Data Mining on Social Interaction Networks

Martin Atzmueller

University of Kassel, Knowledge and Data Engineering Group,
Wilhelmshöher Allee 73, 34121 Kassel, Germany
atzmueller@cs.uni-kassel.de

Abstract

Social media and social networks have already woven themselves into the very fabric of everyday life. This results in a dramatic increase of social data capturing various relations between the users and their associated artifacts. In such settings, data mining and analysis plays a central role: Predictive data mining targets the acquisition and learning of specific models in order to support the users, e. g., for classification or inference of parameters for future cases. Furthermore, descriptive data mining aims at obtaining patterns which summarize and characterize the data.

From an application perspective, there is a variety of computational social systems – with an increasing use of mobile and ubiquitous technologies. The various direct and indirect interactions between the users in the online networks as well as the real-world human interactions using ubiquitous devices can then be represented using social interaction networks.

In this article, we consider social interaction networks from a data mining perspective – also with a special focus on real-world face-to-face contact networks: We combine data mining and social network analysis techniques for examining the networks in order to improve our understanding of the data, the modeled behavior, and its underlying emergent processes. Furthermore, we adapt, extend and apply known predictive data mining algorithms on social interaction networks. Additionally, we present novel methods for descriptive data mining for uncovering and extracting relations and patterns for hypothesis generation and exploration by the user, in order to provide characteristic information about the data and networks. The presented approaches and methods aim at extracting valuable knowledge for enhancing the understanding of the respective data, and for supporting the users of the respective systems. We consider data from several social systems, like the social bookmarking system BibSonomy, the social resource sharing system flickr, and ubiquitous social systems: Specifically, we focus on data from the social conference guidance system *Conferator* and the social group interaction system *MyGroup*.

This work first gives a short introduction into social interaction networks, before we describe several analysis results in the context of online social networks, as well as real-world face-to-face contact networks. Next, we present predictive data mining methods making use of the social interactions, i. e., for localization, recommendation and link prediction. After that, we present novel descriptive data mining methods for mining communities and patterns on social interaction networks.

1 Introduction

The emergence of new social systems and organizational social applications has created a number of novel social and ubiquitous environments. By interacting with such systems, the user is leaving traces within the different databases and log files, e. g., by updating the user's status via Twitter, commenting an image in Flickr, copying a post in BibSonomy, or connecting to other users via ubiquitous devices.

Ultimately, each type of such a trace gives rise to a corresponding network of user relatedness, where users are connected if they interacted either explicitly (e. g., by a direct encounter via an RFID tag, or by establishing "friendship" in an online social network) or implicitly (e. g., by visiting a user's profile page). We consider a link within such a network as evidence for user relatedness and interaction and call it accordingly *social interaction network*. This connects but also transcends private and business applications featuring a range of different types of networks, organizational contexts and corresponding interactions, e. g., networks that involve spatial proximity relations like co-location or face-to-face proximity. In this context, physical devices, e. g., mobile phones or RFID devices, can help to link relations in the digital domain to relations in physical and/or social space, and vice-versa. With the growth and availability of the collected data, there is also an increasing interest in the analysis of such social interaction networks.

This article considers data mining in social interaction networks, specifically human behavioral (offline) networks, that is, networks of face-to-face proximity, e. g., during a conversation. We call these networks *face-to-face contact networks* in the following. We include social (online) networks contructed from subject-centric or object-centric sociality [63], considering data from social bookmarking systems such as BibSonomy[1] [24] and resource sharing systems like Flickr[2], but especially focus on real-world face-to-face contact networks.

The capture of human face-to-face contacts in social interaction networks and the analysis of both offline and online data is receiving increasing interest. While there has been foundational work on the analysis of face-to-face contact networks, e. g., [22, 30, 117], data mining on those networks is still a rather new field of research, for which we provide novel analyses resulting in new insights into their structure and relations. In addition, we adapt, extend and apply known data mining methods to the collected social interaction networks, especially the face-to-face contact networks. Furthermore, we propose novel methods and approaches for describing and characterizing the networks and properties of their nodes, respectively. Integrated into different systems [14] and applications [9, 10], the techniques and methods are also deployed in a practical real-world setting.

This work, an adapted and substantially extended revision of [7], provides an overview on those previously published articles [10, 12, 15, 17, 71, 80, 81, 90, 91, 107, 108] as grouped together in the author's habilitation thesis [8].

[1] `http://www.bibsonomy.org`
[2] `http://www.flickr.com`

The face-to-face contact networks in our context are acquired using the CONFERATOR[3] [10] and MYGROUP[4] [9] systems. CONFERATOR, a social conference guidance system, and MYGROUP, a social workgroup support system, are ubiquitous social systems for enhancing social interactions in the context of conferences and working groups, respectively. They are built on top of the RFID-based proximity sensing hardware developed by the SocioPatterns[5] collaboration, and on top of the UBICON software platform[6] [9]. Both systems allow the collection of real-world networks of human face-to-face interactions – as behavioral networks – as well as the utilization of additional online social networks. Since these are our own systems, we were able to comprehensively perform our experiments using all the available data.

Data mining provides approaches for the identification and discovery of non-trivial patterns and models hidden in large collections of data [6, 39, 48]. While there exist several process models for data mining [64], a prominent model is given by the CRISP-DM process [31, 130] – an industry standard for data mining. CRISP-DM consists of six phases: (1) *Business Understanding* (2) *Data Understanding* (3) *Data Preparation* (4) *Modeling* (5) *Evaluation*, and (6) *Deployment*.

We combine data mining and social network analysis techniques for analyzing social interaction networks in order to improve our *understanding* of the data, the modeled behavior, and its underlying processes, in Section 3. This corresponds to the business and data understanding phases in the CRISP-DM process. Furthermore, for the analysis itself we can also apply techniques for explorative analysis discussed below.

Next, we focus on elements of the *modeling* phase, i. e., the core data mining step in Sections 4-5: The applied data mining methods can be divided into descriptive and predictive methods [48]: While descriptive methods are used for summarizing the data, for identifying hidden information in the form of patterns, and for exploration, predictive methods are used for constructing models for inferring future properties given new data, e. g., for classification: We adapt and extend known predictive data mining algorithms on social interaction networks for supporting the users in typical tasks such as recommendation and localization in the context of the mentioned systems. Additionally, we present novel methods for descriptive data mining for uncovering and extracting relations and patterns for hypothesis generation and exploration by the user, in order to provide characteristic information about the data and networks.

The methods and criteria in the *evaluation* phase depend on the applied model type, and are therefore specific for a data mining technique. Therefore, we discuss these in the respective subsections of Sections 4-5. In addition, Section 3.1 describes a novel community assessment technique, while Section 5.3 summarizes interactive techniques for the evaluation of mined patterns in social interaction networks. Finally, the *deployment* phase is tackled by the integration of the methods into practical applications, e. g., into CONFERATOR and MYGROUP as described in Section 2.3. Furthermore, especially the descriptive data mining methods have been integrated into the pattern mining and analytics system VIKAMINE [7, 14].

[3] http://www.conferator.org
[4] http://ubicon.eu/about/mygroup
[5] http://www.sociopatterns.org
[6] http://www.ubicon.eu

Related Work. Overall, data mining in the context of social interaction networks concerns core elements of data mining and knowledge discovery itself, e. g., [48], but also includes techniques from social network analysis, e. g., [114,124], as well as mining social media, e. g., [104, 120], complex network analytics [5, 27, 88, 97, 118], and mining the ubiquitous web [53, 112, 134].

Specifically, community detection [66,98], analysis of roles [110,111], contact patterns [30,56], localization [51,86,101], recommendations [19,28,57,113], link prediction [20,77,122], descriptive pattern mining [103,126], exceptional model mining [34, 62,70], but also techniques for reality mining, e. g., [35,89] are prominent topics in this area. Especially for face-to-face contact networks, Cattuto et colleagues [21,22,30,56] provide an overview on social dynamics in those networks. Their experiments include applications and analysis at conferences [4,22,56,119], schools [116], and in epidemiology [105,117]. In the following sections of this article, we will discuss related work in the respective sections in more detail.

Structure of this Article. The remainder of this work is structured as follows: Section 2 outlines basics of social interaction networks, real-world face-to-face contact networks, and describes the CONFERATOR and MYGROUP systems. Section 3 summarizes several analysis directions concerning interrelations in social interaction networks, communities and roles in human face-to-face contact networks, and structure and dynamics of interactions of conference participants. Next, Section 4 discusses adaptations, extensions, and applications of predictive methods in social interaction networks. After that, Section 5 presents novel efficient descriptive methods for community mining and exceptional model mining, as well as an exploratory pattern mining approach on social interaction networks. Finally, Section 6 concludes with a summary and outlook.

2 Basics of Social Interaction Networks

In this section, we focus on basics of social interaction networks. We first provide an introduction to social interaction networks. Next, we focus on human face-to-face contact networks. After that, we present the CONFERATOR and MYGROUP systems.

2.1 A Brief Introduction to Social Interaction Networks

With the rise of social software and social media, a wealth of user-generated data and user interactions is being created in online social networks [50] – as a technical platform. Often, these are also called (online) social network services, cf. [3]. In the following, we adopt an intuitive definition of social media, regarding it as online systems and services in the ubiquitous web, which create and provide social data generated by human interaction and communication [7].

A *social network* is a core concept of social network analysis [109,124]: According to Wassermann and Faust [124], a social network is a social structure consisting of a set of actors (such as individuals or organizations) and a set of dyadic ties between these actors. There are various types of ties such as, for example, friendship, kinship, or

organizational position. Usually, these relations are modeled as a graph, with the actors as nodes, and the ties as edges connecting the nodes.

In this article, we focus on *social interaction networks* [90–93], i. e., user-related social networks in social media capturing social relations inherent in social interactions, social activities and other social phenomena which act as proxies for social user-relatedness. Therefore, according to the categorization of Wassermann and Faust [124, p. 37 ff.] social interaction networks focus on *interaction* relations between *people* as the corresponding actors. This also includes interaction data from sensors and mobile devices, as long as the data is created by real users. Consider, for example, users who connect their mobile phones via Bluetooth or NFC (Near Field Communication), look at similar pictures in Flickr, talk about similar topics in Twitter, or explicitly establish "contacts" within certain applications. Furthermore, we consider real-world contacts as determined by other ubiquitous applications, based on principles of ubiquitous computing [127, 128] or the ubiquitous web [53, 112, 134].

Networks in ubiquitous and social applications, for example, in RFID-based systems, can be derived according to the detected contacts between the respective RFID tags: A tie between two actors can be derived, for example, if their assigned RFID tags were in contact, possibly weighted with the number of contacts, or their durations. Such social interation networks are often observed during certain events, for example, during conferences [10, 81, 133, 136], at work [9, 80], or other group-based activities. In Section 2.2, we specifically discuss human face-to-face contact networks. Furthermore, we summarize according applications in Section 2.3.

Overall, we consider social interactions in an online and offline context, that is, connections and relations in online systems as well as real-world face-to-face contacts. Furthermore, we also consider social relations implemented using specific resources or artifacts, according to the principle of object-centric sociality [63], where objects of a specific actor, e. g., resources, mediate connections to other actors. This is especially relevant in the scope of social bookmarking and social resource sharing systems. Examples are given by the "Favorite" relation in Flickr [87], or resource - resource relations, e. g., by a common set of tagged images. This also transcends to further collaborative systems, for example, data of versioning systems like CVS [80].

Typical analytical questions concern the analysis of key actors, roles and communities, ranging from different measures of centrality, cf. [25] to the exploration of topological graph properties, e. g., [42] or structural neighborhood similarities, e. g., [72]. The analysis of communities intuitively considers densely connected subgroups of actors, represented as nodes in the social network, e. g., [40, 45, 66, 99]. The detection and characterization of communities is a typical descriptive data mining task [17], as well as the description of patterns given properties of the nodes, e. g., roles, tagging information, or common geo-location. Predictive approaches include, for example, link prediction techniques, as well as recommendation, or utilizing the network for improving localization methods. Both descriptive methods and predictive modeling approaches will be discussed in Sections 4 and 5, respectively, below.

