Mining Social Interaction – Ubiquitous Sensors and Social Media

Martin Atzmueller

Interdisciplinary Research Center for Information System Design
Knowledge and Data Engineering Group, University of Kassel
Willhelmshöher Allee 73, 34121 Kassel, Germany
`atzmueller@cs.uni-kassel.de`

Abstract

Ubiquitous social media has woven itself into the very fabric of everyday life. This results in a dramatic increase of social data. Essentially, various links between the users and their associated artifacts capture different facets of interactions and relationships. The emergence of ubiquitous and mobile devices enables the capture of such interactions also in offline contexts, and the generation of according interactions networks. This paper focuses on the analysis and mining of such interaction networks, specifically focussing on data collected using ubiquitous RFID sensors as well as complementing social media. We discuss analysis results in the context of real-world systems.

Keywords: data mining, social media, social computing, social network analysis, rfid

MSC: 68Q87, 68T05, 68U35

1. Introduction

The analysis of roles [48, 56] and community structure [38, 11, 10], link prediction [12, 29], and investigating the relation between online and offline networks [15, 4] are exemplary prominent research topics in mining ubiquitous data and social media analytics. In this paper, we summarize work on mining social interactions using ubiquitous sensors and social media; we adopt a rather intuitive definition of social media, regarding it as online systems and services in the ubiquitous web. The key point is social data generated by human interaction

and communication [2]. In this way, we analyze user interaction formalized in so-called social interaction networks [4, 3]: These refer to user-related social networks in social media that are capturing social relations inherent in social interactions, social activities and other social phenomena which act as proxies for social user-relatedness.

While there has been foundational work on the analysis of face-to-face contact networks, e. g., [16, 15, 50], data mining and analytics 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. For those, we apply data from different real-world systems and applications, e. g., [7, 5, 6, 35].

The remainder of this paper is structured as follows: We start with a discussion of offline and online interactions, exemplified using real-world social systems. Next, we sketch approaches for role analysis [8, 32] and temporal dynamics [2, 9] in social interaction networks. After that, we summarize methods for community detection and analysis on heterogenous interaction networks [11, 25]. Then, we discuss adaptations, extensions, and applications of predictive modeling approaches for multi-modal link prediction on evolving networks [46, 45, 44, 47]. Finally, we conclude with a summary and outlook.

2. Social Media and Social Interaction Networks

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 [2]. Therefore, 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; we include interaction data from sensors and mobile devices, as long as the data is created by real users.

We focus on *social interaction networks* [4, 3], 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 [53, p. 37 ff.] social interaction networks focus on *interaction* relations between *people* as the corresponding actors. Networks in ubiquitous and social applications, for example, in RFID-based systems, can then be derived according to the detected contacts between the respective RFID tags, possibly weighted with the number of contacts, or their durations. Such social interation networks are often observed during certain events, for example, during conferences [7, 32, 54].

CONFERATOR [7][1] is a social conference guidance system for efficiently managing face-to-face contacts at a conference, and collectively building a personalized conference program. CONFERATOR is built on top of the RFID-based proximity sensing system developed by the SocioPatterns[2] collaboration. The applied RFID

---

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

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.

3. Role Analysis and Temporal Dynamics

The analysis of interaction and groups, and their evolution, respectively, are prominent topics in social sciences, e. g., [52, 18, 25, 47, 9]. The temporal evolution of contact networks is analyzed, for example, in [14, 13]. 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.

Role mining, e. g., [48, 56] concerning communities mainly analyzes the relations between them for a specific actor. [48] 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 [17] for such a community-oriented analysis.

In this context, we analyze different time-based link patterns using offline and online information. We ground user-interaction and community structure accordingly using different online and offline properties in [8]. 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, for example, 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., [22].

In a threshold-based analysis, e. g., using different minimal contact durations of the contact data, we analyze general structural properties of the contact network, investigate the stability and dynamics of community structures, and examine different explicit and implicit roles of conference participants, cf. [8, 32].

4. Community Analytics

Community mining is a prominent approach for identifying densely connected subgroups of the nodes contained in a network, e. g., [39, 20, 26, 55]. 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.

