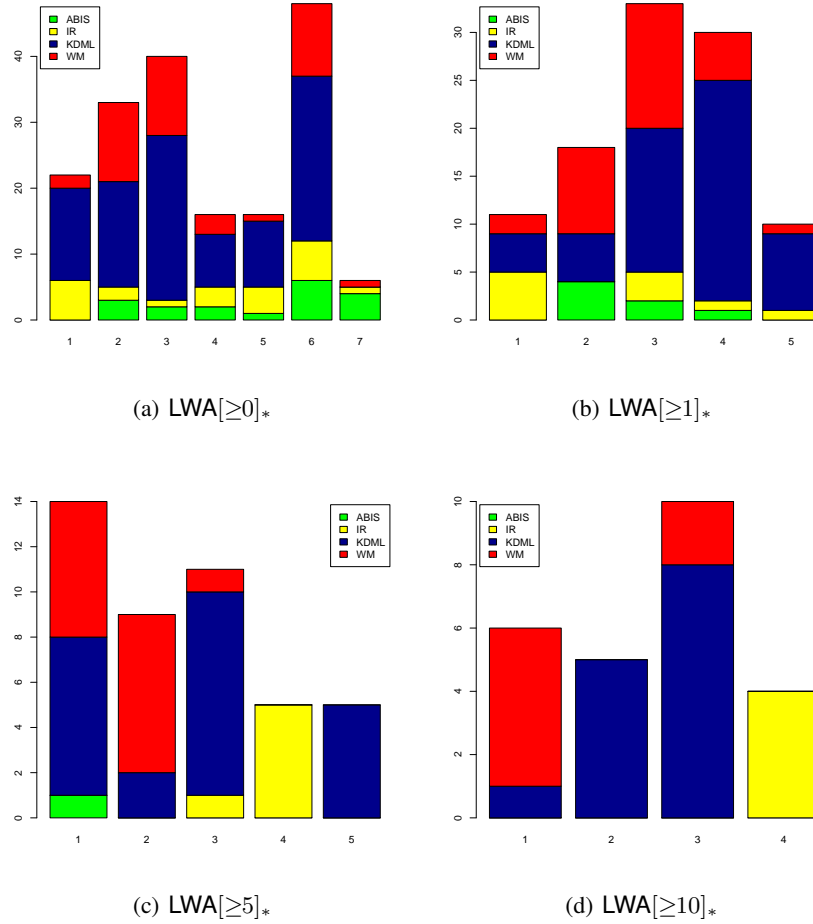
(a) LWA[≥0]∗

(b) LWA[≥1]∗

(c) LWA[≥5]∗

(d) LWA[≥10]∗

Fig. 6. Exemplary community detection results (overlapping communities) using the MOSES [27] algorithm. The different communities are colored according to their special interest track distributions.

Poster Session As part of the conference, a joint poster session with accompanied buffet was scheduled. Typically, there is a lot of interaction at a poster session, leading

Table 3. Density in the the contact graph LWA[≥0]∗.

	ABIS	IR	KDML	WM
ABIS	**0.62**	0.23	0.19	0.28
IR	0.23	**0.44**	0.21	0.20
KDML	0.19	0.21	**0.38**	0.31
WM	0.28	0.20	0.31	**0.58**

to many contacts. Similar to the global inter-group densities in Table 3 the according inter-group densities for the poster session are shown in Table 4. Since the contacts of the poster session are a subset of all contacts for the whole conference, the density values are lower than those for the global densities.

However, we observe the same trends as for the global density, namely, that the intra-group densities are larger than the inter-group conversations – with one exception, i.e., the KDML special interest group. One possible explanation is, that KDML was the largest group present, and therefore the chances for increasing their inter-density shares with other special interest groups are better. Additionally, the large share of the density values compared to the global ones seem to indicate the impact and importance of the poster session for networking.

Table 4. Density in the complete contact graph of the poster session.