2.2 Real-World Interactions: Face-to-Face Contact Networks

In contrast to online interactions, face-to-face contacts represent interactions between actors in the real-world. Consider conference contact networks, for example: Using the SocioPatterns RFID tags and the CONFERATOR system described below, face-to-face contacts can be measured. When the RFID tags are worn on the chest of the conference participants, tag-to-tag proximity is a proxy for face-to-face communication, since the range of the signals is approximately 1.5 meters if not blocked by the human body. For detecting a contact, we apply the technique described in [119], which is based on a typical conversational setting: A face-to-face contact is observed 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.

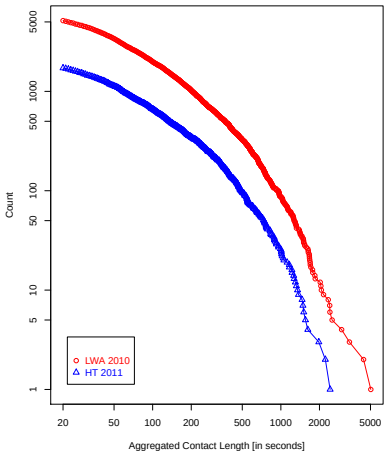


Fig. 1. Example [107] of a typical cumulated contact length distribution of human face-to-face contacts, collected at two conferences (LWA 2010 and HT 2011). The x-axis displays the minimum length of a contact (in seconds), the y-axis the number of contacts having at least this contact length.

Face-to-face contact networks are a special sort of social interaction networks. Using face-to-face contacts between pairs of actors, we can derive a face-to-face contact networks. The networks are constructed such that links connect actors, if these were in face-to-face contact, possible weighted with the contact count or the (normalized) contact duraction. Other networks that indicate real-world interactions, e. g., by co-authorship or frequent co-location, are obtained by considering co-visited talks, posters, or co-authored papers. Networks based on co-location can be constructed, for example, based on the counts or (normalized) durations that pairs of participants were observed in the same location, for example, at room-level. Figure 1 shows an example of the typical contact distribution that we observe in conference face-to-face contact networks: Confirming previous findings, e. g., by the Sociopatterns collaboration [56], most of the contacts take less than one minute and the contact durations of both conferences show a long-tailed distribution.

In contrast to the work presented by the SocioPatterns collaboration discussed above, we aim at a larger research focus concerning data mining and social network analysis: We examine social dynamics in conferences and the contexts of working groups. Additionally, we especially investigate communities, and roles of actors in these networks grounded using background information. Furthermore, we do not primarily approach the *modeling* of the networks phenomena from a social network analysis perspective, but focus on the analysis, and especially on approaches and methods for mining and extracting useful models and patterns from the data.

2.3 CONFERATOR and MYGROUP: Enhancing Social Interactions

CONFERATOR [10] and MYGROUP [9] are ubiquitous social systems for enhancing social interactions: CONFERATOR is a social conference guidance system for efficiently managing face-to-face contacts at a conference, and collectively building a personalized conference program. MYGROUP is a similar system in the context of working groups that aims to enhance interactions and knowledge exchange between the individual team members. Both systems are built on top of the RFID-based proximity sensing system developed by the SocioPatterns[7] collaboration. The applied RFID tags allow the coupling of real world (offline) data, i. e., face-to-face contacts, with the online social world, e. g., given by online interactions within the system or in linked online social networks. In particular, these RFID proximity tags can collect face-to-face contacts. This allows for highly personalized profiles in the systems which can be applied, e. g., for community mining, recommendations, or for improving the localization.

In the following, we first give an overview on the systems and the applied technical platform UBICON. After that, we summarize the most important features of the systems.

Overview. CONFERATOR [10] is a social and ubiquitous conference guidance system, aiming at supporting conference participants during conference planning, attendance and their post-conference activities. It features the ability to observe and to manage social and face-to-face contacts during the conference and provides a number of features for supporting social networking.

In a similar context, the conference navigator by Brusilovsky [37, 131] allows attendees of a conference to organize the conference schedule. However, it is not connected to the real live activity of the user during the conference. Hui et al. [55] describe an application using Bluetooth-based modules for collecting mobility patterns of conference participants. Furthermore, Eagle and Pentland [35] present an approach for collecting proximity and location information using Bluetooth-enabled mobile phones, and analyze the obtained networks. Similarly, the Find-And-Connect [133] system utilizes bluetooth and passive RFID for obtaining locations of participants, and infers *encounters* based on the co-location of participants as a proxy for contacts between participants. However, no direct face-to-face contacts are measured. In contrast to these systems, CONFERATOR is able to collect the real-world face-to-face contacts using the SocioPattern RFID tags described above. The setup requires a number of RFID readers at fixed positions in the target area; the participants are then equipped with RFID tags. The technology allows for the localization of tags and for detecting tag-to-tag proximity. When the tags are worn on the chest, tag-to-tag proximity is a proxy for a face-to-face contact, since the range of the signals is approximately 1.5 meters if not blocked by the human body. For more details, we refer to Barrat et al. [30].

Utilizing RFID proximity tags, social interaction data can be collected at a much more accurate level than, e. g., based on co-location information [35, 133]. From the data, we can derive social interaction networks, apply these for mining the collected data, and put the discovered patterns and learned models to use by the system. With CONFERATOR, we provide a broad application spectrum and features: Compared to

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

previous RFID-based approaches, we increased the precision of the localization component and linked together tag information and the conference schedule. Furthermore, we implemented a light-weight integration with BibSonomy, and added connectors to other social systems used by participants, e. g., Facebook, Twitter, Xing or LinkedIn. This provides the basis for new insights into the behavior of all participants concerning their real-world (offline) and online social interactions. CONFERATOR has been successfully applied at LWA 2010[8] [12], LWA 2011[9], and LWA 2012[10] – conferences for special interest groups of the German Computer Science Society (GI), at the Hypertext 2011[11] conference [80], and at a technology day of the VENUS project.[12]

MYGROUP aims at supporting members of working groups. It employs the same technology as CONFERATOR for localizing the members and for monitoring their social contacts. Additionally it provides profile information including links to (external) social software, e. g., BibSonomy [24], Twitter, Facebook, or XING. MYGROUP has been applied at a number of different events: Since December 2010, it is being continuously applied by the Knowledge and Data Engineering (KDE) group at the University of Kassel, and is currently being extended towards a larger research cluster. In addition, MYGROUP has also been utilized at a large student party,[13] for supporting organizational processes, at the First International Changemaker-Camp[14] at the University of Kassel for profiling group processes, and at a CodeCamp for supporting software development in the context of the VENUS project.

CONFERATOR and MYGROUP have been implemented using the UBICON platform – a platform for enhancing ubiquitous and social networking. From a technical point of view, the UBICON platform consists of the application logic, components for privacy management and database management, a (customizable) set of data processors that process the incoming (raw) data, a set of data processors for more subsequent sophisticated processing, and a storage architecture based on a MySQL database.[15] The set of data processors include, e. g., the localization component for determining the location of RFID tags. The system is implemented with a model-view-controller pattern using the Spring framework.[16] UBICON can be deployed using a standard servlet container, e. g., Apache Tomcat.[17]

Features of CONFERATOR and MYGROUP. CONFERATOR and MYGROUP enable ubiquitous social networking using profile information, data from social contacts, as well as from monitoring the current activity streams, e. g., ongoing contacts, visits of talks, BibSonomy posts, Twitter tweets, etc. From a data mining perspective, we can

[8] http://www.kde.cs.uni-kassel.de/conf/lwa10
[9] http://lwa2011.dke-research.de
[10] http://lwa2012.cs.tu-dortmund.de
[11] http://ht2011.org
[12] http://www.uni-kassel.de/eecs/iteg/venus
[13] http://wintersause.de
[14] http://www.knowmads.nl
[15] http://mysql.com
[16] http://springsource.org
[17] http://tomcat.apache.org

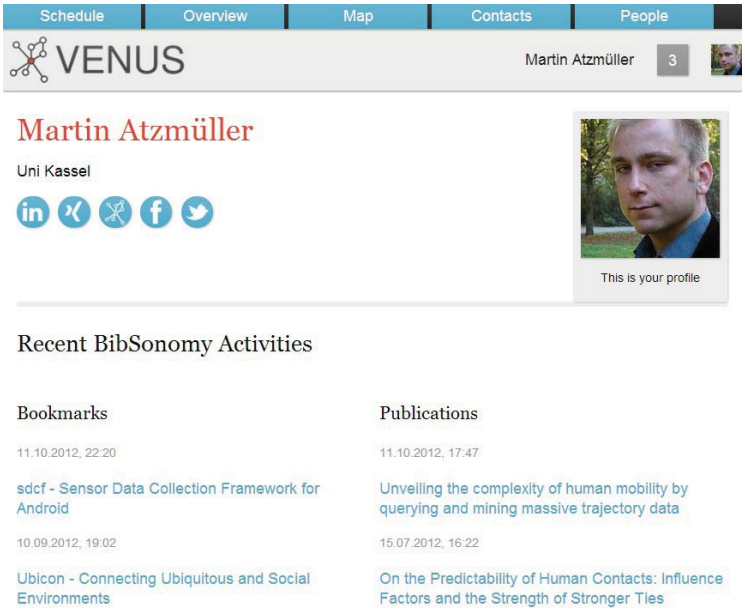


Fig. 2. A screenshot [9] of a CONFERATOR user profile, with contact information and latest BibSonomy activities, i. e., publication and bookmark posts. Further profile information (not shown) includes, for example, a personalized social tag cloud and location and contact context information.

exploit the collected social structures provided by both social networks and social resource sharing systems for supporting complex and structured interactions: We can recommend persons, for example, based on joint research topics or contacts. We apply data mining methods on the collected data to make this information visible to our users. Different trust and privacy settings, e. g., concerning the visibility of contacts and locations, allow a selective distribution of sensitive information which is important for increasing trust in the system [33, 115].

Conference participants can recall their individual contacts, e. g., as virtual business cards, cf. Figure 2. The system allows the setup of a complete profile, social networking to other participants, and the management and personalization of the conference schedule, by providing helpful information about the individual talks, upcoming talks, and the ability to pick talks for a personal conference program. In addition, recommendations for contacting *interesting* persons are provided. CONFERATOR and MYGROUP feature the following basic options: They allow participants to recall their contacts (in the *contact view*, depicting the shares of contacts to other users), to observe the (online) social live around them (*timeline view*, as a configurable list of time-based events and activities), to help in finding participants via the localization component (*map view*, see Figure 3) and to browse the individual profiles (*profile view*, see Figure 2). Furthermore, recommendations and context-specific notifications are provided [9, 10], e. g., concerning other conference participants or the contact history.

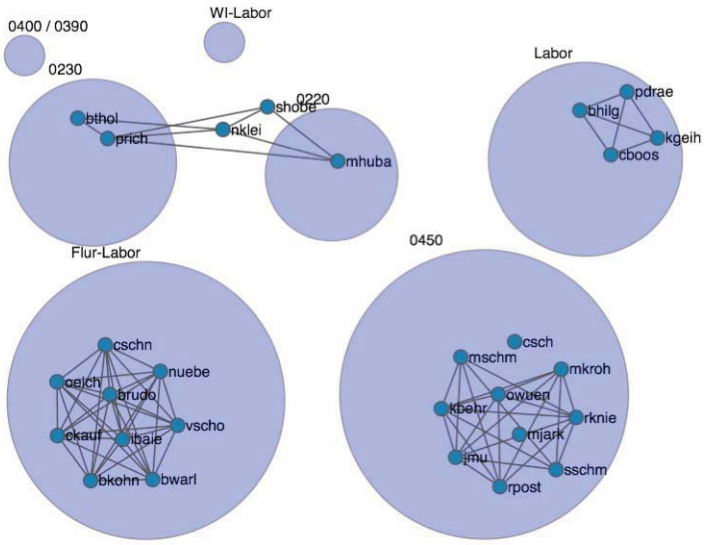


Fig. 3. A screenshot [9] of the *map view* of MYGROUP. The large circles denote individual rooms, the smaller circles participants; connections between those indicate ongoing face-to-face contacts.

