Focus Article

Mining social media: key players, sentiments, and communities

Martin Atzmueller*

Social media is the key component of social networks and organizational social applications. The emergence of new systems and services has created a number of novel social and ubiquitous environments for mining information, data, and, finally, knowledge. This connects but also transcends private and business applications featuring a range of different types of networks and organizational contexts. Important structures concern subgroups emerging in those applications as communities (connecting people), roles and key actors in the networks and communities, and opinions, beliefs, and sentiments of the set of actors. Collective intelligence can then be considered as an emerging phenomenon of the different interactions. This focus article considers mining approaches concerning social media in social networks and organizations and the analysis of such data. We first summarize important terms and concepts. Next, we describe and discuss key actor identification and characterization, sentiment mining and analysis, and community mining. In the sequel we consider different application areas and briefly discuss two exemplary ubiquitous and social applications—the social conference guidance system Conferator, and the MyGroup system for supporting working groups. Furthermore, we describe the VIKAMINE system for mining communities and subgroups in social media in the sketched application domains. Finally, we conclude with a discussion and outlook.

How to cite this article:
WIREs Data Mining Knowl Discov 2012, 2: 411–419 doi: 10.1002/widm.1069

INTRODUCTION

Social networks and social media services grow dramatically: Google+ that only started in June 2011 has shown an exponential growth (10 million users in 16 days, and 20 million in 24 days).[1] The importance of social networks and social media—both in a private but also organizational setting, can be noted at several levels. The impact of social media has been shown on several large-scale evolutions such as the Arab spring or the UK riots. A number of current events have also been twittered first before the 'classic media' distributed the information, for example, the emergency landing of a US Airways' Flight 1549 in the Hudson River.[2] From a business perspective, social media is being applied for marketing and product management, for identifying customer opinions or aggregating knowledge from their organizational portals. In internal company portals, social media can also help in finding experts and for recommending knowledgeable persons.

What constitutes the fascination of social media? Social networks and social media are currently enabling technologies for a variety of applications considering individuals, groups, and organizations. Social media captures user-generated content of 'classic' social services, for example, Twitter; but also content and data generated by more ubiquitous systems, for example, by radio-frequency identification (RFID)-based applications.[3–6] In this way, social media includes a number of other dimensions that are present in sensor networks, mobile devices, and the ubiquitous Web. From an application perspective, the analysis of communities plays a dominant role for determining common interests or special competences. These can then be applied, for example, for personalization[7,8] or for generating recommendations.[3,5,7] In addition, from a business perspective there is often the question how product-related customer opinions can be rated, ranked, and assessed.[9,10]

*Correspondence to: atzmueller@cs.uni-kassel.de

Knowledge and Data Engineering Group, University of Kassel, Kassel, Germany

DOI: 10.1002/widm.1069

How can we extract interesting information, patterns, and knowledge from this set of local, distributed, and diverse social media data that are created by human communication and interaction—also including data from ubiquitous and sensor-enabled devices? How can we mine social media for meaningful patterns or entities, for complementing intelligent information systems? In the following, we first define important concepts and terms, before we consider specific techniques and methods. After that, we discuss exemplary applications and describe the VIKAMINE tool in more detail. Finally, we conclude with a discussion and outlook.

DEFINITIONS

In the following, we introduce some important terms, that is, social media, user-generated content, social networks and social network analysis (SNA), and finally communities.

Social Media

Kaplan and Haenlein define social media as a set of internet-based applications that apply principles of the Web 2.0.[11] In addition, these systems are based on user-generated content, that is, content that is being produced and shared by users online. We adopt a similarly intuitive definition and consider social media as Internet-based systems and services in the ubiquitous Web that use and provide all kinds of social data of human interaction and communication. This includes data from sensor networks and mobile devices, as long as the data is created by real users.

User-Generated Content

Social media services such as blogs, microblogging services (Twitter), wiki-based applications (Wikipedia) or resource-sharing services (YouTube) are just a few examples for applications that allow users to create content, annotate it, and share it with their (personal) network. In turn, friends and contacts can comment the shared content and communicate in this way. Such interactions let communities form and grow.

