An Ontology-based Approach to Social Networks Mining

Viacheslav Lanin[a], Lyudmila Lyadova[b], Elena Zamyatina[c] and Nikita Vostroknutov
Department of Information Technologies in Business, HSE University,
Perm, Russian Federation

Keywords: Social Network, Social Media, Multifaceted Ontology, Event Classification, Event Logs, Process Mining.

Abstract: The article presents an approach to the analysis of processes in social networks based on using multifaceted ontologies. An overview of existing tools for analyzing social networks is provided and the results of studying social networks are presented. A multifaceted ontology describing social networks has been developed based on the research findings. The main result for this study is the ontology of events, which can be used to pre-process data, extracted from social networks, to generate event logs in a form suitable for export to Process Mining tools to analyze networks (identification of user behavior patterns, analysis of information distribution, etc.). Examples illustrating the proposed approach are given.

1 INTRODUCTION

From the user's point of view any social network is an interactive multi-user website, the content of which is filled by the network participants themselves. It is an automated social environment that allows a group of users to communicate about their common interests. Communication is carried out via different tools (for example a web service of internal mail or instant messaging, and so on).

On the one hand, social networks help solve many tasks, but on the other hand, they might become a source of problems. This fact has motivated a large number of researchers to study social networks. The greatest interest is caused by the problems of dissemination of information on the network, issues of community formation, etc. The methods of static analysis of networks using graph models, statistical methods and machine learning methods are better developed.

When analyzing social networks, economists receive information about transactions, the influence of others on human behavior, while political scientists investigate the formation of political preferences. Static and dynamic methods may be used for these purposes.

This article focuses on the methods that allow researchers to study the dynamic behavior of users in networks. The most promising methods are based on the analysis of event logs. Logs can be formed based on data obtained from real social networks. But studies can also be carried out using simulation modeling tools. The ontological approach expands the possibilities of studying social networks with existing tools. It is proposed to use multifaceted ontology (Abrosimova, 2018; Shalyaeva, 2016; Shalyaeva, 2017) to retrieve information on processes in social networks, to prepare event logs used by Process Mining tools to solve many tasks of analyzing various systems.

2 RELATED WORKS

Consider approaches to solving the most interesting problems of social network analysis according to goal of the research presented in this paper.

Previous research works provide many examples of using social networks to distribute content among users. This is how marketers try to spread information about products in order to make profit (Kang, 2015; Yang, 2018; Bindi,2017; Zhao, 2018). On the other hand, attackers try to spread malicious or fake information (Dang-Pham, 2020, Dmitriev, 2020,

[a] https://orcid.org/0000-0002-0650-2314
[b] https://orcid.org/0000-0001-5643-747X
[c] https://orcid.org/0000-0001-8123-5984

Lanin, V., Lyadova, L., Zamyatina, E. and Vostroknutov, N.
An Ontology-based Approach to Social Networks Mining.
DOI: 10.5220/0010716600003064
In *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2021) - Volume 2: KEOD*, pages 234-239
ISBN: 978-989-758-533-3; ISSN: 2184-3228

Ilieva, 2018; Tumbinskay, 2017). Thus, it is necessary to develop algorithms or strategies that either contribute to the fastest possible dissemination of information, or, conversely, hinder dissemination.

In the article (Van der Aalst, 2005), event logs are used to identify social connections between employees of a company, to build a sociogram that reflects the structure of social connections for users of the CRM system of the company. These tools are useful to study connections between users of mass media too. In order to use the Process Mining tools, it is needed to get an event log.

An overview of the technologies for extracting information on events is given in the paper (Zhan, 2019).Event logs can be built on the base of data stored in databases (Calvanese, 2016).

Event logs can be constructed with using data of network messages (Carrasquel, 2021) and event data of different applications (Mukala, 2015). Methods of extracting data to generate event logs from unstructured or semistructured sources are also discussed. News events can be extracted from social media (Peña-Araya, 2015).

The article (Zavarella, 2014) describes a system that automatically identifies certain types of global events, such as natural disasters, epidemics, and military conflicts through the analysis of news sites and social networks. The article describes several experimental approaches to semantic integration of user content published in social networks with existing information systems.

The development of a system used to automatically receive and categorize events based on user posts on Twitter is described in the article (Ritter, 2012). The author uses machine learning algorithms to work with publications and identifies various event groups from comments, including entities, events, dates, and categories.

Semantic technologies extend data pre-processing capabilities at preparing event logs for the analysis with process mining tools. These technologies have become the kernel for process analysis software developing (De Medeiros, 2008) as they allow solve the tasks of information retrieval, data extraction and analysis, in particular, when searching for and analyzing facts (Vokhmintsev, 2013).

Simulation methods, their capabilities and advantages are discussed in the articles (Mikov, 2013; Zamyatina, 2020; Dmitriev, 2020). The results of application of various methods and tools of simulation modeling to solve problems of social networks analysis are presented.

3 SOCIAL NETWORKS ONTOLOGY DEVELOPMENT

In this research, developed ontology should be based on the combination of two aspects: functional and structural (Kietzmann, 2011) views to social media. According to the purpose of this research, the key concept must be an *Event*.

It is necessary to define possible events happening in social media via their functional and structural elements. So, event-based approach to the definition of social media can be discussed.

According to the selected approach possible types of events in social media should be defined first and generalized social media types could be explored then via *event logs*.

3.1 Social Networks Studying

The main properties of social networks to be identified to prepare information for modeling.