In addition to the profile pages, the timeline visualization [9] is one of the main visualizations of CONFERATOR. It is an aggregation of different events and activities of the participants arranged in a time-oriented list. This includes, for example, contacts, BibSonomy posts or Twitter tweets: It provides an aggregated view on the currently active topics published on Twitter or Bibsonomy by the participants and the conversations that recently happened. MYGROUP utilizes the same visualization technique: The timeline, for example, which is displayed on a large LCD screen at the KDE group at the University of Kassel, often stimulates interesting research discussions and enables enhanced dissemination and exchange of knowledge. The systems are continuously refined according to user feedback and usability studies, leading to continuous improvement of the systems and implementation of new useful features.

Altogether, CONFERATOR and MYGROUP create a ubiquitous and social environment where large groups (implicitly) collaborate together using electronic media to accomplish certain tasks. Ultimately, this enables a form of *Collective Intelligence* [69, 82, 83, 129]: A very intuitive notion captures the term as "groups of individuals doing things collectively that seem intelligent" [82]. In the context of conferences and working groups, for example, talk and contact recommendations can be improved using the collected interaction data. Furthermore, interesting topics can be identified at a conference, for example, by taking the top visited talks and their descriptions into account.

3 Analysis of Social Interaction Networks

The analysis of online social network data has received significant attention: As a prominent example, Mislove et al. [88], applied methods from social network analysis as well as complex network theory and analyzed large scale crawls from prominent social networking sites. They worked out properties common to all considered social networks and contrasted these to properties of the web graph. Broder et al. [27] utilized complex network theory for analyzing (samples from) the web graph. Kwak et al. [65] provide an an analysis of fundamental network properties and interaction patterns in Twitter, while Ahn et al. [3] provide an analysis of the topological characteristics of networks in online social networking services. Furthermore, Newman [100] analyzed many real life networks, summing up their characteristics. Similarly, an analysis of fundamental network properties and interaction patterns in Twitter can be found in [65].

In this section, we investigate structural interrelations between social interaction networks, and analyze communities and roles in face-to-face contact networks at conferences. Furthermore, we inspect and investigate the interactions and dynamics of conference participants. In this way, we aim at gaining a better understanding of the behavior and its underlying processes. The extracted knowledge can then be applied, for example, for optimizing processes, or the integration into applications.

For enhancing our understanding of social interaction networks, we focus on the network structures and the properties of the nodes. In the context of conferences, for example, we aim to analyze the behavior and relations of the participants in order to ultimately uncover common patterns and trends throughout the conferencing scenario. We combine data mining and social network analysis techniques for examining social interaction networks and summarize specific methods and analysis results in the context of those interaction networks. We present analysis results in the context of social bookmarking and social resource sharing systems, as well as in human face-to-face contact networks – using the CONFERATOR and MYGROUP systems.

In the following, we first present an analysis of structural interrelations on social interaction networks. After that, we discuss dynamics of communities and roles in human contact networks at conferences and present an analysis of dynamic and static behavior of conference participants.

3.1 Structural Network Interactions and Correlations

Social interaction networks are of large interest for major applications, such as recommending contacts in online social networks or for identifying groups of related users, cf. [90, 91]. There, we provide a detailed structure and semantic based analysis of three sample social media applications: Twitter (microblogging), Flickr (social resource sharing), and BibSonomy (social bookmarking). In each application, we identify various implicit and explicit social interaction networks, also called evidence networks, focusing on explicit and implicit user traces.

In the following, we outline two main issues: Are there interrelations and correlations between the interaction networks? Furthermore, can these be applied for the analysis and data-driven assessment of communities? The second question is especially important, since one of the main problems of community detection [40, 66, 98, 99] is the

non-trivial evaluation and validation of the identified communities. The assessment of the quality of a given community is always application dependent and *relative* to certain aspects of user relatedness, e. g., race of individuals in [97], shared topical interests in social bookmarking systems, or social traces manifested in the social interaction networks. Often there is no gold-standard evaluation data at hand in order to validate the discovered groups.

Interrelations in Social Interaction Networks. As a starting point for the analysis of the considered social interaction networks, we perform a structural interrelation analysis in order to identify general properties and to compare different networks focusing on common network structures, community structures, and user-relations within these networks. We examine and compare social interaction networks applying user data from the real-world social bookmarking application BibSonomy, Flickr and Twitter. We analyze general structural properties of the obtained networks and comparatively discuss major structural characteristics in order to show that there are structural and semantic inter-network correlations between the different evidence networks.

In particular, we examine several general structural properties, the degree distribution and the degree correlation, indicating significant similarities of the networks. Furthermore, we analyze topological and semantical distances, the dependencies of the networks' neighborhood, and the inter-network correlations, and collect evidences for strong correlations and interrelations between the considered social interaction networks. Specifically, we analyze inter-network correlations between such user-generated networks and show that these relations are strong enough for inferring reciprocal conclusions between the networks.

Analysis of Community Structure. The social interaction networks are thoroughly analyzed with respect to the contained community structure. Using standard community measures, e. g., modularity [98], segregation index [41], and conductance [73], we show that there is a strong common community structure accross different social interaction networks. Furthermore, we analyze the rankings between a large set of communities mined on the different networks, and show, that the induced rankings are reciprocally consistent. Therefore, since the correlations and dependencies are strong enough for assessing structural analysis techniques, e. g., community mining methods, implicitly acquired social interaction networks can be applied for a broad range of analysis methods instead of using expensive gold-standard information. We therefore propose an approach for (relative) community assessment: The presented approach is based on the idea of *reconstructing existing social structures* [113] for the assessment and evaluation of a given clustering. This provides for a simple yet cost-efficient assessment option, since we can apply secondary data, i. e., implicitly acquired social interaction networks capturing user relatedness, for assessing community structures for another network in the same application context. Specifically, the presented analysis is thus not only relevant for the evaluation of community mining techniques, but also for implementing new community detection or user recommendation algorithms, among others.

3.2 Communities and Roles in Face-to-Face Contact Networks

Gaining a better understanding of communities and roles in social interaction networks helps for a number of applications, for example, for personalization, community detection, or recommendations.

In [12], we investigate user-interaction and community structure according to different special interest groups during a conference. For the analysis, we considered the contact network of the LWA 2010 conference that we obtained using the CONFERATOR system. The analysis thus utilizes real-world conference data capturing community information about participants and their face-to-face contacts, and is grounded using information about membership in special interest groups and academic status/position. We analyze various general structural properties of the contact graph, confirming previous results of the SocioPattern experiments concerning typical face-to-face contact networks at conferences e. g., [56].

Furthermore, we examine different explicit and implicit roles of conference participants in the context of community structure at the conference. Role mining concerning communities mainly analyzes the relations between the communities for a specific actor. Scripps et al [110, 111] present a method for assessing roles with respect to the membership in the communities and the potential to bridge or to connect different communities. In this way, different actor profiles concerning their centrality prestige and their community importance can be derived. Chou and Suzuki present a similar method considering a set of given communities [32] for such a community-oriented analysis.

Community Dynamics. For analyzing the community structure and its dynamics, we consider community detection methods in a time-based analysis using the special interest group communities as a ground truth. We analyze, whether there is a detectable community structure in the investigated face-to-face networks that is consistent with the one given through the groups. Using the community measure *segregation index* [41] and modularity [96], for example, we can observe the trend, that more relevant (i. e., longer) conversations are biased towards dialog partners with common interests, as captured by the interest group membership: Members of a special interest group tend to talk more frequently and longer within their interest group, as analyzed on the accordingly weighted social interaction networks.

Furthermore, with increasing minimal conversation lengths the induced social interaction networks show a more pronounced community structure, both considering automatically mined communities as well as the communities defined by the special interest groups. This finding is also supported by analyzing the densities in the respective subgraphs induced by the community detection method, and the respective ground-truth communities. Therefore, the face-to-face contacts show inherent community structure, which is also consistent with the special interest groups.

Roles and Communities. For analyzing the dynamics of roles of the participants, we also utilize the ground-truth community information. We provide a time-based analysis of the roles, and different role profiles. For example, we consider *bridges* connecting communities, or *ambassadors* as important bridging actors. As an example, Figure 4

shows the *Ambassador* role in a time-based analysis concering increasing minimal conversation length thresholds. Intuitively, an ambassador has a high node degree and is able to connect different communities. In Figure 4, we observe the trend that the organizers start out as ambassadors for shorter conversations, but become less and less important for longer conversations. In contrast, more and more professors are categorized as ambassadors, for longer conversations. For organizers, this is in line with our expectations, since they need to be involved in a lot of shorter conversations. On the other hand, especially professors seem to be able to bridge communities with longer (more meaningful) conversations. Similar results concerning the communities and roles were also obtained for other conferences, e. g., for the Hypertext 2011 conference [81].

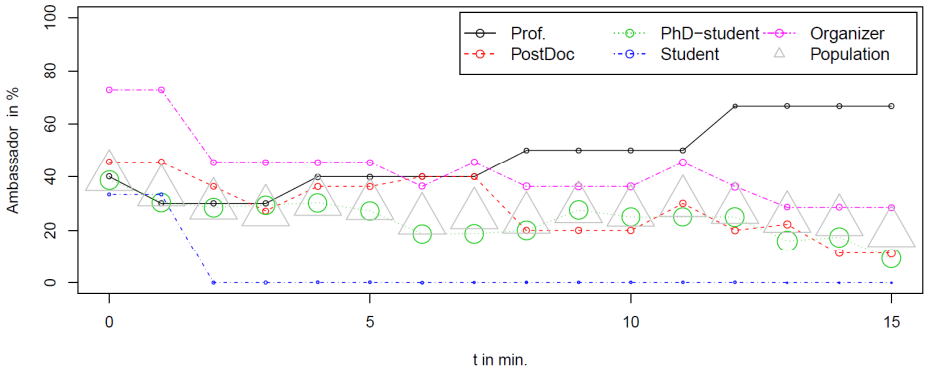


Fig. 4. Fraction of the participants that assume the role Ambassador at LWA 2010 when considering all conversations with a length $\geq t$, cf. [12]. Ambassadors are characterized by a high degree and by a high connectivity between different communities, cf. [110].