Social Networks and Social Network Analysis

In the context of social media, we consider social (online) networks as a special sort of technical platform, for which the interaction and communication between the actors is provided based on the infrastructure of the Internet. The connecting elements of the different communities are, for example, common goals, interests, or other needs.[12] These examples show the strong connection to (online) communities that manifest such connecting elements. As a core concept SNA[13] considers a set of nodes (actors) and their relationships (connections), which are modeled by links between the nodes. For example, we could consider friendship relationships or the 'Follower' relation in Twitter. Using SNA, we can analyze, for example, how human contact patterns[14] advance or inhibit the distribution of diseases (e.g., flu), from an epidemiological point of view.[15]

Communities

Communities essentially represent very densely connected sets of nodes in a social network, similar to clusters or dense 'subgroups'. The links or connections between the nodes can denote or be motivated by different interests, needs, and so on. In this way, communities denote better connected clusters in the network that have strong connection (links) to each other. There exist different definitions of communities; the above notion includes the core of most definitions focusing on the density of the connections within the community.[16–22] In general, usually not only the density within the community is assessed, but the connection density of the community is also compared to the density of the rest of the network.[22] Then, cuts between communities are established in such a way as to maximize the community evaluation function.[20–22]

METHODS

For the analysis of social media we can consider its content in more detail, for example, using natural language processing (NLP) techniques, or linguistic analysis. Furthermore, we can focus on the induced network structure of the utilized social media data, for example, for a friendship network we consider all friendship links in the network.

In the following, we consider three methods for mining social networks and organizations in more detail: We first describe SNA techniques for the identification and characterization of key players. After that, we focus on assessing opinions of social media content: Sentiment mining considers the content in more detail and is applied for extracting opinions, moods, or sentiments from social media. Finally, we focus on the mining, analysis and assessment of communities in social media: Community mining aims at identifying densely connected groups (communities); it mainly utilizes the (network) structure of the applied social media.

Key Players and Roles

Key players are actors that are important for the network in terms of connectivity, number of contacts, and the paths that are passing through the corresponding node. The identification and characterization of key players is interesting for prestige and reputation mining, identifying hubs in the network and for social monitoring. The assessment can happen on different layers: For instance, we can consider the network as a whole for discovering individual roles.[13] In addition, we can consider roles in specific communities.[23] Finally, descriptive pattern mining methods for characterization can be applied to both layers.[24,25]

For the first case, standard SNA methods can be applied for the analysis concerning the complete network structure,[13] for example, for determining the mean path length between nodes, or for discovering the diameter of the network. In addition, on the level of the whole network we can determine different centrality measures, to identify important nodes (hubs).[4,13] Examples are given by the degree centrality as the number of connections to the neighbors of the node, the betweenness centrality as the number of shortest path of all node pairs that go through a specific node, or the closeness centrality that considers the length of these shortest paths. For the degree and betweenness centrality, high values indicate a higher importance, whereas the reverse is true for the closeness centrality.[13]

Role mining concerning communities mainly considers the relations between the communities for a specific actor. Scripps et al.[23] present a method for assessing roles considering 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[26] present a similar method considering given communities for a community-oriented analysis.

Although the above methods mainly focus on the network and community structure, a simple characterization or description is usually not provided by standard methods for role and community mining. To this end, the characterization of actors and their roles is provided by descriptive pattern mining techniques[4,24,25] that utilize different centrality measures and allow the characterization of role-specific nodes given the respective centrality measures. Descriptive pattern mining focuses on identifying exceptional patterns for a certain target concept, for example, belonging to a certain role; the mined patterns consist of a combination of features, in our case, for example, individual centrality measures. In this way, roles can get an intuitive interpretation given the network characteristics.

