Analyzing and Grounding Social Interaction in Online and Offline Networks

Martin Atzmueller

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

Abstract. In social network analysis, there are a variety of options for investigating social interactions. This paper reviews our recent work on analyzing and grounding social interactions in online and offline networks considering distributional semantics, structural network correlation and network inter-dependencies. Specifically, we focus on the analysis of user relatedness, community structure, and relations on online and offline networks. We discuss findings and results that justify the use of even implicitly accruing social interaction networks for the analysis of user-relatedness, community structure, etc. Furthermore, we provide insights into recent work on analyzing and grounding offline social networks.

Keywords: social network analysis, social interaction networks, mining social media, distributional semantics, community structure, social distributional hypothesis.

1 Introduction

The analysis of user relatedness [10, 13, 14], community structure [11, 12, 15], and the relation between online and offline networks [7, 16] are prominent research topics in data mining and social network analysis. In this context, this paper summarizes our recent work on analyzing and grounding social interaction. We analyze user interaction formalized in so-called social interaction networks [2, 14]: 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. Essentially, social interaction networks focus on *interaction* relations between *people*, see [19, p. 37 ff.], that are the corresponding actors.

First, we present the *social distributional hypothesis* [13] – a pragmatic proxy for homophily [10] – stating that users with similar interaction characteristics are related, and provide supporting evidence. Second, we extend this to the analysis of *communities* [11, 12] showing structural correlations between implicit networks of user interactions. Third, we investigate the *structural grounding* considering both online and offline network properties [5, 16]. In this way, we provide novel insights into the *grounding of offline behavior*. Our analysis results justify the analysis of even implicitly accruing social interaction networks with respect to user-relatedness, semantics and community structure, and provide for valuable insights, e. g., for the development of link analysis methods, community detection, and the connection of online and offline information.

T. Calders et al. (Eds.): ECML PKDD 2014, Part III, LNCS 8726, pp. 485–488, 2014.

2 Analysis of 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. 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 [1, 3]. 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.

In the following, we focus on the analysis of such social interaction networks. Figure 1 provides an overview on the analysis and grounding approaches, while Table 1 further summarizes the methods, applied techniques and results which we discuss below in more detail, reviewing our recent work. In particular, we propose the *social distributional hypothesis* [13] stating, that users with similar interaction characteristics tend to be related. Considering users as (social) entities, their distributional characteristics can be observed utilizing social interaction networks. The social distributional hypothesis is postulated similar to the *distributional hypothesis* [8] in linguistics; it states that words with similar distributional characteristics tend to be semantically related, i. e., that words occurring in similar contexts have a similar meaning.

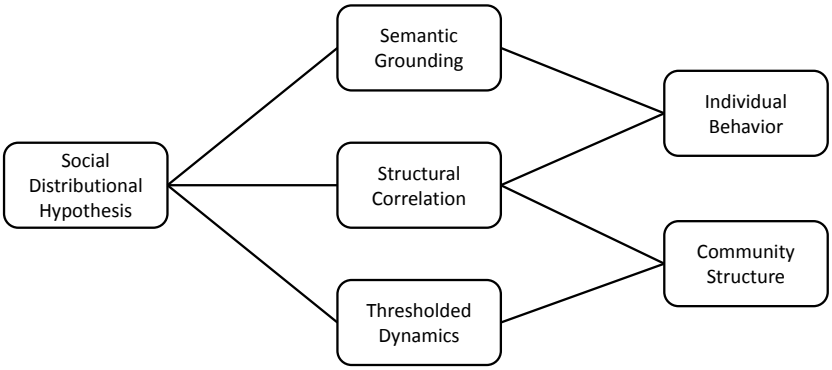

Fig. 1. Overview on the analysis and grounding setup: Starting with the social distributional hypothesis, we apply several methods for analyzing individual behavior and community structure

In [13, 14] we conduct a series of experiments on social interaction networks from Twitter, Flickr and BibSonomy and investigate the user-relatedness concerning the interactions, their frequency, and the specific interaction characteristics. The results indicate interrelations between structural similarity of interaction characteristics and semantic relatedness of users, supporting the social distributional hypothesis. This also grounds methods for analyzing social interaction networks in general.

On a structural level, we investigate two further issues in [11, 12] on the social interaction networks: Are there interrelations and correlations between the interaction networks? Furthermore, can these be applied for the analysis and data-driven assessment of communities? We analyze general structural properties of the obtained networks and comparatively discuss major structural characteristics in order to show that

Table 1. Overview on the applied methods and specific analysis techniques

Analysis	Method	Results
Semantic Grounding [13, 14]	Similarity Covariate Analysis (tag-based, location-based)	Interrelations between structural similarity of interaction characteristics and semantic user-relatedness
Structural Correlation [11, 12]	Degree Correlation, Neighborhood, Graph Covariance	Structural inter-network correlations; consistent community structure and ranking accross networks
Time-based Link Patterns [5, 9, 16]	Community Analysis, Role Analysis, Link Prediction	Semantically consistent community and role structures; indicators for complementing network structures

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 dependencies of the networks' neighborhood, and inter-network correlations. The results indicate strong correlations and interrelations between the considered social interaction networks, that are strong enough for inferring reciprocal conclusions between the networks. Based on these results, we propose an approach for (relative) community assessment based on the idea of *reconstructing existing social structures* [18] for the assessment and evaluation of a given clustering.