3.3 Structure and Dynamics on Interactions of Conference Participants

Understanding the mechanisms of conference interactions and their dynamics, e. g., using a time-based analysis of conference participants, can help in many ways: It can increase the efficiency and effectiveness of individual networking, support the conference organization, be utilized for process optimization or be incorporated into advanced data mining methods and tools. However, the analysis of conference interactions is not easy if conventional tools like questionnaires are used, cf. [124, p. 45 ff.], 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 [81], we present an in-depth analysis of the static and dynamic nature of a conference, exemplified by the ACM Hypertext conference 2011 in Eindhoven, The Netherlands, where we collected RFID face-to-face contact data using the CONFERATOR system. Figure 5, for example, depicts the conference contact dynamics for all coffee and lunch breaks, and gives a first impression of the contact hotspots during the conference. Since the setup of the CONFERATOR system started in the middle of the first day

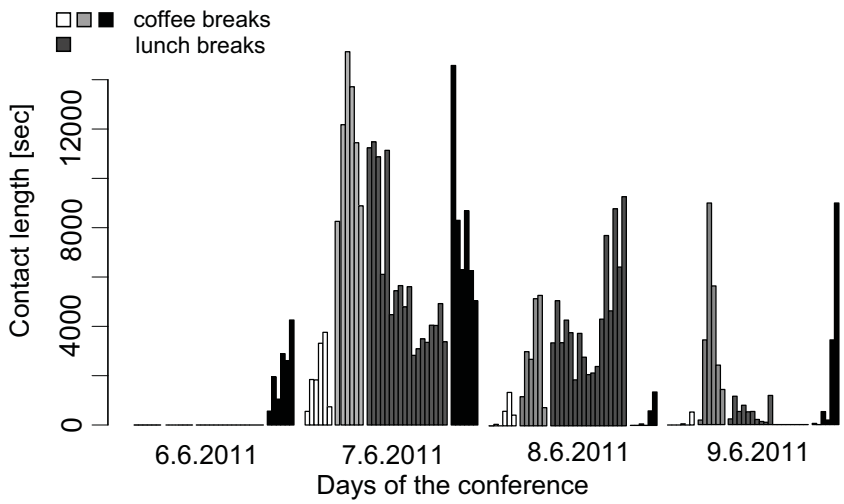


Fig. 5. Dynamic contact activitiy during the ACM Hypertext 2011 conference [81] - focusing on the coffee and lunch breaks. The time slices contain contact durations for the complete conference except for the sessions. The start times of the coffee breaks were as follows: 8:30, 10:30, 15:30 and 16:00. 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 window summing up all durations of all face-to-face contacts contained in this window; adjacent bars belong to the same coffee or lunch break.

(6.6.2011), 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 departing 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.

In the following, we specifically consider the *individual* behavior of participants at the conference, and their *community* interactions, e. g., insights into the communication in tracks. One of the first experiments in a similar context at conferences was performed by Cattuto and colleagues, cf. [4, 22, 119]. We extend their findings with a number of results for homophily and session attendance of the participants, their communication behavior and an analysis concerning their submitted papers.

Individual Behavior. Our analysis focuses on individual behavior considering the communication within tracks, after the end of a session, and the behavior of special roles, e. g., presenters. Focusing on these, we also include the content of the submitted papers using a bag-of-words model. In contrast to intuition, in the analysis of the presenters, we cannot confirm our assumption, that these were more involved in talks with participants presenting similar work based on the content of their papers. For the track visiting behavior, we find that all tracks focus on their own community, concerning the

session attendance of the individual members of the track. Furthermore, we provide an in-depth analysis of participants, presenters, session chairs, concerning their roles at the conference, by mining role patterns, e. g., for the ambassador or bridge role, cf. Section 3.2 for details. The strength of the affiliation of the conference turns out to be one of the strongest features in the patterns that determines to connect different communities that are present at the conference.

Community Interactions. For the analysis of interactions within different communities, we investigate different partitionings, e. g., concerning the individual tracks and sessions, but also automatically mined communities with respect to their contact behavior. Figure 6 shows an example of community structure concerning different 'organizational' aspects, i. e., *track*, *country*, *affiliation to the conference* and *academic status* communities: These are partitionings according to the visited track, the country of origin, the affiliation to the conference, and the academic status, e. g., Professor, PhD, or student. In Figure 6 we observe, for example, that the length of the conversations has a high impact for the track community: The longer the conversation, the higher the probability of having a contact within the same track community. Vice versa, we also observed, that longer conversations are more probable, if the dialog partners are both members of the same track.

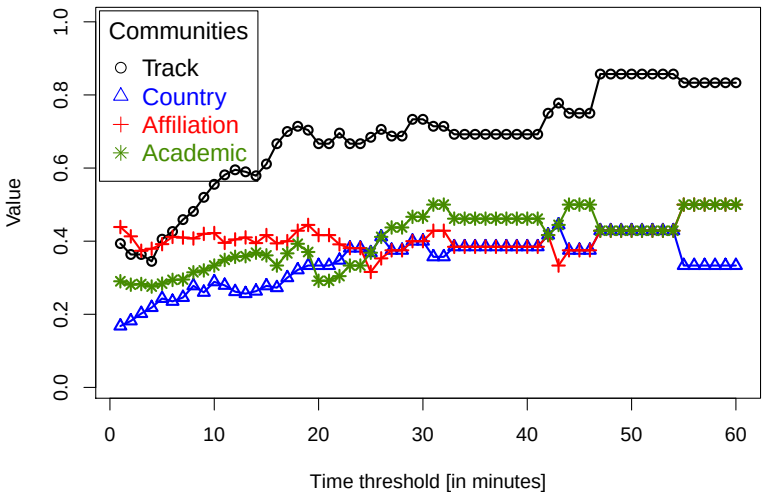


Fig. 6. Overview: Community quality indicator (p-value [110]) for the partitionings track, country, affiliation and academic status, in a threshold-based analysis using different minimal conversation lengths, cf. [81]. The higher the p-value, the higher the probability for a contact within a community.

4 Predictive Modeling

Approaches for predictive modeling aim to learn models for later deployment, e. g., for estimating a function that predicts a certain value for future cases [38, 39]. Classification or regression are such typical predictive tasks. In this section, we present predictive approaches in the context of social interaction networks utilizing data from social and ubiquitous systems, specifically the CONFERATOR and MYGROUP systems. For these, standard predictive methods, e. g., classification using random forests, as well as PageRank-based and link mining approaches were adapted and extended as needed.

Below, we present approaches for resource-aware localization, for recommendations of experts, and for analyzing the predictability of links in face-to-face contact networks. For the localization, we combined tracking signals and contact information from social interaction networks for enhancing the performance of common machine learning techniques using different voting mechanisms. Further, we adapted the well-known PageRank method [26] on special social interaction networks. These are mediated by software (source code) resources using information from revision control systems, extended with social contact information. Using this information, the performance of the recommendation method is significantly improved compared to standard approaches. Finally, we consider a link prediction approach in social interaction networks where we analyze the impact of stronger ties and identify influence factors for the prediction.

4.1 Resource-Aware RFID Indoor Localization utilizing Social Interactions

In the context of CONFERATOR and MYGROUP, knowing where attendees and colleagues are, respectively, supports group organization and thus facilitates everyday work processes. The localization component provides the locations of all users, and shows where their conversations take place. During conferences, for instance, CONFERATOR offers the possibility of observing who is visiting a given talk, thus facilitating the academic exchange during the subsequent coffee breaks.

Furthermore, it is possible to identify hotspots, e. g., conference rooms where a large number of conference participants is listening to – apparently interesting – talks, and to potentially recommend those to undecided participants. Capturing and visualizing live interactions of individual users is an important task for CONFERATOR and MYGROUP, essentially in order to enable collective intelligence. Therefore, a localization framework is a central component for such a system.

Resource-Aware Localization Setting. In [108], we present an approach for a resource-aware and cost-effective indoor localization method in the context of RFID based systems. While approaches for outdoor localization can utilize various existing sources, e. g., GPS signals, mobile broadcasting signals, or wireless network signatures, methods for indoor localization usually require special installations (e. g., RFID or Bluetooth readers), and/or require extensive training and calibration efforts.

The proposed cost-effective and resource-aware solution requires only a small number of RFID readers. Furthermore, our method can be applied to installations, where readers cannot be positioned freely. The latter constraint is encountered often, especially in historical buildings under monumental protection.

In contrast to typical experiments that examine RFID localization in laboratory experiments, e. g., [51, 101], we present an analysis of data collected in a real-life context, that poses additional problems in terms of signal quality and noise: We consider a real-life localization problem at room-level, i. e., the task to determine the room, that a person is in at a given point in time.

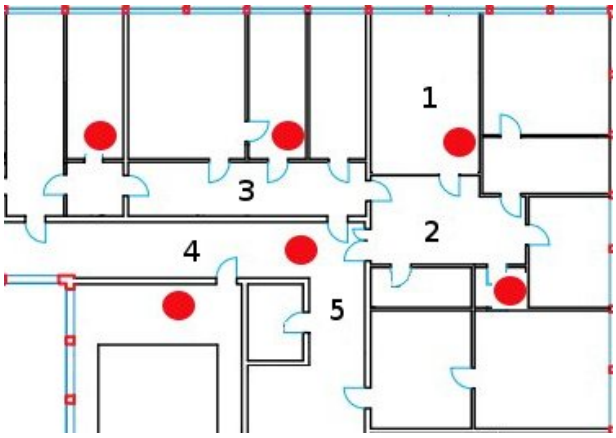


Fig. 7. Example conference area [108]: The numbered rooms were used by participants during the poster-session of LWA 2010, the circles mark the positions of RFID readers.

Social Boosting: Improving Localization using Face-to-Face Contacts. We present an analysis of the contact and proximity data in order to prove the validity and applicability for the sketched application. Additionally, we evaluate the benefits of several state-of-the-art machine learning techniques for predicting the locations of participants at the room-level. We propose to utilize the (proximity) contacts of participants for improving the predictions of a given core localization algorithm using different voting mechanisms. We evaluate the impact of different strategies considering the top performing machine learning algorithm. The real-world evaluation data was collected at LWA 2010, cf. Section 3.2. Figure 7 shows the covered conference area with 6 readers put in adequate positions.

We evaluated several state-of-the-art machine-learning algorithms in this context, complemented by novel techniques for improving these using the RFID interaction contacts. The results of the experiments yielded several reasonable values for the applicable parameters. For the simpler algorithms, they could also have been learned in a short preceding training phase, which demonstrates the broad applicability of the approach in the sketched resource-aware setting. Overall, using the social face-to-face contact information in the *Social Boosting* algorithm [108], we could improve the localization accuracy from 84 % using a baseline algorithm to nearly 90 %, as evaluated during the poster session of the LWA 2010 conference.

4.2 Combining Interaction Networks for Expert Recommendation

Recommendations play an important role in supporting software development, especially in larger teams: Locating experts for a given problem is one of the main challenges when working in a large team.

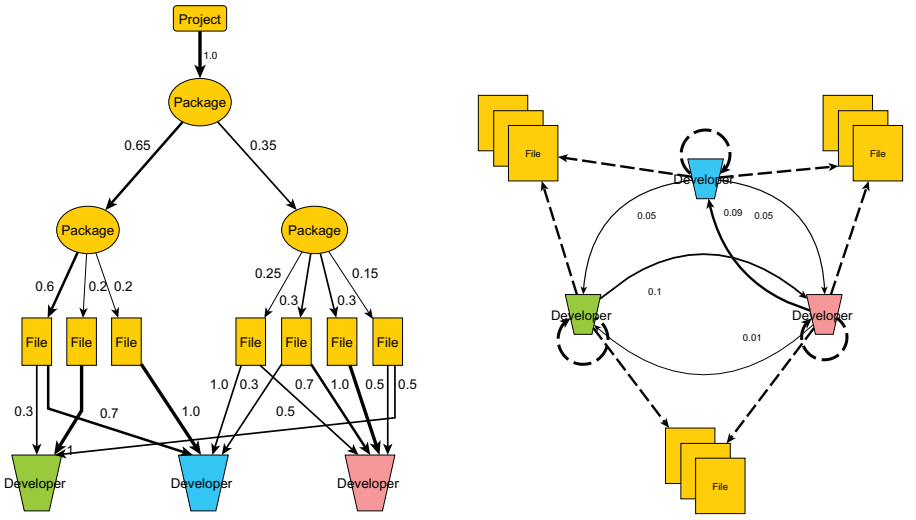
Expert Profiling. Our work [80] presents an approach for expert profiling and recommendation in such contexts: First, specific profiles of developers concerning resources, packages, and projects provide an overview on the area that the respective developer is working on, e. g., for an overview on the activity of a team. Second, predictive models, e. g., modeling the familiarity of developers with specific resources, can increase the effectiveness of other team members. In the case of specific questions, persons can be suggested that are especially familiar with these resources. In this way, knowledge management and knowledge transfer, e. g., transfer of projects, instructing new team members, or participation in open-source projects, can be successfully implemented.