Sentiment Mining and Analysis

Sentiment mining (sentiment analysis, opinion analysis)[9,10,27] aims at extracting subjective information from textual data using NLP, linguistic methods, and text mining approaches. It has a broad range of applications, for example, in customer relationship management, brand analysis, and market research. Liu[27] provides an overview on key aspects of sentiment mining and opinion analysis and distinguishes the following elements: The *opinion holder*, the *object* of the *opinion*, the *features*, their synonyms, and different *opinion orientations* (e.g., positive/negative) of these. Then, the mining aims at identifying such opinion elements in textual information.

Thus, on a general level sentiment mining focuses on identifying the polarity (positive/negative) of a text (fragment) considering the aspects above. On a more detailed level the polarity can be differentiated in more detail, for example, distinguishing 'sad' or 'disappointed' in the negative case or 'happy' or 'cheerful' for a positive sentiment. Furthermore, sentiments can be detected on different levels, for words, phrases, sentences, text passages, or entire (portions of) documents.

The applied methods for sentiment analysis and opinion mining are based on machine-learning techniques, for example, latent semantic analysis and support-vector machines, which are used for learning a sentiment classificatory.[28] In addition, also lexicon-based approaches are applied, which contain specific sentiment values for individual words. Then, the individual sentiment values of the contained words are aggregated to obtain the total sentiment value of a text or text passage. For evaluation the predictions are usually compared to human annotations (usually by a number of different annotators) to assess the classification accuracy.

Applications of sentiment mining and analysis range from the classification of product comments, feature-based opinion summaries of different products, to an assessment of blog entries. Hu and Liu,[29] for example, describe an approach for mining and summarizing customer reviews which integrates sentiment analysis and text summarization techniques. A practical view on the problem of identifying spam in user-generated content is described by Jindal and Liu[30] again exemplified by the domain of customer product reviews.

Community Mining, Analysis, and Assessment

Community mining aims at discovering and analyzing (cohesive) subgroups, clusters, or communities that are 'densely' connected with each other in a network. Techniques for the mining of communities include graph-based approaches,[18] clustering according to features of the nodes, or pattern mining techniques for optimizing a community evaluation function,[16] which we describe in the applications section below. The core idea of the evaluation function is to apply an objective evaluation criterion, for example, 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 locally, or together with other community-splits globally,[16,20,21] respectively.

The discovered communities can then be applied for recommendations,[5,7,8,31] facetted browsing[7] or for personalization of intelligent systems,[3,6,7,31] for which the community assignment is made explicit or the information about other members of the community is used implicitly for the adaptation of the application. This is implemented, for example, for friend recommendations in social networks such as Google+ or Facebook, for personalization in Web search,[32] or for facetted browsing. Furthermore, interesting topics and contacts can be recommended, for example, in ubiquitous applications such as Conferator described below. For the evaluation and assessment of communities usually manually collected test data and user studies are applied, which are often expensive to conduct. For data-centric approaches[33] secondary data, that is, available secondary networks are applied for a relative comparison of communities; this allows for a simple and cost-effective assessment of communities.

APPLICATIONS AND TOOLS

There is a variety of social systems using social media. These include, for example, microblogging services (Twitter), wiki-based applications (Wikipedia), resource-sharing services (YouTube, Flickr), social conferencing applications (Conference Navigator,[13,31] Conferator[3,4,6]) or bookmarking and publication management systems such as BibSonomy.[34] With the growing availability of low-cost sensor and mobile devices, more and more ubiquitous and social applications emerge that combine 'classic' social media with sensor data in a broad sense, that is, also including data from mobile devices such as smartphones.

Examples for the application of social media mining are intelligent social monitoring, reputation management, customer profiling, product management, and advertising. Sentiment mining, for example, can be applied for identifying positive or negative product assessments to improve product and/or reputation management. Pattern mining can be applied for role and key actor characterization, for example, for describing spammers in social bookmarking systems[25] or for describing the maturity of tags being used in social media.[35] Furthermore, community mining can be applied for identifying implicit communities (e.g., people with similar interests). These can then be used for product placement and advertising, to recommend products that people with similar interests bought, or products that were bought by friends. Similarly, explicit or implicit communities can be applied for a number of recommendation options, for example, publications, interesting news or contacts. Social conferencing systems such as Conference Navigator[31,36] and Conferator[3,4,6] apply community detection, for example, for personalization and recommendations—in the context of talks, contacts, news, or other resources.