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 [38], e. g., using the modularity [38, 39], the segregation index [21] or the conductance [27] as an evaluation function. Then, cuts between communities are established in such a way as to maximize the community 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. There are a variety of algorithms, e. g., *InfoMap* [43, 42], *Label Propagation* [40], or *Leading Eigenvector* [39]. The discovered communities can then be applied, for example, for recommendations or for personalization.

The CONFERATOR system, for example, provides the respective community to the user in a special section of the site. This enables a personal discovery of the own community. In addition, description-oriented community mining methods can be applied, such that the communities can be characterized in terms of common interests, or other describing features. In [11], we present such 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., [37, 26, 28], that only identify communities as subsets of users. The proposed 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. For longitudinal effects, e. g., [25] provides details on the dynamics of structural properties of face-to-face contact networks, including the predictability of community membership.

5. Link Prediction

Link prediction [12, 29] 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. Link prediction in social interaction networks has a number of prominent applications, including the prediction of missing links, cf. [29], for improving collaborative filtering or for recommending new contacts.

There is already a large body of research for link prediction concerning *online* social networks, e. g., [1, 29, 36, 31, 24]. 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., [30], mainly rely on questionaires, diaries, or recordings, and usually only consider rather small groups,

cf. [53, p. 45 ff.]. Different factors contribute to the creation of links, e. g., existing connections [46], and their respective tie strengths [31]. Identifying connections between evolutionary processes, link creation, and factors for strengthening existing ties, are also of great interest.

In [46, 45], we focus on real-world offline networks of *human contacts*, that is, *face-to-face* contacts between persons. Our application context is given by academic conferences, e. g., [32]. We aim to predict *new* contacts based on network properties, as an adaptation of methods for online social networks. 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. 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*, i. e., longer contacts, for new and recurring contacts. The results of the analysis indicate that stronger ties have a rather strong influence on the contact behavior and the prediction performance, cf. [46, 45]. 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. 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.

On a structural level, we furthermore study triadic closures in face-to-face contact networks [45]. The fundamental concept of a triadic closure was first introduced in [41]: Essentially, the basic principle of triadic closure is that there is an increased likelihood that two people become friends in the future, if they have a friend in common. Our analysis showed that the link strength and academic status of participants has influences on closing triangles (depending on the type of the conference, e. g., for smaller and larger ones, and their specific interaction structures, respectively); we also complemented the analyses by considering reversed triadic closures, cf. [45]. In addition, we analyze the predictability of links grounded using different online and offline information [44]. Our results show semantically grounded consistent community and role structure. In addition, we observe that a set of heterogenous networks, i. e., different online and offline networks can complement each other well for improving link prediction methods, also cf. [51].

Overall, the results of the analysis provides interesting insights especially concerning the impact of the contact durations, the strength of stronger ties, and the impact of the network structure.

6. Conclusions and Outlook

This paper summarized work on data mining and analytics on social interactions using ubiquitous sensors and social media: We first discussed offline and online

social interactions. Then, we summarized recent methods and analysis results, from roles and temporal dynamics to community analytics and link prediction methods.

Overall, our analysis results are not only relevant for gaining justifications and important insights into structural and semantic features of social interaction networks. They can also help, e. g., for implementing further new link mining, community detection or user recommendation algorithms.

For future work, we aim to leverage our analysis results in order to embed the indicators, patterns, influence factors, and predictors into more advanced models in the context of offline and online interaction networks. Furthermore, we plan to extend the analysis towards more dynamic approaches including heterogenous network data with extended attribute information. As a general extension, the concept of *reality mining* [19, 34] is a related research direction of several of the presented techniques and methods. Overall, these also open up opportunities towards the *ubiquitous web* [49], and *collective intelligence* [33, 23].