	ABIS	IR	KDML	WM
ABIS	**0.19**	0.06	0.05	0.08
IR	0.06	**0.24**	0.07	0.08
KDML	0.05	0.07	0.10	**0.12**
WM	0.08	0.08	0.12	**0.36**

Discussion In summary, our results exemplify, that community structure as induced by interest groups in conferences can be measured in face-to-face communication networks quite well. Furthermore, standard community detection algorithms show promising results for mining user communities from social interaction patterns in both on-line and face-to-face communication.

Typically, conference conversations tend to cluster around interest groups. For the LWA we could observe this effect for the present special interest groups. Intelligent analysis and mining techniques can then help, for example, for establishing more dynamic conference events.

4.4 Roles and Key Players

The assignment of roles to nodes in a network is a classification process that categorizes the players by common patterns. In this section, we discuss the connection between the academic status of the conference participants and the classic centrality measures and a community based role assignment. The latter was introduced in [28] together with the $rawComm$ measure that it is based on. The measure is defined as:

$$rawComm(u) = \sum_{v \in N(u)} \tau_u(v),$$

where the contribution of a node v in the neighborhood of u is given by:

$$\tau_u(v) = \frac{1}{1 + \sum_{v' \in N(u)} \left(I(v, v') * p + \bar{I}(v, v')(1 - q) \right)}.$$

Table 5. Group size and average graph centralities per academic position and for organizers and non-organizers in LWA[≥ 5]$_*$: degree deg, strengths $str_\#$ and str_Σ, eigenvalue centralities eig_*, $eig_\#$ and eig_Σ, betweenness bet, closeness clo and the average community metric $rawComm$.

Position	size	deg	$str_\#$	str_Σ	eig_*	$eig_\#$	eig_Σ	bet	clo	$rawComm$
Prof.	10	7.500	16.700	11893.200	0.310	0.285	0.337	49.565	0.407	0.525
PostDoc	11	7.727	15.545	9793.364	0.303	0.213	0.198	75.973	0.419	0.675
PhD-student	33	7.152	15.091	9357.182	0.309	0.201	0.165	46.221	0.409	0.567
Student	5	3.600	12.400	6514.400	0.099	0.068	0.027	17.989	0.347	0.417
Other	6	6.667	14.333	8920.000	0.288	0.211	0.209	38.234	0.413	0.581
Organizer	11	10.000	23.727	15227.545	0.459	0.424	0.417	94.497	0.447	0.699
Non-Organizer	54	6.370	13.389	8408.056	0.256	0.162	0.144	39.565	0.397	0.542

In the formulas $N(u)$ is the neighborhood of a node u; $I(v, v') = 1$ if there is an edge between v and v' and 0 else; $\bar{I} = 1 - I$, p is the probability that an edge in the graph connects two communities and q is the probability that two non-linked nodes are in different communities. The $rawComm$ score can be interpreted as a measure of how much a node connects different communities (for details, cf., [28]). To estimate the probabilities p and q, we applied the method suggested in [28] using the contact graph and the special interest groups as grounding. The measure utilizes the (dynamically constructed) contact graphs since we focus on the time-dependent analysis of communities and roles, and the implied communication and interaction structure.

Although we considered a set of given communities as grounding (the special interest groups), it becomes clear by considering the results of the Infomap algorithm for automatic community detection (see Section 4.3) that some communities are subsets of the special interest groups, while others transcend the bounds of these groups. We therefore have chosen the probabilistic approach and used the clustering given through the special interest groups only for estimating the parameters p and q.

Global Characterization Table 5 displays the average values of several graph node centralities of LWA[≥ 5]$_*$ aggregated by academic position and for the conference organizers and non-organizers (regular conference participants), respectively. Note, that while the categories referring to academic status are disjoint (the category *other* includes all participants that do not fit one of the other four) organizers and non-organizers both include participants from all the 'status' categories.