The described applications can be generalized and applied to many other application contexts as well, in which the discovery and analysis of key actors, roles, and communities plays an important part. Collaborative systems of companies and organizations, for social media-driven organizational networks such as wikis or version control systems[5] allow a broad range of further applications. The identification of process chains, identification of experts, and workflow optimization provide further interesting application examples.

In the following, we briefly discuss two exemplary ubiquitous and social applications, that is, the Conferator[3,6] as a social conference guidance system, and the MyGroup application for making working environments smarter using collective intelligence techniques.[5] From a tool perspective, we also briefly describe the VIKAMINE[24] system for analysis and mining of communities and subgroups in social media and briefly summarize exemplary results from applications utilizing data from last.fm, Flickr, BibSonomy, and Conferator.

Conferator/MyGroup

The emergence of social media and ubiquitous computing has created new environments where large groups work collectively together using electronic media to accomplish certain tasks in an intelligent way. This is often referred to as 'Collective Intelligence',[37] and can be applied to tackle a number of problems, for example, group memory, organizational design, or new technologies for making groups smarter. Conferator and MyGroup are two systems[3,5] using social media for collective intelligence. Conferator offers conference participants the option to organize and manage their social contacts during conferences: For this purpose, active SocioPattern RFID tags are applied that allow to localize participants and to collect their face-to-face contacts. For these, the system allows the setup of a complete profile, the annotation of their own social contacts, social networking to other participants, and the management and personalization of the conference schedule. A similar application, MyGroup allows the support of social communication in the context of working groups using social interaction—awareness by utilizing the same technology.

Conferator has been applied at a number of conferences, for example, at the LWA 2010 and LWA 2011 conferences of the German association of computer science, and at the ACM Hypertext 2011. MyGroup is continuously running in the Knowledge and Data Engineering research group at the University of Kassel, and has also been applied at a number of different events, for example, at a software development code camp to enhance social interactions and communication in developing software. Several of the applied data mining methods are based on the community mining and key actor analysis techniques described above: The created social media data, for example, are used for applying different methods for the discovery of professional communities, and for generating a number of useful recommendations: Conferator recommends interesting contacts and people that have similar interests or a similar 'conferencing' behavior.[3] For this purpose data mining is applied on the contact graph of all conference participants, and other social network and publication data. An expert recommendation component[5] provides a mining approach for identifying software developers for question answering. The method utilizes the MyGroup contact graph and individual conversations (contacts) between the respective software developers and additionally their check-in behavior in version control systems. Both approaches can be applied for a faster identification of professional contacts. Influence and roles in a conference context and their characterizations can then be analyzed[4] considering, for example, professional roles in this context. In an academic context, these are typically the roles 'Professor', 'Post-Doc', 'PhD Student', and 'Graduate/Undergraduate Student'.

At the LWA 2010 conference an analysis[4] showed, for example, that professors usually have higher centrality values compared to the other groups. An exception is given by the betweenness centrality, for which the Post-Docs showed the highest scores. This is an indication of their potential influence in hierarchical relationships, for which they are contained in many shortest paths between the participants, and thus have a very important function as bridges between these. Tables 1 and 2 show exemplary analysis results of the LWA 2010 and the Hypertext 2011 conferences. While Table 1 focuses on the characterization of different status roles given a set of network properties, Table 2 describes the community-oriented roles by different conference-oriented properties. In both cases, the VIKAMINE tool for pattern mining and subgroup analytics was applied for obtaining these characteristic patterns.

Pattern Mining with VIKAMINE

VIKAMINE[24] is an open environment for pattern mining and analytics including subgroup discovery and community mining. According to the methods discussed above, it can be applied for key actor and role characterization and descriptive community mining. Figure 1 shows a screenshot of the main user interface.