In the context of MYGROUP, we focussed on supporting software development groups [80]. The presented approach can potentially be generalized for any organization using revision control systems, e. g., for recommending collaborators based on changes in documents, papers, or wikis, etc. We propose an approach for analyzing the communication and commit structure of a development group. In our case study, the development group uses CVS as a code versioning system; additionally, conversations between developers are captured using the RFID tags applied by MYGROUP as described above. In the sketched scenario, the two basic assumptions are the following: The number of added and removed lines of code that a developer commits for a specific resources serves as a proxy for her *familiarity* with this specific portion of code, e. g., [44, 52]. Additionally, conversations between developers serve as a way for transferring *knowledge* from one developer to another. Therefore, such interactions also help to increase the familiarity of developers with the source code.

Combining Interactions for Recommendation. We provide two novel classes of graphs built from structural data for mining developer and resource profiles concerning the 'familiarity' with specific resources, packages and/or projects. We analyze code changes and the structure of the software projects. In this way, we create resource trees resembling the hierarchic organization of source files, cf. Figure 8 for a simple example. The contribution of each developer is then measured by the number of changed lines of code. We combine this information with the RFID face-to-face contact network measuring the face-to-face contacts of the developers for modeling the real-life communication, especially conversations that take place before a commit. Using this social interaction network, we are able to store the time and duration of a conversation, so that this information can be analyzed further. This enables the capture of conceptual knowledge which is mostly propagated via conversations and cannot be extracted by mining software repositories. In addition to weighted edges which reflect the relative amount of changed lines of code, we consider edges between the developers. These edges are weighted by the cumulative duration of the face-to-face contacts of the developers within the last eight hours before committing changes to the source code. Thereby,

we connect their real-life communication and social interactions using the MYGROUP system. The resulting structure captures important knowledge of a social group: Exogenous information, e. g., developers writing code by themselves, and endogenous information representing the knowledge transfer from one individual to another by means of communication.

Utilizing the graph structure, the *PageRank* [26] algorithm is then applied: We can calculate either a set of developers that are familiar with a specific portion of source code, or we can analyse the quality of code coverage by a given subgroup of software developers. We evaluated the approach in the context of a developer group with a medium sized project comparing the proposed approach against a ground-truth expert-ranking. As shown in the evaluation, the explanatory and predictive power of exogenous experience captured by CVS logs combined with the endogenous flow of experience captured by logged communication shows strong improvements concerning an approach only based on a lines-of-code analysis. The evaluation results demonstrate the effectiveness and impact of the proposed approach: The contact information (RFID) usually improves the performance of the proposed approach compared to a baseline using only a lines-of-code analysis.



(a) Exemplary resource tree with edges from seven files to three developers [80]. The edge weights are determined by the hierarchical partitioning of the lines of code in each resource, and by the contribution of each developer (at the bottom of the tree), respectively.

(b) Example developer contact graph [80] (the dashed lines correspond to those in the resource tree). Edges between developers are weighted by their relative conversation durations, normalized by a parameter (0.1 in our example) which aims to model the amount of knowledge transferred in communication.

Fig. 8. Exemplary resource and developer contact graphs [80]. For recommending software developers, the PageRank algorithm [26] is then applied to their combination.

4.3 Link Predictability in Face-to-Face Contact Networks

Link prediction [77] in social network considers the dynamics and mechanisms in the creation of links between the actors of social networks. The goal is to learn a model for predicting new and/or recurring links accurately. This also relates to mobility [20, 122] and dynamic behavior [121, 125]. Link prediction in social interaction networks has a number of prominent applications, including the prediction of missing links, cf. [77], for improving collaborative filtering, e. g., [54], or for recommending new contacts, e. g., [76, 102]. A method for recommending interesting contacts, for example, has been deployed in the CONFERATOR system.

There is already a large body of research for link prediction concerning *online* social networks, e. g., [1, 60, 77, 79, 94, 135]. However, important aspects of face-to-face contact networks, i. e., interactions that happen offline, still remain largely unexplored. Sociological experiments and approaches, e. g., [36, 43, 78], mainly rely on questionaires, diaries, or recordings, and usually only consider rather small groups, cf. [124, p. 45 ff.]. In contrast, the CONFERATOR system is able to collect face-to-face contacts (and the according networks) at much better precision and for rather larger groups. The analysis of such networks can potentially provide more direct answers to fundamental questions, e. g., how do personal links get established, what are influence factors in this contexts, what is the impact of stronger ties in face-to-face contact networks.

Predicting New and Recurring Links. In [107], we aim at providing first insights for answering such questions. We focus on face-to-face contact networks. For the analysis, we apply real-world data collected at the LWA 2010 and Hypertext 2011 using the CONFERATOR system, cf. Section 3.2-3.3. Using this data, we can observe and analyze social interaction at a very detailed level, including the specific event sequences and durations. We aim to predict *new* contacts based on network properties of face-to-face contact networks, e. g., for a recommendation setting, as an adaptation of methods for online social networks. For that, we apply and extend basic link prediction measures utilizing the network structure, and their weighted variants.

In addition, we extend the analysis in two important directions: First, we consider the length of the contacts in more detail, and analyze the impact of longer conversations. Second, we consider the prediction of future *recurring* contacts, i. e., renewed contacts between specific actors, e. g., on the first day of the conference vs. the subsequent days. For these, we analyze influence factors and patterns for establishing such contacts, and also consider their specific *durations* in a fine-grained dynamic analysis. Essentially, this leads to the analysis of the impact of *stronger ties* for new and recurring contacts. We estimate the influence of stronger ties for the prediction and show its impact using real-world data of two conferences.

Link Predictability and Stronger Ties. The results of the analysis indicate that stronger ties have a strong influence on the contact behavior and the prediction performance. As depicted in Figure 9, we observe, for example, that a longer contact, i. e., a conversation, is more likely, the longer the contact on the first day of the conference. An interesting further question is to find typical features to predict renewed contacts and their lengths.

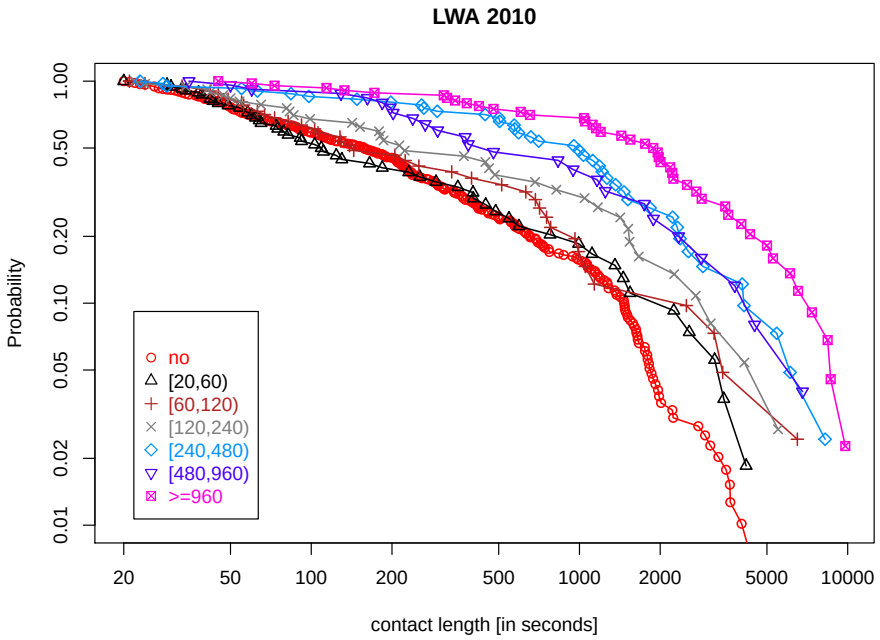


Fig. 9. Example of predicting recurring contacts at LWA 2010 showing the impact of the contact duration between two participants on the first day, and the contact length of a recurring contact at the second and third day, cf. [107]. The red line labeled with 'no' (circle symbol) in the LWA 2010 plot, for example, shows the distribution of all contacts between participants at days two and three, which had no contact at the first day. The line labeled with $[60, 120)$ (cross symbol) shows the distribution of all contacts between participants at days two and three, which had a contact with contact duration between 60 and 120 seconds at the first day of the conference.

We show, that there are clear influence patterns of the contact durations, depending on roles such as academic status, the strength of the link to the conferences, and affiliation with the respective conference tracks. Furthermore, considering the contact durations in the ranking of the predicted contacts significantly improves the performance.

Overall, the results of the analysis provides interesting insights especially concerning the impact of the contact durations and the strength of such stronger ties. These insights are a first step onto predictability applications for human contact networks. The indicators, patterns, and influence factors can then be integrated into more advanced prediction models in the context of face-to-face contact networks: New features can then be constructed for supervised or unsupervised link prediction methods.

5 Descriptive Pattern Mining

Descriptive data mining aims to uncover certain patterns for characterization and description of the data. Typically, the goal of the methods is not only an actionable model, but also a human interpretable set of patterns [84, 123]. Descriptive pattern mining especially supports the goal of explanation-aware data mining [18], due to its more interpretable results, e. g., for characterizing a set of data [11], for concept description [16], and for providing regularities and associations between elements in general.

In the following, we present novel methods for descriptive data mining for uncovering and extracting relations and patterns. These are applied for hypothesis generation and exploration by the user, in order to provide characteristic information about the data and networks. We focus on the mining of patterns that describe interesting communities and subsets of nodes contained in social interaction networks.

We first describe an efficient method for descriptive community mining based on the novel COMODO algorithm [17]: Descriptive community mining aims to identify interesting communities according to a community evaluation function using the network structure, and a set of descriptive attributes. Next, we discuss a generalized setting for identifying descriptive patterns utilizing exceptional model mining [34,62,70], and provide the GP-GROWTH algorithm [71] for fast exhaustive exceptional model mining: It can then be applied both for characterizing the network structure using an approach similar to descriptive community mining, but also for describing interesting subgroups of nodes in the network based on their properties. After that, an exploratory approach [15] on social interaction networks using geo-tagged social media is presented, which utilizes both techniques presented above. All approaches can be applied for supporting the user, e. g., for recommendations, facetted browsing, or for interactive exploration.

5.1 Efficient Descriptive Community Mining

Community mining is a prominent approach for identifying densely connected subgroups of the nodes contained in a network. A community is intuitively defined as a set of nodes that has more and/or better links between its members compared to the rest of the network. Classic community mining, e. g., [45, 66, 75, 96], works on a set of nodes connected by links. In general, usually not only the density within the community is assessed but the connection density of the community is compared to the density of the rest of the network [98], e. g., using the modularity [45, 98, 99], the segregation index [41] or the conductance [74] as an evaluation function. Then, cuts between communities are established in such a way as to maximize the community evaluation function The core idea of the evaluation function is to apply an objective evaluation criterion, for example, for the modularity the number of connections within the community compared to the statistically "expected" number based on all available connections in the network, and to prefer those communities that optimize the evaluation function. Usually, an optimization approach is taken that partitions the whole graph subsequently into a number of parts – each of them is then considered as a community. The discovered communities can then be applied, for example, for recommendations or for personalization of intelligent systems.

Descriptive Community Mining. In [17], we present an approach for mining descriptive community patterns according to standard community evaluation measures: The proposed method collects patterns that describe communities by combinations of features, e. g., tags or topics for social bookmarking systems. We can consider, for example, groups of users interested in the topics *web mining*, *computer* and *java*.