A first observation is that the organizers have significantly higher scores in all nine measures under observation. In the considered conference scenario this is highly plausible due to the nature of an organizer's job during a conference – which in the case of LWA 2010 also included the supervision and maintenance of the RFID-experiment and the CONFERATOR. Among the four academic positions, striking differences can be noticed. First of all, the student scores in all centralities are lower than those of the other categories. We attribute this phenomenon to the fact, that students are less established in their scientific communities than scientists in higher academic positions

and usually have little conference experience. This example motivates the use of social tools that could assist participants in initiating contact to their communities and persons of interest.

Within the categories "Prof.", "PostDoc" and "PhD-student" the eigenvalue centralities show a particular behavior. While the unweighted eigenvalue centrality eig_* does not fluctuate much, the weighted versions eig_Σ and $eig_\#$ increase strongly from one position to the next higher one. Eigenvalue centralities are considered a measure of importance. It seems plausible, that in a contact graph among scientists, the players with longer scientific experience – including a higher and broader degree of knowledge within scientific areas and more previous contacts and collaborations with their colleagues – are considered more important and that this attitude is reflected in their contacts. The node strength measures show similar results. While the degree deg is only slightly different among the three positions, the weighted versions $str_\#$ and especially str_Σ show large differences and increase together with the position. The considerable difference between the weighted and unweighted measures indicates the relevance of the frequency and the length of the contacts: Professors, for example, have longer and more contacts to other participants than PostDocs.

Another aspect is illustrated by the betweenness (bet) scores: Relatively to the other groups, a lot of shortest paths of $\mathsf{LWA}[\geq 5]_*$ run through nodes of PostDocs. We attribute this to the structure of scientific institutes, where usually one professor supervises several PostDocs who again each supervise several PhD-students. PostDocs are the connection between professors and the postgraduates and thus assume the role of gatekeepers in their working environment.

Explicit Roles For the $rawComm$ measure it is harder to come up with a plausible explanation for the different values and the order of the academic positions. However, as described in [28], it can be combined with $ndeg$ – the degree divided by the maximum degree – to gain a role classification for the network's nodes in the following way: One out of four roles is assigned to a node v according to

$$
role(v) := \begin{cases}
\text{Ambassador} & ndeg(v) \geq s, rawComm(v) \geq t \\
\text{Big Fish} & ndeg(v) \geq s, rawComm(v) \lessgtr t \\
\text{Bridge} & ndeg(v) \lessgtr s, rawComm(v) \geq t \\
\text{Loner} & ndeg(v) \lessgtr s, rawComm(v) \lessgtr t
\end{cases}
$$

where s and t are thresholds that we chose as $s = t = 0.5$ – the same choice as in [28].

Ambassadors are characterized by high scores in both degree and $rawComm$ which means that they connect many communities in the graph. A *Big Fish* has contacts to a lot of other nodes, however, mostly within the same community. *Bridges* connect communities, however, not as many as ambassadors. Finally, *Loners* are those with low scores in both measures.