VIKAMINE features a variety of state-of-the-art techniques for subgroup discovery, pattern mining, and analytics. Subgroup discovery aims at detecting interesting subsets of a population according to a given target property of interest. This task can be flexibly formalized using different interestingness measures, for example, relating to increased shares of a property in a subgroup compared to the whole dataset. Community mining can be implemented by targeting subgroups that are more densely connected than expected by chance, thus implementing a local modularity measure.[16,18]

VIKAMINE has been successfully applied for a broad range of applications in mining social media. In the context of social bookmarking systems, for example, VIKAMINE can be applied for obtaining descriptive profiles of spammers, that is, for their characterization.[25] The mined patterns

TABLE 1 | Exemplary Results of Pattern Analysis for Characterizing Nonorganizers and PhD Students at the LWA 2010 Conference[4]

Target	#	Lift	*P*	Size	Description
Nonorganizer	1	1.06	0.88	51	$clo = \{low; medium\}$
	2	1.05	0.87	61	$eig^* = \{low; medium\}$
	3	1.04	0.86	59	$deg = \{low; medium\}$
	4	1.10	0.92	12	$clo = \{low; medium\}$ AND $deg = \{high; medium\}$
	5	1.12	0.93	30	$clo = \{high; medium\}$ AND $eig^* = \{low; medium\}$
PhD student	1	1.07	0.54	59	bet $= \{high; low\}$
	2	1.07	0.54	48	$str = \{high; low\}$
	3	1.14	0.58	26	$deg = high$
	4	1.31	0.67	12	$bet = \{high; low\}$ AND $eig^* = high$
	5	1.38	0.70	20	$deg = high$, AND $bet = \{high; low\}$
	6	1.58	0.80	10	$deg = high$, AND $bet = \{high; low\}$ AND $eig^* = \{high; low\}$

The rows of the table show the lift of the pattern comparing the fraction of nonorganizers and PhD students covered by the respective pattern compared to the fraction of the whole dataset, the size of the pattern extension (number of described nonorganizers/PhD students), and the description itself. *Clo*, *eig*, *deg*, *bet*, and *str* denote the closeness, eigenvector, degree, betweenness, and weighted degree (strength) centralities, respectively, with the values *low*, *medium*, *high*.

TABLE 2 | Exemplary Results of Pattern Analysis at the Hypertext 2012 Conference for Describing Individual Roles Using Pattern Mining and Subgroup Discovery[6]

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

The rows of the table show the lift of the pattern comparing the fraction of ambassadors/bridges covered by the respective pattern compared to the fraction of the whole dataset, the size of the pattern extension (number of participants), and the description itself. The patterns are described by (combinations) of properties of the participants, e.g., being session chairs or having a strong affiliation to the Hypertext conference.

capturing certain spammer subgroups can provide explanations and justifications for marking or resolving spammer candidates in the social bookmarking system BibSonomy.[34] For a variety of social media applications it is usually useful to identify high-quality tags, that is, tags with a certain maturity.[35] Pattern mining using VIKAMINE was applied for obtaining maturity profiles of tags based on a set of graph centrality features on the tag—tag co-occurrance graph, which are simple to compute and to assess. For community mining, we implemented a descriptive community mining approach[16] as a plugin. Furthermore, VIKAMINE implements analysis and mining methods for geotagged social media for discovering descriptive tags for certain geolocations.[38] VIKAMINE provides a plugin with a specialized user interface for handling, presenting, and visualizing geoinformation. The interactive exploration also can utilize background knowledge concerning the provided tags, which is entered either in a textual or graphical form. Furthermore, descriptive community mining (as described above) is enabled using a specialized plugin. Some illustrative results of a last.fm case study are shown in Table 3.