In this way, we aim to *identify and describe* interesting communities, in contrast to standard community mining approaches, e. g., [66, 75, 96], that only identify communities as subsets of users. In contrast to such global approaches, we focus on the discovery of local community patterns. According to the idea of local pattern mining, we do not try to find a complete (global) partitioning of the network. Instead, we consider local patterns describing local communities, so-called "nuggets" in the data, cf. [61]. The patterns should be as exceptional as possible with respect to a given community quality measure. The pattern formalization provides an intuitive description of the community, i.e., a characterization in terms of their descriptive features. This is usually not achieved by classical community mining methods that consider the nodes of a network (e. g., users in a social network) as mere strings or ids – and provide no easily interpretable description.

COMODO Algorithm. Our proposed approach combines local pattern mining using exceptional model mining, see Section 5.2, and community detection: We present an efficient algorithm for mining the top-k community patterns with respect to a number of standard community evaluation functions, i. e., the novel COMODO algorithm: Using *extended frequent pattern trees* [17], COMODO conducts an exhaustive search by traversing a representation of the pattern search space compiled into a *community pattern tree* (CP-tree). The CP-tree is a compact version of the dataset, that also contains relevant information about the graph structure. Using this tree, the patterns can be efficiently computed using only the information contained in the tree. For pruning, COMODO utilizes optimistic estimates of the community quality functions, in a branch-and-bound fashion. Therefore, we propose suitable optimistic estimates [47, 132] which are efficient to compute.

Our approach also tackles typical problems that are not addressed by standard approaches for community detection such as pathological cases like small community sizes. Furthermore, we focus on interpretable patterns that can easily be incorporated in a practical application, for example, for recommendations. Since in practice the entities in a network tend to belong to a number of different communities, the presented method captures overlapping community allocations.

We demonstrate our approach on networks from BibSonomy. The presented approach is not limited to social bookmarking systems and can be applied to any kind of graph-structured data for which additional descriptive features are available, e. g., certain activity in telephone networks or interactions in face-to-face contacts that also utilize tags or topic descriptions for the contained relations. The applied optimistic estimates allow a reduction of the search space by orders of magnitude, especially using the modularity quality function. Overall the proposed optimistic estimates show huge pruning potential for many applications, especially considering the local modularity measure as an effective tool for fast descriptive community mining.

5.2 Fast Exhaustive Exceptional Model Mining

In the context of descriptive pattern mining, the concept of *exceptional model mining* has recently been introduced [34, 68, 70]. It can be considered as a generalization of typical descriptive approaches like association rule mining [2], subgroup discovery [61] or frequent pattern mining [46], and enables more complex target properties, cf. [68]. Exceptional model mining tries to identify interesting patterns with respect to a local model derived from *a set* of attributes, e. g., a correlation or a linear regression model. The interestingness can be flexibly defined, e. g., by a significant deviation from a model that is derived from the total population or the respective complement set of instances within the population. There exist heuristic algorithms [67] for exceptional model mining; however, these cannot guarantee any optimality of its results, in contrast to exhaustive methods. Then, efficient exhaustive methods are required due to the size of the large (exponential) pattern search space. Possible applications include the identification of characteristic patterns [11, 13, 23], analysis of node information in social interaction networks [12, 81, 107], or descriptive community mining approaches [17].

GP-GROWTH Algorithm. In [71], we present the novel *GP-growth* algorithm that can be used for mining patterns with exceptional target models *exhaustively*. We propose the concept of valuation bases allowing us to derive a new algorithm capable of performing efficient exhaustive search for many different classes of exceptional models. We extend the well-known FP-tree data structure [49] by replacing the frequency information stored in each node of the tree by the more general concept of valuation bases: These are dependent on a specific *model class* and allow for an efficient computation of the target model parameters. Intuitively, in the generalized tree (*GP-tree*), each node stores a valuation basis that locally provides all information necessary to compute the target model parameters. We characterize the scope of the presented approach, describe properties of possible target models, and discuss its instantiations for model classes presented in literature: The applicability of the proposed approach is discussed by drawing an analogy to data stream mining, providing a constructive proof for the adaptation of the valuation base approach to other possible model classes.

Evaluation. An evaluation of the presented approach utilizes publicly available UCI data [95], as well as a social data from Flickr. Our runtime experiments show improvements of more than an order of magnitude in comparison to a naive exhaustive depth-first search. This enables the application in large social network datasets. As an example, we used a dataset obtained from flickr containing about 1.1 million instances and about 1200 tags that were used as describing attributes. Then, we aimed at identifying combinations of tags (as descriptions), for which their correlation is especially strong. As a result, even for a search depth of 2, a simple benchmark algorithm, i. e., depth-first-search, did not finish the task within two full days. In contrast, the same task performed by GP-growth finished in about 8 minutes, using a standard office PC with a 2.2 GHz CPU and 2 GB RAM. Furthermore, even for an increased search depth of 3, the task could be completed within 10 minutes. Thus, GP-GROWTH provides for a fast efficient method for exceptional model mining. Further speedups can then be achieved using optimistic estimates, e. g., as discussed above in Section 5.1.

5.3 Exploratory Pattern Mining on Social Media

Since descriptive pattern mining characterizes and summarizes the data, it can also be applied for hypothesis generation in order to derive semi-automatic and interactive exploratory approaches.

Exploratory Pattern Mining. In [15], we present such an interactive exploratory approach on social interaction networks incorporating location and tagging information. In this context, location information is considered as a proxy for social relatedness and interaction, cf. [3,58,85,92,106] for more details. Furthermore, using tagging information assigned to the nodes of the network, we can also explore tag-similarity measures, e. g., [29], as a proxy for the relatedness and interaction.

We present a two way perspective on exploring locations, tags, resources and their induced interaction networks: First, we aim to describe sets of related resources (e. g., photos) using location-information and tags, which are semantically related as well as focused on certain locations. We can imagine, for example, browsing the map of Germany and taking an overview on the general Berlin/Brandenburg area in terms of tag descriptions. Second, we characterize given locations using tagging patterns and photos for interactive browsing. A user may click on a map to specify his point of interest, for example, and is then provided a set of tags that are used for that region.

Implementation. We propose an iterative two step approach for the exploration of locations and resources: The first step uses pattern mining techniques, e. g., [13, 17] to automatically generate a candidate set of potentially interesting descriptive tags. For a flexible characterization of locations at different levels the search can be adapted by employing different location-based target measures for pattern mining. The result of the first step is thus a set of descriptive *interesting* patterns. In the second step, a human explores this candidate set of patterns and introspects interesting patterns manually by browsing and viewing various visualizations. The pattern mining parameters can be adapted in an exploratory fashion. In this way, we obtain an overview on the resources in terms of their location and describing tags. Furthermore, we can characterize different regions, areas or specific locations in terms of such descriptive information. The resulting patterns can be exploited by providing different visualizations and browsing options. Additionally, they can be filtered according to different interestingness criteria defined by the applied quality function. Furthermore, background knowledge, e. g., on semantically equivalent tags, can be manually refined and included in the process.

We demonstrate the impact and validity of the presented approach in a case study using publicly available data from the social photo sharing application *Flickr*. We apply COMODO for descriptive community mining (see Section 5.1) and the efficient method SD-Map* [13] for descriptive pattern mining. This could also be implemented using the exceptional pattern mining approach described in Section 5.2.

In the case study, we show on a structural level, that the proposed approach allows us to obtain more significant communities compared to standard community detection methods as a baseline. Furthermore, the interactive approach provides a good starting point for exploratory analysis as shown by an exemplary case study.

6 Conclusions and Outlook

Data mining on social interaction networks is a rather novel research area, especially considering human face-to-face contact networks. In this article, we presented several analyses and results, examining the interaction networks in order to improve our understanding of the data, the modeled behavior, and its underlying processes. Furthermore, we showed how to adapt, extend and apply known predictive data mining algorithms on social interaction networks. Additionally, we presented novel methods for descriptive data mining for uncovering and extracting relations and patterns for hypothesis generation and exploration by the user, in order to provide characteristic information about the data and networks. We introduced the CONFERATOR and MYGROUP applications for enhancing social interactions: Both systems have been applied in various conferences and workgroup events.

The presented data mining approaches on social interaction networks tackle several emerging research directions: First, with the increase of data in ubiquitous and social environments, the analysis and mining of this data becomes more important – and social interaction networks provide for a valuable modeling tool in that area. Additionally, with the increasing data volume in different pervasive services and applications, the handling of big data, i. e., large volumes of data, requires efficient algorithms. Furthermore, sensors are transcending into private and personal domains of life. The derivation and construction of social (interaction) networks based on the sensor measurements requires both systematic analysis and efficient and effective algorithms. The concept of *reality mining* [35, 89] is then a related research direction, as a general extension of several of the presented techniques and methods. Overall, these also open up opportunities towards the *ubiquitous web* [53, 112], and *collective intelligence* [59, 82].

References

1. Adamic, L.A., Adar, E.: Friends and Neighbors on the Web. Social Networks 25(3), 211–230 (2003)
2. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.) Proc. 20th Int. Conf. Very Large Data Bases, (VLDB). pp. 487–499. Morgan Kaufmann (1994)
3. Ahn, Y.Y., Han, S., Kwak, H., Moon, S., Jeong, H.: Analysis of Topological Characteristics of Huge Online Social Networking Services. In: Proc. 16th Intl. Conf. on the World Wide Web (WWW). pp. 835–844. ACM, New York, NY, USA (2007)
4. Alani, H., Szomszor, M., Cattuto, C., den Broeck, W.V., Correndo, G., Barrat, A.: Live Social Semantics. In: Intl. Semantic Web Conference (ISWC). pp. 698–714 (2009)
5. Albert, R., Barabási, A.: Statistical Mechanics of Complex Networks. Reviews of Modern Physics 74(1), 47 (2002)
6. Atzmueller, M.: Applied Natural Language Processing and Content Analysis: Advances in Identification, Investigation and Resolution, chap. Data Mining. IGI Global, Hershey, PA, USA (2011)
7. Atzmueller, M.: Mining Social Media: Key Players, Sentiments, and Communities. WIREs: Data Mining and Knowledge Discovery 2, 411–419 (2012)
8. Atzmueller, M.: Data Mining on Social Interaction Networks (2013), habilitation thesis, University of Kassel