References

[1] Adamic, L.A., Adar, E.: Friends and Neighbors on the Web. Social Networks 25(3), 211–230 (2003)

[2] Atzmueller, M.: Mining Social Media: Key Players, Sentiments, and Communities. WIREs: Data Mining and Knowledge Discovery 2, 411–419 (2012)

[3] Atzmueller, M.: Analyzing and Grounding Social Interaction in Online and Offline Networks. In: Proc. ECML/PKDD 2014: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. LNCS, vol. 8726, pp. 485–488. Springer, Berlin (2014)

[4] Atzmueller, M.: Data Mining on Social Interaction Networks. JDMDH 1 (June 2014)

[5] 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)

[6] Atzmueller, M., Becker, M., Kibanov, M., Scholz, C., Doerfel, S., Hotho, A., Macek, B.E., Mitzlaff, F., Mueller, J., Stumme, G.: Ubicon and its Applications for Ubiquitous Social Computing. New Review of Hypermedia and Multimedia 20(1), 53–77 (2014)

[7] 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 53(3) (2011)

[8] Atzmueller, M., Doerfel, S., Hotho, A., Mitzlaff, F., Stumme, G.: Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles. In: Modeling and Mining Ubiquitous Social Media, LNAI, vol. 7472. Springer, Berlin (2012)

[9] Atzmueller, M., Ernst, A., Krebs, F., Scholz, C., Stumme, G.: On the Evolution of Social Groups During Coffee Breaks. In: Proc. WWW 2014 (Companion). IW3C2 / ACM (2014)

[10] Atzmueller, M., Lemmerich, F.: Exploratory Pattern Mining on Social Media using Geo-References and Social Tagging Information. IJWS 2(1/2) (2013)

[11] Atzmueller, M., Mitzlaff, F.: Efficient Descriptive Community Mining. In: Proc. 24th International FLAIRS Conference. pp. 459 – 464. AAAI Press, Palo Alto, CA, USA (2011)

[12] Barabasi, A.L.: Linked. The New Science of Networks (2002)

[13] Barrat, A., Cattuto, C., Colizza, V., Gesualdo, F., Isella, L., Pandolfi, E., Pinton, J.F., Rava, L., Rizzo, C., Romano, M., Stehle, J., Tozzi, A., Broeck, W.: Empirical Temporal Networks of Face-to-Face Human Interactions. Eur. Phys. J. Special Topics 222(6), 1295–1309 (2013)

[14] Barrat, A., Cattuto, C.: Temporal Networks, chap. Temporal Networks of Face-to-Face Human Interactions. Understanding Complex Systems, Springer (2013)

[15] Barrat, A., Cattuto, C., Szomszor, M., den Broeck, W.V., Alani, H.: Social Dynamics in Conferences: Analyses of Data from the Live Social Semantics Application. In: Proc. ISWC. LNCS, vol. 6497, pp. 17–33. Springer, Berlin (2010)

[16] 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)

[17] Chou, B.H., Suzuki, E.: Discovering Community-Oriented Roles of Nodes in a Social Network. In: Proc. DaWaK. pp. 52–64 (2010)

[18] Coleman, J.: Foundations of Social Theory. Belknap Press of Harvard Univ. Press, Cambridge, Mass. (2000)

[19] Eagle, N., Pentland, A.S.: Reality Mining: Sensing Complex Social Systems. Pers. Ubiquit. Comput. 10(4), 255–268 (Mar 2006)

[20] Fortunato, S., Castellano, C.: Community Structure in Graphs (2007), arxiv:0712.2716 Chapter of Springer's Encyclopedia of Complexity and System Science

[21] Freeman, L.: Segregation In Social Networks. Sociological Methods & Research 6(4), 411 (1978)

[22] Isella, L., Stehle, 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)

[23] 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

[24] Katz, L.: A New Status Index Derived from Sociometric Analysis. Psychometrika 18(1), 39–43 (March 1953)

[25] Kibanov, M., Atzmueller, M., Scholz, C., Stumme, G.: Temporal Evolution of Contacts and Communities in Networks of Face-to-Face Human Interactions. Science China Information Sciences 57(3) (2014)

[26] Lancichinetti, A., Fortunato, S.: Community Detection Algorithms: A Comparative Analysis. Physical Review E 80 (2009)

[27] 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)