CONCLUSION

This paper has presented concepts, approaches, and applications for mining social media focussing on key

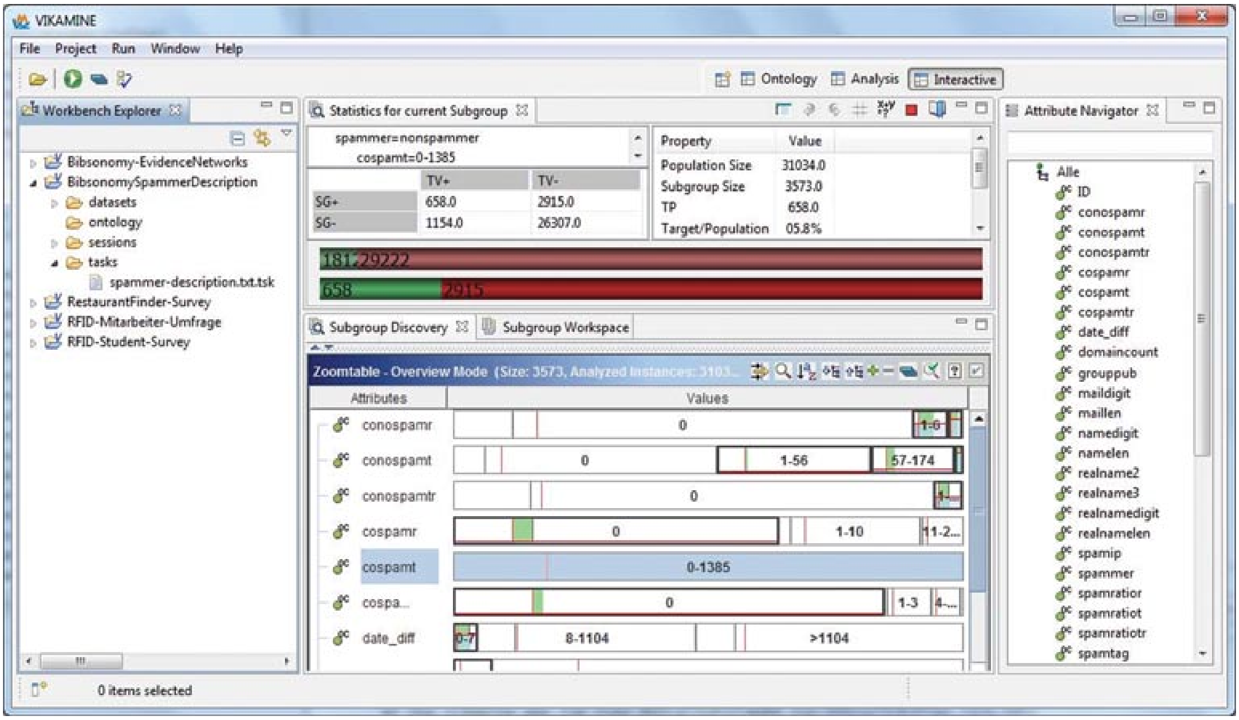

FIGURE 1 | Main user Interface of the VIKAMINE[24] system (available at http://www.vikamine.org). The screenshot shows the workbench explorer on the left, and the attribute navigator on the right. The panel in the middle contains the zoomtable showing value distributions of individual attributes below the current pattern view including a statistical characterization.

TABLE 3 | Example of Different Community Patterns Denoting Densely Connected Subgroups from a last.fm Case Study

Size	Community Description
519	80s
240	gregorian_chant AND 80s
215	girl_groups AND 80s
171	atmospheric
122	synth_pop
32	psychedelic AND minimal
16	psychedelic AND 80s
10	psychedelic AND brit_rock AND classic_rock
10	death_rock AND minimal AND 80s
10	death_rock AND 80s AND doom_metal

The individual communities are described by conjunctions of tags assigned by the respective users contained in the communities.

players, sentiments, and communities. With all information and data that is being collected in social media systems, and the continuously growing number of users, there is still an increasing number of options and challenges for the analysis of such data. Extended personalization and recommendations are just two exemplary applications of mining social networks and organizations.

As a further step of extensively extending mining social media to mobile and ubiquitous environments, 'reality mining'[39,40] applies data mining on 'everyday' sensor data comprehensively, including smartphones, and ubiquitous sensor networks. However, there is a growing challenge between technical analysis options, user interests, and their privacy. Legal guidelines for system design,[41] functions for anonymization,[42] and privacy-preserving data mining methods[41,43] provide approaches for including privacy aspects into such intelligent systems.