9. Atzmueller, M., Becker, M., Doerfel, S., Kibanov, M., Hotho, A., Macek, B.E., Mitzlaff, F., Mueller, J., Scholz, C., Stumme, G.: Ubicon: Observing Social and Physical Activities. In: Proc. 4th IEEE Intl. Conf. on Cyber, Physical and Social Computing (CPSCom). (2012)
10. 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), 101–107 (2011)
11. Atzmueller, M., Benz, D., Hotho, A., Stumme, G.: Towards Mining Semantic Maturity in Social Bookmarking Systems. In: Proc. 4th Intl. Workshop on Social Data on the Web, 10th Intl. Semantic Web Conference. Online at CEUR-WS.org/Vol-830/ (2011)
12. Atzmueller, M., Doerfel, S., Hotho, A., Mitzlaff, F., Stumme, G.: Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles. In: Atzmueller, M., Chin, A., Helic, D., Hotho, A. (eds.) Modeling and Mining Ubiquitous Social Media, LNAI, vol. 7472. Springer, Berlin (2012)
13. Atzmueller, M., Lemmerich, F.: Fast Subgroup Discovery for Continuous Target Concepts. In: Proc. 18th International Symposium on Methodologies for Intelligent Systems (ISMIS). LNCS, vol. 5722, pp. 1–15. Springer, Heidelberg, Germany (2009)
14. Atzmueller, M., Lemmerich, F.: VIKAMINE - Open-Source Subgroup Discovery, Pattern Mining, and Analytics. In: Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD). LNCS, vol. 7524, pp. 842–845. Springer, Berlin (2012)
15. Atzmueller, M., Lemmerich, F.: Exploratory Pattern Mining on Social Media using Geo-References and Social Tagging Information. International Journal of Web Science 2(1/2) (2013)
16. 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)
17. Atzmueller, M., Mitzlaff, F.: Efficient Descriptive Community Mining. In: Proc. 24th International FLAIRS Conference. pp. 459 – 464. AAAI Press (2011)
18. Atzmueller, M., Roth-Berghofer, T.: The Mining and Analysis Continuum of Explaining Uncovered. In: Proc. 30th SGAI Intl. Conference on Artificial Intelligence (AI) (2010)
19. Balby Marinho, L., Hotho, A., Jäschke, R., Nanopoulos, A., Rendle, S., Schmidt-Thieme, L., Stumme, G., Symeonidis, P.: Recommender Systems for Social Tagging Systems. SpringerBriefs in Electrical and Computer Engineering, Springer (Feb 2012)
20. Barabasi, A.L.: Linked. The New Science of Networks (2002)
21. Barrat, A., Cattuto, C., Colizza, V., Pinton, J.F., den Broeck, W.V., Vespignani, A.: High Resolution Dynamical Mapping of Social Interactions with Active RFID 5(7) (2010)
22. 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: Proceedings Intl. Semantic Web Conference. Lecture Notes in Computer Science, vol. 6497, pp. 17–33 (2010)
23. Behrenbruch, K., Atzmüller, M., Evers, C., Schmidt, L., Stumme, G., Geihs, K.: A Personality Based Design Approach Using Subgroup Discovery. In: Winckler, M., Forbrig, P., Bernhaupt, R. (eds.) Human-Centered Software Engineering, LNCS, vol. 7623, pp. 259–266. Springer Berlin Heidelberg (2012)
24. Benz, D., Hotho, A., Jäschke, R., Krause, B., Mitzlaff, F., Schmitz, C., Stumme, G.: The Social Bookmark and Publication Management System BibSonomy – A Platform for Evaluating and Demonstrating Web 2.0 Research. J. VLDB 19(6), 849–875 (2010)
25. Brandes, U., Erlebach, T. (eds.): Network Analysis: Methodological Foundations [outcome of a Dagstuhl seminar, 13-16 April 2004], LNCS, vol. 3418. Springer (2005)
26. Brin, S., Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems 30, 107–117 (1998)

27. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J.: Graph Structure in the Web. Computer Networks 33(1-6), 309–320 (2000)
28. Burke, R., Gemmell, J., Hotho, A., Jäschke, R.: Recommendation in the Social Web. AI Magazine 32(3), 46–56 (2011)
29. Cattuto, C., Benz, D., Hotho, A., Stumme, G.: Semantic Grounding of Tag Relatedness in Social Bookmarking Systems. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T.W., Thirunarayan, K. (eds.) The Semantic Web – ISWC 2008, Proc.Intl. Semantic Web Conference 2008. LNAI, vol. 5318, pp. 615–631. Springer, Heidelberg (2008)
30. 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)
31. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step-by-step Data Mining Guide. Retrieved from http://www.crisp-dm.org/CRISPWP-0800.pdf
32. Chou, B.H., Suzuki, E.: Discovering Community-Oriented Roles of Nodes in a Social Network. In: Proc. Intl. Conf. on Data Warehousing and Knowledge Discovery (DaWak). pp. 52–64 (2010)
33. Comes, D.E., Evers, C., Geihs, K., Hoffmann, A., Kniewel, R., Leimeister, J.M., Niemczyk, S., Roßnagel, A., Schmidt, L., Schulz, T., Söllner, M., Witsch, A.: Designing Sociotechnical Applications for Ubiquitous Computing - Results from a Multidisciplinary Case Study. In: Proc. 12th IFIP International Conference Distributed Applications and Interoperable Systems (DAIS. pp. 194–201. Springer, Berlin (2012)
34. Duivesteijn, W., Knobbe, A., Feelders, A., van Leeuwen, M.: Subgroup Discovery Meets Bayesian Networks–An Exceptional Model Mining Approach. In: 10th IEEE Intl. Conference on Data Mining (ICDM). pp. 158–167. IEEE (2010)
35. Eagle, N., Pentland, A.S.: Reality Mining: Sensing Complex Social Systems. Pers. Ubiquit. Comput. 10(4), 255–268 (Mar 2006)
36. Etzioni, Amitai Etzioni, O.: Face-to-Face and Computer-Mediated Communities, A Comparative Analysis. The Information Society 15(4), 241–248 (1999)
37. Farzan, R., Brusilovsky, P.: Community-based Conference Navigator. In: Proc. SociUM Workshop, 11th Intl. Conf. User Modeling (UM), Corfu, Greece (2007)
38. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery in Databases. AI Magazine 17, 37–54 (1996)
39. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery: An Overview. In: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.) Advances in Knowledge Discovery and Data Mining, pp. 1–34. AAAI Press (1996)
40. Fortunato, S., Castellano, C.: Community Structure in Graphs (2007), arXiv:0712.2716 Chapter of Springer's Encyclopedia of Complexity and System Science
41. Freeman, L.: Segregation In Social Networks. Sociological Methods & Research 6(4), 411 (1978)
42. Gaertler, M.: Clustering. In: Brandes and Erlebach [25], pp. 178–215
43. Gatica-Perez, D.: Automatic Nonverbal Analysis of Social Interaction in Small Groups: A Review. Image Vis. Comput. (2009)
44. Girba, T., Kuhn, A., Seeberger, M., Ducasse, S.: How Developers Drive Software Evolution. In: International Workshop on Principles of Software Evolution. pp. 113–122. IEEE Computer Society, Los Alamitos, CA, USA (2005)
45. Girvan, M., Newman, M.E.J.: Community Structure in Social and Biological Networks. PNAS 99(12), 7821–7826 (June 2002)
46. Goethals, B.: Survey on Frequent Pattern Mining. Tech. rep. (2003), manuscript

47. Grosskreutz, H., Rüping, S., Wrobel, S.: Tight Optimistic Estimates for Fast Subgroup Discovery. In: Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD). pp. 440–456. Springer, Berlin (2008)
48. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd Edition. Morgan Kaufmann, San Francisco, USA (2006)
49. Han, J., Pei, J., Yin, Y.: Mining Frequent Patterns Without Candidate Generation. In: Intl. Conf. on Management of Data. pp. 1–12. ACM, New York, NY, USA (2000)
50. Heidemann, J.: Online Social Networks. Inf. Spektrum 33(3), 262 – 271 (2010)
51. Hightower, J., Vakili, C., Borriello, G., Want, R.: Design and Calibration of the SpotON Ad-Hoc Location Sensing System. Tech. rep., UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA, (2000)
52. Hindle, A., German, D.M., Holt, R.: What Do Large Commits Tell Us?: A Taxonomical Study of Large Commits. In: Proc. Intl. Working Conference on Mining Software Repositories. pp. 99–108. MSR '08, ACM, New York, NY, USA (2008)
53. Hotho, A., Stumme, G.: Towards the Ubiquitous Web. Semantic Web 1, 117–119 (2010)
54. Huang, Z., Li, X., Chen, H.: Link Prediction Approach to Collaborative Filtering. In: Proc. 5th ACM/IEEE-CS Joint Conference on Digital Libraries. pp. 141–142. JCDL '05, ACM, New York, NY, USA (2005)
55. Hui, P., Chaintreau, A., Scott, J., Gass, R., Crowcroft, J., Diot, C.: Pocket Switched Networks and Human Mobility in Conference Environments. In: Proc. ACM SIGCOMM Workshop on Delay-tolerant Networking. pp. 244–251. ACM, New York, NY, USA (2005)
56. 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. CoRR 1006.1260 (2010)
57. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag Recommendations in Social Bookmarking Systems. AI Communications 21(4), 231–247 (Dec 2008)
58. Kaltenbrunner, A., Scellato, S., Volkovich, Y., Laniado, D., Currie, D., Jutemar, E.J., Mascolo, C.: Far From the Eyes, Close on the Web: Impact of Geographic Distance on Online Social Interactions. In: Proc. ACM SIGCOMM Workshop on Online Social Networks (WOSN 2012). Helsinki, Finland (2012)
59. Kapetanios, E.: Quo Vadis Computer Science: From Turing to Personal Computer, Personal Content and Collective Intelligence. Data and Knowledge Engineering 67(2), 286 – 292 (2008), special Jubilee Issue: DKE 25 Years
60. Katz, L.: A New Status Index Derived from Sociometric Analysis. Psychometrika 18(1), 39–43 (March 1953)
61. Klösgen, W.: Explora: A Multipattern and Multistrategy Discovery Assistant. In: Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.) Advances in Knowledge Discovery and Data Mining, pp. 249–271. AAAI Press (1996)
62. Knobbe, A., Feelders, A., Leman, D.: Data Mining: Foundations and Intelligent Paradigms. Vol. 2, chap. Exceptional Model Mining, pp. 183–198. Springer, Berlin (2011)
63. Knorr-Cetina, K.: Sociality with Objects: Social Relations in Postsocial Knowledge Societies. Theory, Culture and Society 14(4), 1–43 (1997)
64. Kurgan, L.A., Musilek, P.: A Survey of Knowledge Discovery and Data Mining Process Models. The Knowledge Engineering Review 21, 1 – 24 (2006)
65. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a Social Network or a News media? In: Proc. 19th Intl. Conf. on World Wide Web (WWW). pp. 591–600. ACM (2010)
66. Lancichinetti, A., Fortunato, S.: Community Detection Algorithms: A Comparative Analysis. Physical Review E 80 (2009)
67. van Leeuwen, M.: Maximal Exceptions with Minimal Descriptions. Data Mining and Knowledge Discovery 21(2), 259–276 (2010)
68. van Leeuwen, M., Knobbe, A.J.: Diverse Subgroup Set Discovery. Data Mining and Knowledge Discovery 25(2), 208–242 (2012)

69. Leimeister, J.M.: Collective Intelligence. Business and Information Systems Engineering 2(4), 245–248 (2010)
70. Leman, D., Feelders, A., Knobbe, A.: Exceptional Model Mining. Machine Learning and Knowledge Discovery in Databases pp. 1–16 (2008)
71. Lemmerich, F., Becker, M., Atzmueller, M.: Generic Pattern Trees for Exhaustive Exceptional Model Mining. In: Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD). LNCS, vol. 7524, pp. 277–292. Springer, Berlin (2012)
72. Lerner, J.: Role Assignments. In: Network Analysis, Lecture Notes in Computer Science, vol. 3418, pp. 216–252. Springer, Heidelberg (2005)
73. Leskovec, J., Huttenlocher, D.P., Kleinberg, J.M.: Signed Networks in Social Media. In: Mynatt, E.D., Schoner, D., Fitzpatrick, G., Hudson, S.E., Edwards, W.K., Rodden, T. (eds.) Proc. SIGCHI Conference on Human Factors in Computing Systems. pp. 1361–1370. ACM, New York, NY, USA (2010)
74. Leskovec, J., Lang, K.J., Dasgupta, A., Mahoney, M.W.: Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters (2008), http://arxiv.org/abs/0810.1355
75. Leskovec, J., Lang, K.J., Mahoney, M.: Empirical Comparison of Algorithms for Network Community Detection. In: Proc. 19th Intl. Conference on World Wide Web (WWW). pp. 631–640. WWW '10, ACM, New York, NY, USA (2010)
76. Li, X., Chen, H.: Recommendation as Link Prediction: A Graph Kernel-based Machine Learning Approach. In: Proc. 9th ACM/IEEE-CS Joint Conference on Digital Libraries. pp. 213–216. JCDL '09, ACM, New York, NY, USA (2009)
77. Liben-Nowell, D., Kleinberg, J.M.: The Link Prediction Problem for Social Networks. In: Proc. 12th Intl. Conference on Information and Knowledge Management (CIKM). pp. 556–559. ACM, New York, NY, USA (2003)
78. Lim, E.M.: Patterns of Kindergarten Children's Social Interaction with Peers in the Computer Area. Intl. Journal of Computer-Supported Collaborative Learning (ijcscl) 7 (2012)
79. Lü, L., Zhou, T.: Link Prediction in Weighted Networks: The Role of Weak Ties. EPL (Europhysics Letters) 89, 18001 (2010)
80. Macek, B.E., Atzmueller, M., Stumme, G.: Profile Mining in CVS-Logs and Face-to-Face Contacts for Recommending Software Developers. In: Proc. IEEE 3rd Intl. Conf. on Social Computing (Socialcom). pp. 250–257. IEEE Computer Society, Boston, MA, USA (2011)
81. Macek, B.E., Scholz, C., Atzmueller, M., Stumme, G.: Anatomy of a Conference. In: Proc. 23rd ACM Conference on Hypertext and Social Media. pp. 245–254. ACM Press, New York, NY, USA (2012 (Awarded with the ACM Douglas Engelbart Best Paper Award))
82. Malone, T.W., Laubacher, R., Dellarocas, C.: Harnessing Crowds: Mapping the Genome of Collective Intelligence. Tech. rep., Center for Collective Intelligence, Massachusetts Institute of Technology (2009)
83. Malone, T.W., Laubacher, R., Dellarocas, C.: The Collective Intelligence Genome. Sloan Management Review 5(3), 21 – 31 (2010)
84. Mannila, H.: Theoretical Frameworks for Data Mining. SIGKDD Explor. Newsl. 1(2), 30–32 (Jan 2000)
85. McGee, J., Caverlee, J.A., Cheng, Z.: A Geographic Study of Tie Strength in Social Media. In: Proc. 20th ACM International Conference on Information and Knowledge Management. pp. 2333–2336. CIKM '11, ACM, New York, NY, USA (2011)
86. Meriac, M., Fiedler, A., Hohendorf, A., Reinhardt, J., Starostik, M., Mohnke, J.: Localization Techniques for a Mobile Museum Information System. In: Proc. WCI (2007)
87. Mislove, A., Koppula, H.S., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Growth of the Flickr Social Network. In: Proc. 1st Workshop on Online Social Networks. pp. 25–30. WOSN '08, ACM, New York, NY, USA (2008)

88. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and Analysis of Online Social Networks. In: Proc. 7th ACM SIGCOMM Conference on Internet Measurement. pp. 29–42. IMC '07, ACM, New York, NY, USA (2007)
89. Mitchell, T.M.: Mining Our Reality. Science 326(5960), 1644–1645 (Dec 2009)
90. Mitzlaff, F., Atzmueller, M., Benz, D., Hotho, A., Stumme, G.: Community Assessment using Evidence Networks. In: Atzmueller, M., Hotho, A., Chin, A., Helic, D. (eds.) Analysis of Social Media and Ubiquitous Data, LNAI, vol. 6904, pp. 79–98. Springer, Heidelberg, Germany (2011)
91. Mitzlaff, F., Atzmueller, M., Benz, D., Hotho, A., Stumme, G.: User-Relatedness and Community Structure in Social Interaction Networks. CoRR/abs 1309.3888 (2013)
92. Mitzlaff, F., Atzmueller, M., Stumme, G., Hotho, A.: Semantics of User Interaction in Social Media. In: Ghoshal, G., Poncela-Casasnovas, J., Tolksdorf, R. (eds.) Complex Networks IV, Studies in Computational Intelligence, vol. 476. Springer, Berlin (2013)
93. 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: Proc. 21st ACM Conference on Hypertext and Hypermedia. ACM, New York, NY, USA (2010)
94. Murata, T., Moriyasu, S.: Link Prediction of Social Networks Based on Weighted Proximity Measures. In: Web Intelligence. pp. 85–88 (2007)
95. Newman, D., Hettich, S., Blake, C., Merz, C.: UCI Repository of Machine Learning Databases, http://www.ics.uci.edu/mlearn/mlrepository.html (1998)
96. Newman, M.E., Girvan, M.: Finding and Evaluating Community Structure in Networks. Phys Rev E Stat Nonlin Soft Matter Phys 69(2), 1–15 (2004)
97. Newman, M.E.J.: The Structure and Function of Complex Networks. SIAM Review 45(2), 167–256 (2003)
98. Newman, M.E.J.: Detecting Community Structure in Networks. Europ Physical J 38 (2004)
99. Newman, M.: Finding Community Structure in Networks Using the Eigenvectors of Matrices. Physical Review E 74(3), 36104 (2006)
100. Newman, M., Park, J.: Why Social Networks are Different from Other Types of Networks. Physical Review E 68(3), 36122 (2003)
101. Ni, L.M., Liu, Y., Lau, Y.C., Patil, A.P.: LANDMARC: Indoor Location Sensing Using Active RFID. Wireless Networks 10(6), 701–710 (2004)
102. Papadimitriou, A., Symeonidis, P., Manolopoulos, Y.: Friendlink: Link Prediction in Social Networks via Bounded Local Path Traversal. In: 2011 International Conference on Computational Aspects of Social Networks (CASoN). pp. 66–71. IEEE (Oct 2011)
103. Petra Kralj Novak, Nada Lavrac, G.I.W.: Supervised Descriptive Rule Discovery: A Unifying Survey of Contrast Set, Emerging Pattern and Subgroup Mining. Journal of Machine Learning Research 10, 377–403 (2009)
104. Russell, M.A.: Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites. O'Reilly Media, 1 edn. (2011)
105. Salathé, M., Bengtsson, L., Bodnar, T.J., Brewer, D.D., Brownstein, J.S., Buckee, C., Campbell, E.M., Cattuto, C., Khandelwal, S., Mabry, P.L., Vespignani, A.: Digital Epidemiology. PLOS Computational Biology 8(7), e1002616 (07 2012)
106. Scellato, S., Noulas, A., Lambiotte, R., Mascolo, C.: Socio-spatial Properties of Online Location-based Social Networks. Proc. 5th International Conference on Weblogs and Social Media (ICWSM) pp. 329–336 (2011)
107. Scholz, C., Atzmueller, M., Stumme, G.: On the Predictability of Human Contacts: Influence Factors and the Strength of Stronger Ties. In: Proc. Fourth ASE/IEEE International Conference on Social Computing (SocialCom). IEEE Computer Society, Boston, MA, USA (2012)

108. Scholz, C., Doerfel, S., Atzmueller, M., Hotho, A., Stumme, G.: Resource-Aware On-Line RFID Localization Using Proximity Data. In: Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD). LNCS, vol. 6913, pp. 129–144. Springer, Berlin (2011)
109. Scott, J.: Social Network Analysis: Developments, Advances, and Prospects. Social Network Analysis and Mining 1, 21–26 (2011)
110. Scripps, J., Tan, P.N., Esfahanian, A.H.: Exploration of Link Structure and Community-Based Node Roles in Network Analysis. In: Proc. 7th IEEE International Conference on Data Mining (ICDM). pp. 649–654. IEEE Computer Society, Washington, DC, USA (2007)
111. 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. pp. 26–35. ACM, New York, NY, USA (2007)
112. Sheth, A.: Computing for Human Experience: Semantics-Empowered Sensors, Services, and Social Computing on the Ubiquitous Web. IEEE Internet Computing 14(1), 88–91 (2010)
113. Siersdorfer, S., Sizov, S.: Social Recommender Systems for Web 2.0 Folksonomies. In: HT09: Proc. 20th ACM Conf. on Hypertext and Hypermedia. pp. 261–270. ACM, New York, NY, USA (2009)
114. Smith, M.A., Shneiderman, B., Milic-Frayling, N., Rodrigues, E.M., Barash, V., Dunne, C., Capone, T., Perer, A., Gleave, E.: Analyzing (Social Media) Networks with NodeXL. In: Carroll, J.M. (ed.) Proc. 4th Intl. Conf. on Communities and Technologies. pp. 255–264. ACM, University Park, PA, USA (2009)
115. Söllner, M., Hoffmann, A., Hoffmann, H., Leimeister, J.M.: How to Use Behavioral Research Insights on Trust for HCI System Design. In: Proc. CHI Conference on Human Factors in Computing Systems (Extended Abstracts Volume). pp. 1703–1708. ACM, New York, NY, USA (2012)
116. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J., Quaggiotto, M., Van den Broeck, W., Régis, C., Lina, B., Vanhems, P.: High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. PLOS ONE 6(8), e23176 (08 2011)
117. Stehle, J., Voirin, N., Barrat, A., Cattuto, C., Colizza, V., Isella, L., Regis, C., Pinton, J.F., Khanafer, N., den Broeck, W.V., Vanhems, P.: Simulation of an SEIR Infectious Disease Model on the Dynamic Contact Network of Conference Attendees. BMC Medicine 9(87) (2011)
118. Strogatz, S.H.: Exploring Complex Networks. Nature 410(6825), 268–276 (Mar 2001)
119. Szomszor, M., Cattuto, C., den Broeck, W.V., Barrat, A., Alani, H.: Semantics, Sensors, and the Social Web: The Live Social Semantics Experiments. In: Proc. 7th Extended Semantic Web Conference (ESWC 2010). LNCS, vol. 6089, pp. 196–210. Springer, Berlin (2010)
120. Tang, L., Liu, H.: Community Detection and Mining in Social Media. Synthesis Lectures on Data Mining and Knowledge Discovery 2(1), 1–137 (2010)
121. Tantipathananandh, C., Berger-Wolf, T.Y.: Constant-Factor Approximation Algorithms for Identifying Dynamic Communities. In: Proc. ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 827–836. ACM, New York, NY, USA (2009)
122. Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabási, A.L.: Human Mobility, Social Ties, and Link Prediction. In: Proc. ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 1100–1108. ACM, New York, NY, USA (2011)
123. Wang, K., Jiang, Y., Tuzhilin, A.: Mining Actionable Patterns by Role Models. In: Proc. 22nd International Conference on Data Engineering (ICDE). pp. 16–. ICDE '06, IEEE Computer Society, Washington, DC, USA (2006)
124. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. No. 8 in Structural Analysis in the Social Sciences, Cambridge University Press, 1 edn. (1994)

125. Watts, D.J., Strogatz, S.H.: Collective Dynamics of 'small-world' Networks. Nature 393(6684), 440–442 (june 1998)
126. Webb, G.: Discovering Significant Patterns. Machine Learning 71, 131–131 (2008)
127. Weiser, M.: The Computer for the 21st Century. Scientific American 265(3), 66–75 (January 1991)
128. Weiser, M.: Some Computer Science Issues in Ubiquitous Computing. Commun. ACM 36(7), 74–84 (1993)
129. Weschsler, D.: Concept of Collective Intelligence. American Psychologist 26(10), 904–907 (1971)
130. Wirth, R., Hipp, J.: CRISP-DM: Towards a Standard Process Model for Data Mining. In: Proc. 4th Intl. Conference on the Practical Application of Knowledge Discovery and Data Mining. pp. 29–39. Morgan Kaufmann (2000)
131. 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)
132. 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). pp. 78–87. Springer, Berlin (1997)
133. Xu, B., Chin, A., Wang, H., Chang, L., Zhang, K., Yin, F., Wang, H., Zhang, L.: Physical Proximity and Online User Behavior in an Indoor Mobile Social Networking Application. In: Proc. 4th IEEE Intl. Conf. on Cyber, Physical and Social Computing (CPSCom) (2011)
134. Zhong, N., Liu, J., Yao, Y.: In Search of the Wisdom Web. Computer 35(11), 27–31 (2002)
135. Zhou, T., Lu, L., Zhang, Y.C.: Predicting Missing Links via Local Information. The European Physical Journal B - Condensed Matter and Complex Systems 71, 623–630 (2009)
136. Zhuang, H., Chin, A., Wu, S., Wang, W., Wang, X., Tang, J.: Inferring Geographic Coincidence in Ephemeral Social Networks. In: Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD). pp. 613–628 (2012)