Furthermore, for analyzing and grounding offline networks we focus on real-world offline networks of *human contacts*, that is, *face-to-face* proximity contacts between persons in [5, 9, 16]. In contrast to virtual networks, the involved contacts were collected using the social conference guidance system CONFERATOR [4] – a ubiquitous RFID-based system that allows us to collect face-to-face contact data [6]. Thus, we can observe and analyze (offline) social interaction at a very detailed level, including the specific event sequences and durations. Also, we complement the analysis of the offline social interaction networks with additional node-level properties and further networks, e. g., utilizing the DBLP co-authorship relations. 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 [5]. 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 [17] of conference participants. Furthermore, we analyze the predictability of links grounded using different online and offline information [16]. Our results show semantically grounded consistent community and role structure. In addition, we observe that different online and offline networks can complement each other well for improving link analysis methods, e. g., concerning link prediction.

3 Conclusions and Outlook

We proposed the social distributional hypothesis as one foundational issue for the analysis of social interaction networks and presented supporting experimental results. Furthermore, we successfully investigated structural correlation and time-based link patterns on online and offline social interaction networks. Overall, our analysis results are not only relevant for gaining justifications and important insights into structural and semantic relations for social interaction networks. They can also help, e. g., for implementing new link mining, community detection or user recommendation algorithms.

References

1. Atzmueller, M.: Mining Social Media: Key Players, Sentiments, and Communities. WIREs: Data Mining and Knowledge Discovery 2, 411–419 (2012)
2. Atzmueller, M.: Data Mining on Social Interaction Networks. JDMDH 1 (June 2014)
3. Atzmueller, M.: Social Behavior in Mobile Social Networks: Characterizing Links, Roles and Communities. In: Mobile Social Networking, pp. 65–78. Springer, Berlin (2014)
4. Atzmueller, M., Benz, D., Doerfel, S., Hotho, A.: Enhancing Social Interactions at Conferences. IT 53(3) (2011)
5. 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.) MUSE 2011 and MSM 2011. LNCS, vol. 7472, pp. 21–39. Springer, Heidelberg (2012)
6. Barrat, A., Cattuto, C., Colizza, V., Pinton, J., den Broeck, W.V., Vespignani, A.: High Resolution Dynamical Mapping of Social Interactions with Active RFID. PLoS ONE 5(7) (2010)
7. Barrat, A., Cattuto, C., Szomszor, M., Van den Broeck, W., Alani, H.: Social Dynamics in Conferences: Analyses of Data from the Live Social Semantics Application. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part II. LNCS, vol. 6497, pp. 17–33. Springer, Heidelberg (2010)
8. Harris, Z.S.: Distributional Structure. Word (1954)
9. Macek, B.-E., Scholz, C., Atzmueller, M., Stumme, G.: Anatomy of a Conference. In: Proc. Hypertext, pp. 245–254. ACM, New York (2012)
10. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology 27(1), 415–444 (2001)
11. Mitzlaff, F., Atzmueller, M., Benz, D., Hotho, A., Stumme, G.: Community Assessment Using Evidence Networks. In: Atzmueller, M., Hotho, A., Strohmaier, M., Chin, A. (eds.) MUSE/MSM 2010. LNCS, vol. 6904, pp. 79–98. Springer, Heidelberg (2011)
12. Mitzlaff, F., Atzmueller, M., Benz, D., Hotho, A., Stumme, G.: User-Relatedness and Community Structure in Social Interaction Networks. CoRR/abs 1309.3888 (2013)
13. Mitzlaff, F., Atzmueller, M., Hotho, A., Stumme, G.: The Social Distributional Hypothesis. Journal of Social Network Analysis and Mining, accepted) (2014)
14. 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, pp. 13–25. Springer, Heidelberg (2013)
15. Newman, M.E.J.: Detecting Community Structure in Networks. Europ Physical J. 38 (2004)
16. 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 (2013)
17. 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 (October 2007)
18. Siersdorfer, S., Sizov, S.: Social Recommender Systems for Web 2.0 Folksonomies. In: Proc. Hypertext, pp. 261–270. ACM, New York (2009)
19. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications, 1st edn. Structural Analysis in the Social Sciences, vol. (8). Cambridge University Press (1994)