Towards Predictive Policing: Knowledge-based Monitoring of Social Networks

Michael Spranger, Florian Heinke, Steffen Grunert and Dirk Labudde
University of Applied Sciences Mittweida
Mittweida, Germany
Email: {*name.surname*}@hs-mittweida.de

***Abstract*—Increasing the resilience of the society against disorders, such as disasters, attacks or threatening groups, is a major challenge. Recent events highlight the importance of a resilient society and steps which are required to be taken in resilience engineering. *A priori* the optimal way to handle such adverse events is to prevent them, or at least provide appropriate courses of preparation. The essential requirement for every kind of preparation is information about relevant upcoming events. Such information can be gained for example from social networks and can form the basis for a long-term and short-term strategic planning by security forces. For that purpose, we here propose an application framework for knowledge-based social network monitoring, which aims at predicting short-term activities, as well as the long-term development of potentially dangerous groups. In this work, a theoretical outline of this approach is given and discussed.**

Keywords–forensic; text processing; resilience engineering

I. INTRODUCTION

The representation and the communication via the Internet, especially in social networks, have become a standard not only for individuals, companies and organizations but also for political groups or gangs using these platforms for planning, appointing and conducting criminal offences [1], [2]. Large events with a relatively large degree of group dynamics, like sport events, demonstrations or festivals, require a high expenditure of staff on the side of the security forces because of unpredictability and uncertainty of associated dynamics. For example, to secure the soccer events in 2014 in Germany approximately two million working hours of police officers were necessary [3]. In order to support decision makers, we outline an application framework for monitoring cliques and groups in social networks, which can be key elements in the emergence of critical events. The monitoring process is facilitated by means of employing general domain-specific endangerer profiles. Such a profile can be deduced from a set of social network sites of known endangerers or perpetrators (in the strict sense). Identifying suspicious activities is realized by group recommendation classifiers.
The following section is structured according to the steps required to generate the proposed framework. First, aspects of ontology definition are outlined, followed by discussions on endangerer profile generation and classifier training. Finally, monitoring strategies are proposed.

II. PROPOSAL

The proposed application framework enables decision makers of security forces to identify threat hot-spots. In this way, they are able to control their human resources. In order to support long-term resource planning, The second aim is to predict the long-term development of groups that pose a threat. The process pipeline consists of three parts:

1) modelling the threat ontology
2) train the general domain-specific endangerers profile
3) monitoring all matching social network sites and calculate a long-term and short-term threat score

A. Threat Ontology

The term ontology in a common understanding means a formal and explicit specification of a common conceptualization. In particular, it is defined as a set of common classified terms and symbols referred to a syntax, and a network of associate relations [4]. Similar to the crime ontology we proposed in recent work [5], an ontology can be used for modelling a complex threat assessment. In this way, knowledge of decision makers is introduced and can be used for extracting semantic information from posts and comments of social network's profiles. In particular, the works of Wimalasuriya and Dou [6], Embley [7] and Maedche [8], show that the use of ontologies is suitable for assisting the extraction of semantic units, as well as their visualization and structures such processes very well.

B. Endangerer Profile

In order to distinguish profiles of interest regarding to a certain threat, a general profile needs to be modelled. Recent work [9], [10] has shown that feature vectors derived from social network profiles are suitable for generating group recommendations. In a similar way, a general classifier can be trained based on the social network profiles of known persons associated with a special threat. For example, Facebook profiles of known hooligans of a specific soccer club can be used to train classifiers that are able to identify social activity of hooligans and peers in social networks.
The generation process is divided into three parts depicted in Figure 1.

C. Monitoring Activities

Once a profile is generated and the threat specific ontology is defined, the social network monitoring can be conducted. At this point a multi-level, information extraction process aims at instantiating the ontology using textual information, like posts and comments. An example of how such a process can be structured is given by Spranger and Labudde [5]. Further text analysis steps, like sentiment analysis (see the discussions

 ISBN: 978-1-61208-415-2

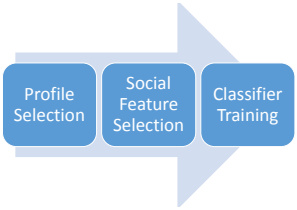

Figure 1. The process of deriving a threat specific general profile.

given in [11] and [12] for details) can complete the instantiated model in different ways. As a short-term benefit, a score can be computed for various time points, signalling whether a threatening event regarding to the specific profile and ontology is directly pointing to a specific location and time frame. These results can be applied to a map to localize short-term hot-spots in terms of security and their dynamics as discussed by Davies and Bishop [13].