In the following, we investigate how nodes in their explicitly given roles like the academic position and the job (organizer) fill those implicitly given graph structure-based roles. Therefore, we applied the role classifier to the graphs $\mathsf{LWA}[\geq 0]_*$ through $\mathsf{LWA}[\geq 15]_*$ to determine – under the assumption, that longer contacts indicate more serious and scientific discussions – how this changes the community roles.
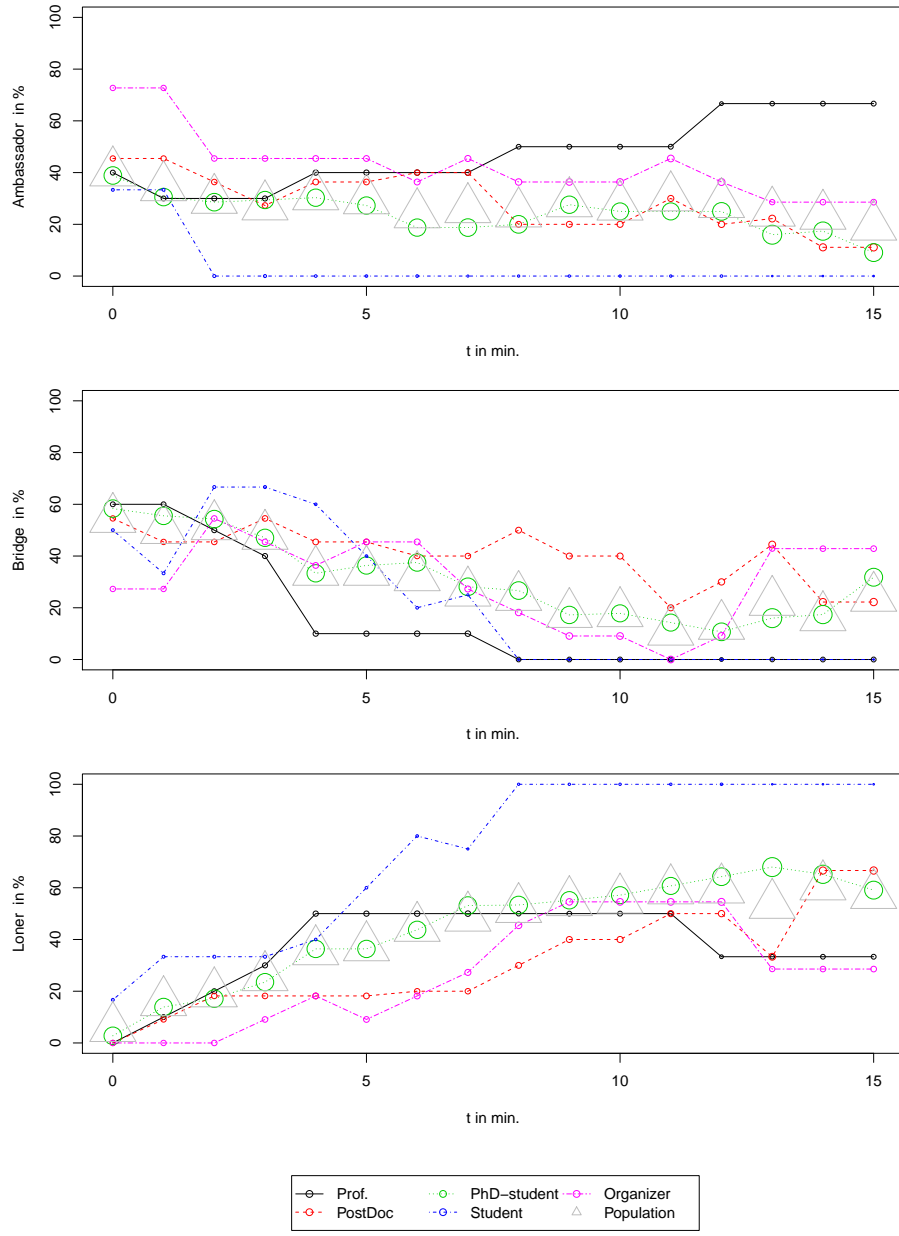
Fig. 7. Fraction of participants that assume the roles Ambassador, Bridge and Loner.

The first immediate finding is, that in none of the graphs any participant was ever classified as Big Fish, i. e., whenever a node has a high degree it also has a high *rawComm* score. We attribute this peculiarity to the fact, that the very nature of social interaction at conferences usually is exchanging ideas with participants outside the own peer group. Especially during the LWA 2010, participants were encouraged to engage in interdisciplinary dialogue for example by including several joint sessions in the schedule and a combined event of social dinner and poster session.

The first of the three diagrams in Figure 7 displays the percentage of participants with a common academic position or job that were classified as Ambassador. The line marked with triangles displays that fraction of all participants together. The second and third diagram display the same fractions for the roles Bridge and Loner. For example in $\mathsf{LWA}[\geq 0]_*$, 40% of the professors were classified as Ambassador, 60% as Bridge and 0% as Loner.

In each diagram the size of the nodes indicates the size of the group of participants with the examined position/job in the respective graph. The PhD-students, for example, are the largest section, while the students form the smallest. For $\mathsf{LWA}[\geq 5]_*$ those sizes are given in Table 5. While all curves in Figure 7 fluctuate, there are several clearly visible tendencies. In all three diagrams, the fractions of PhD-students is very close to the fraction of all participants. The simple reason for that is, that PhD-students are the majority within the conference population and therefore dominate the general behavior.