With sufficient privacy measures, data mining methods can then provide a better understanding of communication, interaction, and collective processes. By integrating the environment, context, and actions of users, social media mining can then support the users during their everyday life in better achieving their goals.

REFERENCES

1. Leon Haland/Google+ 20 mil. Available at: http://goo.gl/S7rNf. (Accessed January 16, 2011).
2. Janis Krums. Miracle on the hudson. Available at: http://twitpic.com/135xa. (Accessed January 16, 2011).
3. Atzmueller M, Benz D, Doerfel S, Hotho A, Jäschke R, Macek BE, Mitzlaff F, Scholz C, Stumme G. Enhancing Social Interactions at Conferences. *Inf Technol* 2011, 3:101–107.
4. Atzmueller M, Doerfel S, Mitzlaff F, Hotho A, Stumme G. Face-to-face contacts at a conference: dynamics of communities and roles. Modeling and mining ubiquitous social media. Heidelberg, Germany: Springer Verlag; 2012.
5. Macek BE, Atzmueller M, Stumme G. Profile mining in cvs-logs and face-to-face contacts for recommending software developers. In: *Proceedings of IEEE Third International Conference on Social Computing (SocialCom)*, Boston, MA; 2011.
6. Macek BE, Scholz C, Atzmueller M, Stumme G. Anatomy of a conference. In: *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*. New York, NY: ACM; 2012.
7. Boratto L, Chessa A, Agelli M, Clemente ML. Group recommendation with automatic identification of users communities. In: *IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies*. Milano, Italy; 2009, 547–550.
8. Gargi U, Lu W, Mirrokni V, Yoon S. Large-Scale Community Detection on YouTube for Topic Discovery and Exploration. In: *Proceedings of 5th International AAAI Conference on Weblogs and Social Media*. Menlo Park, CA: AAAI Press; 2011.
9. Liu B. *Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing*. London, UK: Chapman & Hall; 2010.
10. Pang B, Lee L. Opinion mining and sentiment analysis. *Found Trends Inf* 2008, 2:1–135.
11. Kaplan A, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. *Bus Horiz* 2010, 53:59–68.
12. Heidemann J. Online social networks. *Inf Spektrum* 2010, 33:262–271.
13. Wassermann S, Faust K. *Social Network Analysis*. Cambridge: Cambridge University Press; 1994.
14. Isella L, Stehlé J, Barrat A, Cattuto C, Pinton JF, den Broeck WV. What's in a crowd? Analysis of face-to-face behavioral networks. *J Theor Biol* 2011, 271:166–180.
15. Stehlé J, Voirin N, Barrat A, Cattuto C, Colizza V, Isella L, Régis C, Pinton J-F, Khanafer N, Van den Broeck W, Vanhems P. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Med* 2011, 9:1–15.
16. Atzmueller M, Mitzlaff F. Efficient descriptive community mining. In: *Proceedings of 24th International FLAIRS Conference*. New York, NY: AAAI Press; 2011, 459–464.
17. Fortunato S, Castellano C. *Encyclopedia of Complexity and System Science*. Heidelberg: Springer; 2007.
18. Girvan M, Newman MEJ. Community structure in social and biological networks. *Proc Natl Acad Sci USA* 2002, 99, 7821–7826.
19. Lancichinetti A, Fortunato S. Community detection algorithms: a comparative analysis. *Phys Rev E* 2009, e80. doi:10.1103/PhysRevE.80.056117.
20. Leskovec J, Lang KJ, Dasgupta A, Mahoney MW: Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters. *Internet Math* 2009, 6:29–123.
21. Leskovec J, Lang KJ, Mahoney MW. Empirical comparison of algorithms for network community detection. In: *Proceedings of 19th International Conference on the World Wide Web (WWW)*. New York, NY: ACM Press; 2010, 631–640.
22. Newman MEJ. Detecting community structure in networks. *Eur Phys J* 2004, 38:321–330.
23. Scripps J, Tan PN, Esfahanian AH. Exploration of link structure and community-based node roles in network analysis. In: *Proceedings of ICDM* 2007; 649–654.
24. Atzmueller M, Lemmerich F. VIKAMINE - open-source subgroup discovery, pattern mining, and analytics. Proc. ECML/PKDD 2012: European conference on machine learning and principles and practice of knowledge discovery in databases. Heidelberg, Germany: Springer Verlag; 2012.
25. Atzmueller M, Lemmerich F, Krause B, Hotho A. Who are the spammers? Understandable local patterns for concept description. *In: Proceedings of 7th Conference on Computer Methods and Systems*. Krakow, Poland; 2009.
26. Chou BH, Suzuki E. Discovering community-oriented roles of nodes in a social network. In: *Proceedings of DaWak 2010*. Bilbao, Spain; 2010, 52–64.
27. Liu B. Sentiment analysis: a multifaceted problem. *IEEE Intell Syst* 2010, 25:1–5.
28. Esuli A, Sebastiani F. Sentiment quantification. *IEEE Intell Syst* 2010, 25:72–75.
29. Hu M, Liu B. Mining and summarizing customer reviews. In: *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM Press; 2004.