[28] Leskovec, J., Lang, K.J., Mahoney, M.: Empirical Comparison of Algorithms for Network Community Detection. In: Proc. WWW. pp. 631–640. WWW '10, ACM, New York, NY, USA (2010)

[29] Liben-Nowell, D., Kleinberg, J.M.: The Link Prediction Problem for Social Networks. In: CIKM 2003. pp. 556–559. ACM, New York, NY, USA (2003)

[30] 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)

[31] Lü, L., Zhou, T.: Link Prediction in Weighted Networks: The Role of Weak Ties. EPL (Europhysics Letters) 89, 18001 (2010)

[32] Macek, B.E., Scholz, C., Atzmueller, M., Stumme, G.: Anatomy of a Conference. In: Proc. ACM Hypertext. pp. 245–254. ACM, New York, NY, USA (2012)

[33] 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)

[34] Mitchell, T.M.: Mining Our Reality. Science 326(5960), 1644–1645 (Dec 2009)

[35] Mitzlaff, F., Atzmueller, M., Hotho, A., Stumme, G.: The Social Distributional Hypothesis. Journal of Social Network Analysis and Mining 4(216) (2014)

[36] Murata, T., Moriyasu, S.: Link Prediction of Social Networks Based on Weighted Proximity Measures. In: Web Intelligence. pp. 85–88 (2007)

[37] 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)

[38] Newman, M.E.J.: Detecting Community Structure in Networks. Europ Physical J 38 (2004)

[39] Newman, M.: Finding Community Structure in Networks Using the Eigenvectors of Matrices. Physical Review E 74(3), 36104 (2006)

[40] Raghavan, U.N., Albert, R., Kumara, S.: Near Linear Time Algorithm to Detect Community Structures in Large-Scale Networks. Physical Review E 76(3), 036106 (2007)

[41] Rapoport, A.: Spread of Information through a Population with Socio-Structural Bias: I. Assumption of Transitivity. Bulletin of Mathematical Biology 15(4), 523–533 (Dec 1953)

[42] Rosvall, M., Axelsson, D., Bergstrom, C.T.: The Map Equation. The European Physical Journal Special Topics 178, 13–23 (2009)

[43] Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences 105(4), 1118–1123 (2008)

[44] Scholz, C., Atzmueller, M., Barrat, A., Cattuto, C., Stumme, G.: New Insights and Methods For Predicting Face-To-Face Contacts. In: Proc. ICWSM. AAAI, Palo Alto, USA (2013)

[45] Scholz, C., Atzmueller, M., Kibanov, M., Stumme, G.: How Do People Link? Analysis of Contact Structures in Human Face-to-Face Proximity Networks. In: ASONAM 2013. IEEE, Los Alamitos, CA, USA (2013)

[46] Scholz, C., Atzmueller, M., Stumme, G.: On the Predictability of Human Contacts: Influence Factors and the Strength of Stronger Ties. In: Proc. SocialCom. IEEE (2012)

[47] Scholz, C., Atzmueller, M., Stumme, G.: Predictability of Evolving Contacts and Triadic Closure in Human Face-to-Face Proximity Networks. Journal of Social Network Analysis and Mining 4(217) (2014)

[48] Scripps, J., Tan, P.N., Esfahanian, A.H.: Exploration of Link Structure and Community-Based Node Roles in Network Analysis. In: Proc. ICDM. pp. 649–654 (Oct 2007)

[49] 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)

[50] 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)

[51] Tang, J., Lou, T., Kleinberg, J.: Inferring Social Ties Across Heterogenous Networks. In: Proc. WSDM. pp. 743–752. ACM, New York, NY, USA (2012)

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

[53] 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)

[54] 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. CPSCom (2011)

[55] Yang, J., McAuley, J.J., Leskovec, J.: Community Detection in Networks with Node Attributes. In: Proc. IEEE ICDM. pp. 1151–1156. IEEE, Dallas, TX, USA (2013)

[56] Zhang, J., Tang, J., Zhuang, H., Leung, C.W., Li, J.: Role-Aware Conformity Modeling and Analysis in Social Networks. In: Proc. AAAI. pp. 958–965. AAAI, Palo Alto, CA, USA (2014)