Many of the organizers start out as Ambassador or Bridge. This is again consistent with their job description. However, filtering out short contacts and thus the typical quick organizational conversations, the relevance of the organizers decreases with a higher limit to the minimum contact length. More and more organizers become Loners; in the last graph $\mathsf{LWA}[\geq 15]_*$, they are almost equally distributed among the three roles. One should keep in mind, that the group of organizers contains persons in all academic positions. Therefore, after filtering out most of the contacts that presumably are due to their organizational work, the organizers act mainly in their respective role as conference participants, which might explain the stronger fluctuations in the right part of the curve.

Very consistent with the findings described above is the role distribution among the students. While in the first graphs, where short contacts dominate the longer ones, some of them are classified as Bridge or Ambassador, they quickly disappear from those roles and are classified as Loner. Compared to the PhD-students, the fractions of the PostDocs are with few exceptions higher for the roles Ambassador and Bridge and lower for Loner. This is again consistent with the previous observations concerning the graph structure measures. Due to their greater experience PostDocs seem to have more access to colleagues in other communities. However, with the increasing filter limit, like most of the participants they become classified as loners. Finally, the curve of the professors in the role Ambassador shows the largest deviation from the mainstream. While in that role all other group's fractions decrease, that of the professors increases significantly up to 70% which is far more than any of the other academic positions.

In summary, we observe, that the chosen method of role assignment seems to be highly correlated to the roles like academic position and the organizer job.

Table 6. Role = Non-Organizer / Position = PhD-student for the aggregated count information with an aggregated contact length $\geq$ 5 min. The tables show the lift of the pattern comparing the fraction of non-organizers / PhD-students covered by the pattern p compared to the fraction of the whole dataset, the size of the pattern extension (number of described non-organizers / PhD-students), and the description itself.

target	#	lift	p	size	description
Non-Organizer	1	1.06	0.88	51	$clo=\{low; medium\}$
	2	1.05	0.87	61	$eig_*=\{low; medium\}$
	3	1.04	0.86	59	$deg=\{low; medium\}$
	4	1.10	0.92	12	$clo=\{low; medium\}$ AND $deg=\{high; medium\}$
	5	1.12	0.93	30	$clo=\{high; low\}$ AND $eig_*=\{low; medium\}$
PhD-student	1	1.07	0.54	59	bet=$\{high; low\}$
	2	1.07	0.54	48	s=$\{high; low\}$
	3	1.14	0.58	26	$deg=high$
	4	1.31	0.67	12	$bet=\{high; low\}$ AND $eig_*=high$
	5	1.38	0.70	20	$deg=high$ AND $bet=\{high; low\}$
	6	1.58	0.80	10	$deg=high$ AND $bet=\{high; low\}$ AND $eig_*=\{high; low\}$

Characterization of Explicit Roles In the following, we aim to characterize the roles in more detail; for the dataset, we focus on the majority classes, i.e., we consider the target concept *non-organizer* concerning roles, and the target concept *PhD-students* concerning academic position. For the analysis, we applied a method for mining characteristic patterns [29] based on subgroup discovery techniques, e.g., [30,31,32]: Basically, we aim to discover subgroups of participants described by combinations of factors, e.g., *deg=high AND bet=high* that show a high share of a certain target property, e.g., with respect to organizing or academic roles. For the data preprocessing, we first discretized the numeric features described above into three intervals (low, medium, high) using equal-width discretization.

The most descriptive factors for the role *non-organizer* are shown in Table 6 (upper). They confirm the averaged results shown above, in that the most characteristic *single* factors are given by the closeness, eigenvalue centrality, and the degree of the non-organizers, for which lower values than those of the organizers are measured. However, if we consider combinations of factors, we observe, that there are subgroups regarding the role non-organizer for which extreme values, e.g., of the closeness together with the eigenvalue centrality yield a significant increase in characterization power, as shown by the quality increase in Table 6.