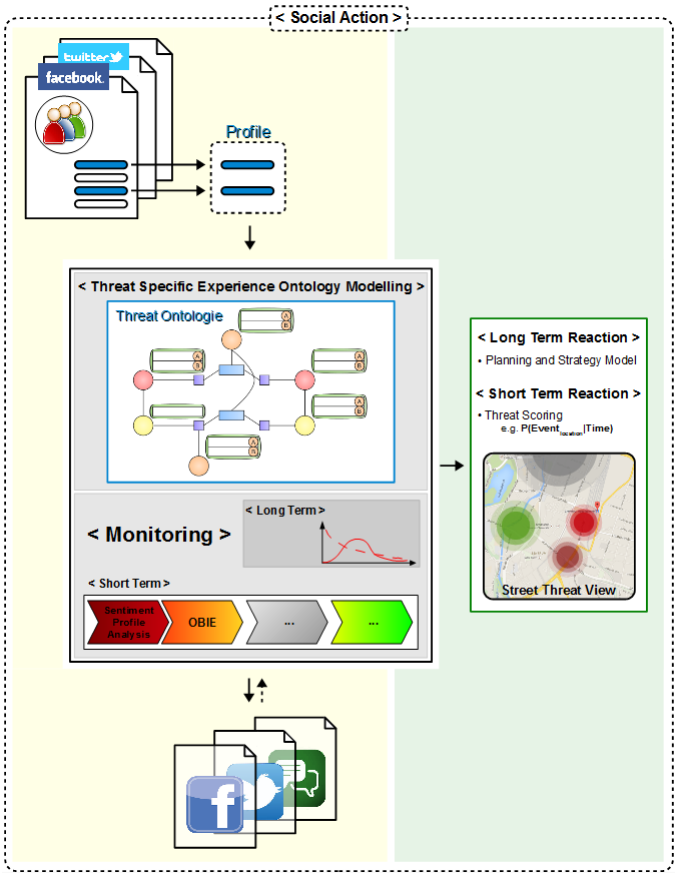

Figure 2. The proposed system. The central, expert-modelled threat-specific ontology describes the environment of a special threat. A general endangerer profile completes the model. In the process the model is used to extract textual information from social network activities. Different scoring functions allow the identification of threat hot-spots or can show the long-term evolution of groups and cliques.

In the age of Big Data and algorithms handling such amounts of information, deducing long term developments of such groups and dynamics is at its early stage. Methodological concepts widely used in modelling complex relations (as for instance systems biology) can be directly transferred to the field of resilience engineering. Especially, employing generic mathematical models to social networks has become computationally feasible, but requires further research. For example, epidemiological models can be efficiently applied to study long term evolutions of groups and sub-networks (see [14]) and study the information transfer between them. Thus, generating valid models and derive predictions from them can be of great value, for instance, in planning personnel and staff demands.

REFERENCES

[1] ITU. Number of worldwide internet users from 2000 to 2014 (in millions). statista. [Online]. Available: http://www.statista.com/statistics/273018/number-of-internet-users-worldwide/ (2015)

[2] eMarketer & American Marketing Association. Number of social network users worldwide from 2010 to 2018 (in billions). statista. [Online]. Available: http://www.statista.com/statistics/278414/number-of-worldwide-social-network-users (2015)

[3] ZIS. Jahresbericht 2013/14. Zentrale Informationsstelle Sporteinsätze. [Online]. Available: http://www.polizei-nrw.de/media/Dokumente/Behoerden/LZPD/ZIS_Jahresbericht_2013_14.pdf (2014)

[4] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," in Formal Ontology in Conceptual Analysis and Knowledge Representation, N. Guarino and R. Poli, Eds. Kluwer Academic Publishers, 1993.

[5] M. Spranger and D. Labudde, "Towards establishing an expert system for forensic text analysis," International Journal on Advances in Intelligent Systems, vol. 7, no. 1/2, 2014, pp. 247–256.

[6] D. C. Wimalasuriya and D. Dou, "Ontology-based information extraction: An introduction and a survey of current approaches," Journal of Information Science, vol. 36, no. 3, 2010, pp. 306–323.

[7] D. W. Embley, "Toward semantic understanding: an approach based on information extraction ontologies," in Proceedings of the 15th Australasian database conference - Volume 27, ser. ADC '04. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2004, pp. 3–12.

[8] A. Maedche, G. Neumann, and S. Staab, "Bootstrapping an ontology-based information extraction system," Studies In Fuzziness And Soft Computing, vol. 111, 2003, pp. 345–362.

[9] M. Manca, L. Boratto, and S. Carta, "Producing friend recommendations in a social bookmarking system by mining users content," in Proc. 3rd. International Conference on Advances in Information Mining and Management, IARIA. ThinkMind Library, 2013, p. 59 to 64.

[10] M. Cheung and J. She, "Bag-of-features tagging approach for a better recommendation with social big data," in Proc. 4th. International Conference on Advances in Information Mining and Management, IARIA. ThinkMind Library, 2014, p. 83 to 88.

[11] S. M. Mohammad, S. Kiritchenko, and X. Zhu, "Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets," in Proceedings of the Second Joint Conference on Lexical and Computational Semantics (SEMSTAR'13), 2013.

[12] X. Wan, "Co-training for cross-lingual sentiment classification," in Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1. Association for Computational Linguistics, 2009, pp. 235–243.

[13] T. Davies and S. Bishop, "Modelling patterns of burglary on street networks," Crime Science, vol. 2, no. 1, 2013, p. 10.

[14] J. Cannarella and J. A. Spechler, "Epidemiological modeling of online social network dynamics," CoRR, vol. abs/1401.4208, 2014.

 ISBN: 978-1-61208-415-2