30. Jindal N, Liu B. Opinion spam and analysis. In: *Proceedings of 1st ACM International Conference on Web Search and Data Mining (WSDM)*. New York, NY: ACM Press; 2008.
31. Farzan R, Brusilovsky P. Community-based conference navigator. Proceedings of 1st workshop on adaptation and personalisation in social systems: groups, teams, communities at the 11th international conference on user modeling, UM 2007. Corfu, Greece, 2007, 30–39.
32. Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. comput. *Networks ISDN* 1998, 30:107–117.
33. Mitzlaff F, Atzmueller M, Benz D, Hotho A, and Stumme G. Community assessment using evidence networks. In: *Analysis of Social Media and Ubiquitous Data*. Heidelberg: Springer; 2011.
34. Benz D, Hotho A, Jäschke R, Krause B, Mitzlaff F, Schmitz C, Stumme G. The Social bookmark and publication management system bibsonomy. *VLDB J* 2010, 19:849–875.
35. Atzmueller M, Benz D, Hotho A, Stumme G. Towards mining semantic maturity in social bookmarking systems. *In: Proceedings of 4th International Workshop on Social Data on the Web*. Bonn, Germany: ISWC; 2011.
36. Wongchokprasitti C, Brusilovsky P, Parra D. Conference Navigator 2.0: community-based recommendation for academic conferences. In: *Proceedings of Workshop on Social Recommender Systems (at IUI 2010)*. Hong Kong, China; 2010.
37. Malone TW, Laubacher R, Dellarocas C. Harnessing crowds: mapping the genome of collective intelligence. *MIT Center for Collective Intelligence, Working Paper 2009-001;* 2009.
38. Lemmerich F, Atzmueller M. Modeling location-based profiles of social image media using explorative pattern mining. In: *Proceedings IEEE SocialCom 2011, Workshop on Modeling Social Media MSM 2011*. Boston: IEEE Computer Society; 2011.
39. Eagle N, Pentland AS. Reality mining: sensing complex social systems. *Pers Ubiquit Comput* 2006, 10:255–268.
40. Mitchell TM. Mining our reality. *Science* 2009, 326:1644–1645.
41. Navarro Bullock B, Lerch H, Roßnagel A, Hotho A, Stumme G. Privacy-aware spam detection in social bookmarking systems. In: *Proceedings of 11th International Conference on Knowledge Management and Knowledge Technologies*. New York: ACM; 2011, 15:1–15:8.
42. Samarati P. Protecting respondents' identities in microdata release. In: *IEEE Trans Knowl Data Eng;* 2001, 1010–1027.
43. Atzori M, Bonchi F, Giannotti F, Pedreschi D: Anonymity preserving pattern discovery. *VLDB J* 2008, 17:703–727.