If we consider the largest group *PhD-student* (concerning the academic positions), we observe the single factors shown in Table 6 (lower), also confirming the averaged results presented above. Similarly to the non-organizers, we see that extreme values, i.e., sets of high and low values, are also very significant for distinguishing PhD students. As expected the combination with other strong influence factors increases the precision of the patterns (indicated by the *lift* parameter).

5 Conclusions

In this paper, we have presented results of an in-depth analysis of the dynamics of community structure and roles of face-to-face contacts during a conference. We have performed various analyses on data collected during the LWA 2010 in Kassel in October 2010 by using a social conference guiding system. We analyzed and described high-level statistics of the collected network data, examined the different communities, the roles and key players concerning these and the conference in total. The analysis was grounded using real-world conference data capturing community information about participants (membership in special interest groups). It was shown that there is detectable community structure in the investigated face-to-face networks that is consistent with the one given through the groups. Further, the structural properties of the contact graphs obtained at the LWA conference reflect different aspects of community interactions and roles. For the latter, we provided characteristic (role) profiles and detected traces of different behavioral patterns for different roles.

As future work, we plan to continue the community related investigations further, since communities play a central role for a social conferencing system and should allow and support emergence and evolution of community structure. Furthermore, identifying key actors according to their roles is an interesting task; results could be used e.g., for creating virtual sessions or recommendations. Additionally, we aim to collect more data in order to generalize and compare the mined profiles and patterns on a broader basis, e.g., concerning more diverse conferences. A first step in this direction has been taken in [33], continuing these analyses seems a very promising task.

Acknowledgements

References

1. Wongchokprasitti, C., Brusilovsky, P., Para, D.: Conference Navigator 2.0: Community-Based Recommendation for Academic Conferences. In: Proc. Workshop Social Recommender Systems, IUI'10. (2010)
2. Atzmueller, M., Benz, D., Doerfel, S., Hotho, A., Jäschke, R., Macek, B.E., Mitzlaff, F., Scholz, C., Stumme, G.: Enhancing Social Interactions at Conferences. it - Information Technology **53**(3) (2011) 101–107
3. Alani, H., Szomszor, M., Cattuto, C., den Broeck, W.V., Correndo, G., Barrat, A.: Live Social Semantics. In: Proc. Intl. Sem. Web Conf. (2009) 698–714

4. Hui, P., Chaintreau, A., Scott, J., Gass, R., Crowcroft, J., Diot, C.: Pocket Switched Networks and Human Mobility in Conference Environments. In: Proc. 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking. WDTN '05, New York, NY, USA, ACM (2005) 244–251

5. Eagle, N., Pentland, A.S.: Reality Mining: Sensing Complex Social Systems. Personal Ubiquitous Comput. **10** (March 2006) 255–268

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

7. Cattuto, C., den Broeck, W.V., Barrat, A., Colizza, V., Pinton, J.F., Vespignani, A.: Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. PLoS ONE **5**(7) (07 2010)

8. Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.F., den Broeck, W.V.: What's in a crowd? analysis of face-to-face behavioral networks. Journal of Theoretical Biology **271**(1) (2011) 166 – 180

9. Barrat, A., Cattuto, C., Szomszor, M., den Broeck, W.V., Alani, H.: Social dynamics in conferences: Analyses of data from the live social semantics application. In: Proc. Intl. Semantic Web Conference. Volume 6497 of Lecture Notes in Computer Science. (2010) 17–33

10. Stehle, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.F., Quaggiotto, M., den Broeck, W.V., Regis, C., Lina, B., Vanhems, P.: High-resolution measurements of face-to-face contact patterns in a primary school. CoRR **abs/1109.1015** (2011)

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

12. Brandes, U., Erlebach, T., eds.: Network Analysis: Methodological Foundations. In Brandes, U., Erlebach, T., eds.: Network Analysis. Volume 3418 of Lecture Notes in Computer Science., Springer (2005)

13. Chou, B.H., Suzuki, E.: Discovering Community-Oriented Roles of Nodes in a Social Network. In: Data Warehousing and Knowledge Discovery. Volume 6263 of Lecture Notes in Computer Science. Springer, Berlin / Heidelberg (2010) 52–64

14. Lerner, J.: Role assignments. In Brandes, U., Erlebach, T., eds.: Network Analysis. Volume 3418 of Lecture Notes in Computer Science. Springer, Berlin / Heidelberg (2005) 216–252

15. Scripps, J., Tan, P.N., Esfahanian, A.H.: Node Roles and Community Structure in Networks. In: Proc. 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web mining and Social Network Analysis, New York, NY, USA, ACM (2007) 26–35

16. Diestel, R.: Graph theory. Springer, Berlin (2006)

17. Gaertler, M.: Clustering. [12] 178–215

18. Newman, M.E., Girvan, M.: Finding and Evaluating Community Structure in Networks. Phys Rev E Stat Nonlin Soft Matter Phys **69**(2) (2004) 026113.1–15

19. Mitzlaff, F., Benz, D., Stumme, G., Hotho, A.: Visit Me, Click Me, Be My Friend: An Analysis of Evidence Networks of User Relationships in Bibsonomy. In: Proceedings of the 21st ACM conference on Hypertext and Hypermedia, Toronto, Canada (2010)

20. Leskovec, J., Lang, K.J., Mahoney, M.W.: Empirical Comparison of Algorithms for Network Community Detection (2010) cite arxiv:1004.3539.

21. Chin, A., Chignell, M.: Identifying Communities in Blogs: Roles for Social Network Analysis and Survey Instruments. Int. J. Web Based Communities **3** (June 2007) 345–363

22. Freeman, L.: Segregation In Social Networks. Sociological Methods & Research **6**(4) (1978) 411

23. Newman, M.E.J.: Analysis of Weighted Networks (2004) http://arxiv.org/abs/cond-mat/0407503.
24. Nicosia, V., Mangioni, G., Carchiolo, V., Malgeri, M.: Extending the Definition of Modularity to Directed Graphs with Overlapping Communities. J. Stat. Mech. (2009)
25. Rosvall, M., Bergstrom, C.: An Information-theoretic Framework for Resolving Community Structure in Complex Networks. Proc. Natl. Acad. of Sciences **104**(18) (2007) 7327
26. Lancichinetti, A., Fortunato, S.: Community Detection Algorithms: A Comparative Analysis (2009) arxiv:0908.1062.
27. McDaid, A., Hurley, N.: Detecting highly overlapping communities with model-based overlapping seed expansion. In: Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining. ASONAM '10, Washington, DC, USA, IEEE Computer Society (2010) 112–119
28. Scripps, J., Tan, P.N., Esfahanian, A.H.: Exploration of Link Structure and Community-Based Node Roles in Network Analysis. In: ICDM. (2007) 649–654
29. Atzmueller, M., Lemmerich, F., Krause, B., Hotho, A.: Who are the Spammers? Understandable Local Patterns for Concept Description. In: Proc. 7th Conference on Computer Methods and Systems. (2009)
30. Wrobel, S.: An Algorithm for Multi-Relational Discovery of Subgroups. In: Proc. 1st European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD-97). (1997) 78–87
31. Atzmueller, M., Puppe, F., Buscher, H.P.: Exploiting Background Knowledge for Knowledge-Intensive Subgroup Discovery. In: Proc. 19th Intl. Joint Conference on Artificial Intelligence (IJCAI-05), Edinburgh, Scotland (2005) 647–652
32. Atzmueller, M., Puppe, F.: SD-Map - A Fast Algorithm for Exhaustive Subgroup Discovery. In: Proc. 10th European Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD 2006). (2006) 6–17
33. Macek, B.E., Scholz, C., Atzmueller, M., Stumme, G.: Anatomy of a Conference. In: Proc. 23rd ACM Conference on Hypertext and Social Media, ACM Press (2012)