All social networks have a property called *user-oriented design*. In the article (Dawot, 2014) the basic principles of user-oriented design are described: interaction between users; recognition (individuality) of community members. These properties allow solve the tasks of generating event logs, where it is necessary to identify not only events, but also objects and processes and their cases. In the article (Kietzmann, 2011), the authors highlight seven *main structural elements*, characterizing social networks based on the principles of user-oriented design:

- *Identity* (possibility of self-expression of the social network user).
- *Conversations* (ways of communication between social network users, both personal and group).
- *Sharing* (methods of obtaining, distributing and modifying public content.
- *Presence* (the ability of users to have the location of other users).
- *Relationships* (different connections between social network users).
- *Reputation* (the ability of users to confirm their popularity or competence).
- *Communities* (user interaction with communities (groups) and sub-communities).

The identification of *functional components* for the listed above structural components leads to information extraction need for analyzing events and their description. The event model of a social network is formed on the basis of the received information on the functional components relevant to the structural components of social networks.

3.2 Designing an Event Model of a Social Network

The main event elements of the social network are listed in Table 1.

Table 1: Main event elements of the social network.

Functional element	Event element
Personal profile	Adding updating or deleting personal information
Content profile (combining various publications from users or groups into one structure, Facebook feed for example)	Changing of user content profile
Activity log	Changing a history of user activity
Social content	Generation of discussions, comments and personal messages
Discussion topic	Creating or changing a topic of discussion
Notification	Notification about post or social connection
Content creation	Publishing text, audio, video post. Rate content (like), sending sticker or emoji.
	Editing text, audio, video message
Content sharing of (personal or third-party)	Rate content (like), repost.
Community	Creating a community
	Deleting a community
Social connection	Applying for social connection
	Changing of social connection
Location status	Changing of location status
Content popularity	Changing of content popularity metric
Subscription	Adding or removing subscription

This general events description is the base for event ontology development.

3.3 Development of the Social Network Event Ontology

The first part of ontology deals with events. The bottom-up method was used for developing an ontology part describing social media events. Respectively the first stage was a class definition for elementary events. The total number of classes of elementary events is more than thirty. After that, at the second stage, elementary events were generalized to appropriate super classes. The top of the event hierarchy is the generalized class named *Social_Media_Event*.

The ontology of a social network includes the general class *Social_Network*, which comprises certain subclasses highlighted on the basis of the analysis of structural elements listed above (Section 3.1) of social networks (Figure 1).

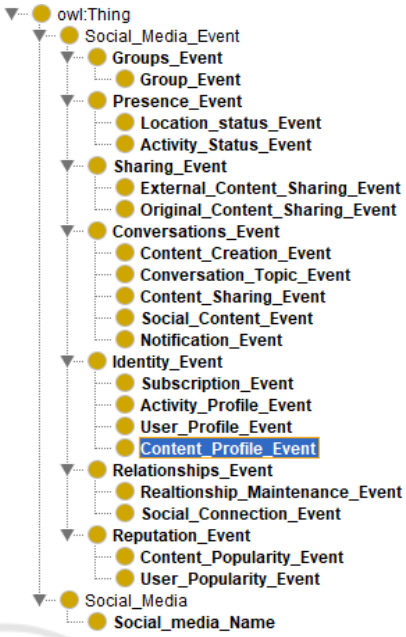

Figure 1: Functional Element Class Hierarchy.

The identified events are presented in ontology as subclasses of functional class (Figure 2).

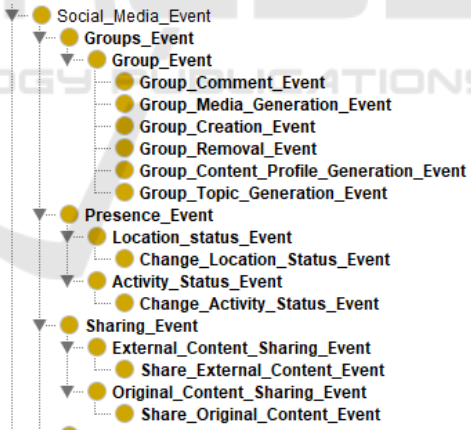

Figure 2: Example of Event Mapping Structure.

Typically, in social media, several elementary events are closely related to each other. For example, the event of sending a message by one user is associated with the event of creating a notification for other users to whom the message is addressed.

The *Complex_Event* class has been added to the ontology to model such related events that are reflective of each other (a “reflected” event is a consequence of the first event). It allows to perform a joint analysis of several related events more

effectively than considering the events independently of each other. Instances of classes can be used to model such complex events. The instances can be adapted for representing common case behavior in a particular social network. For example, when a user clicks the Subscribe button on Youtube, two different events are triggered within the system. The first one is adding the channel to the user's subscription list, and the other is a notification of the channel owner about a new subscriber. The following processes can be described using two classes stored inside the ontology (Figure 3): *Update_User_Popularity_Event* and *Add_Subscription_Event*. To represent such a complex event within the ontology, it is necessary to create instances of classes for the existing events.

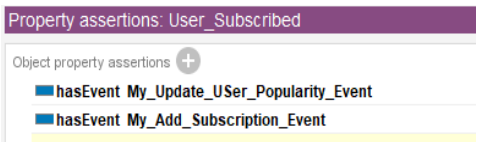

Figure 3: *User_Subscribed* object.

Several more complex events were introduced to show how it is possible to model processes within social networks with the created ontology.

The resulting ontology, describing social network and identified events, is presented using the full ontograph in Figure 4.

4 CREATING EVENT LOGS WITH USING ONTOLOGY

To study social networks a variety of methods of Process Mining can be used. But the event log can be considered the main concept. The event data described in the ontology can be used to automatically generate event logs. To generate a log, it is needed to solve two tasks:

1) to extract event data from social networks using descriptions available in the ontology;
2) to generate a log according to a given form using queries to select the desired events, the data about which is stored in the ontology.

Applicability of the developed tools is illustrated by examples of journals generated with the system (fragments of the event records contained in them).

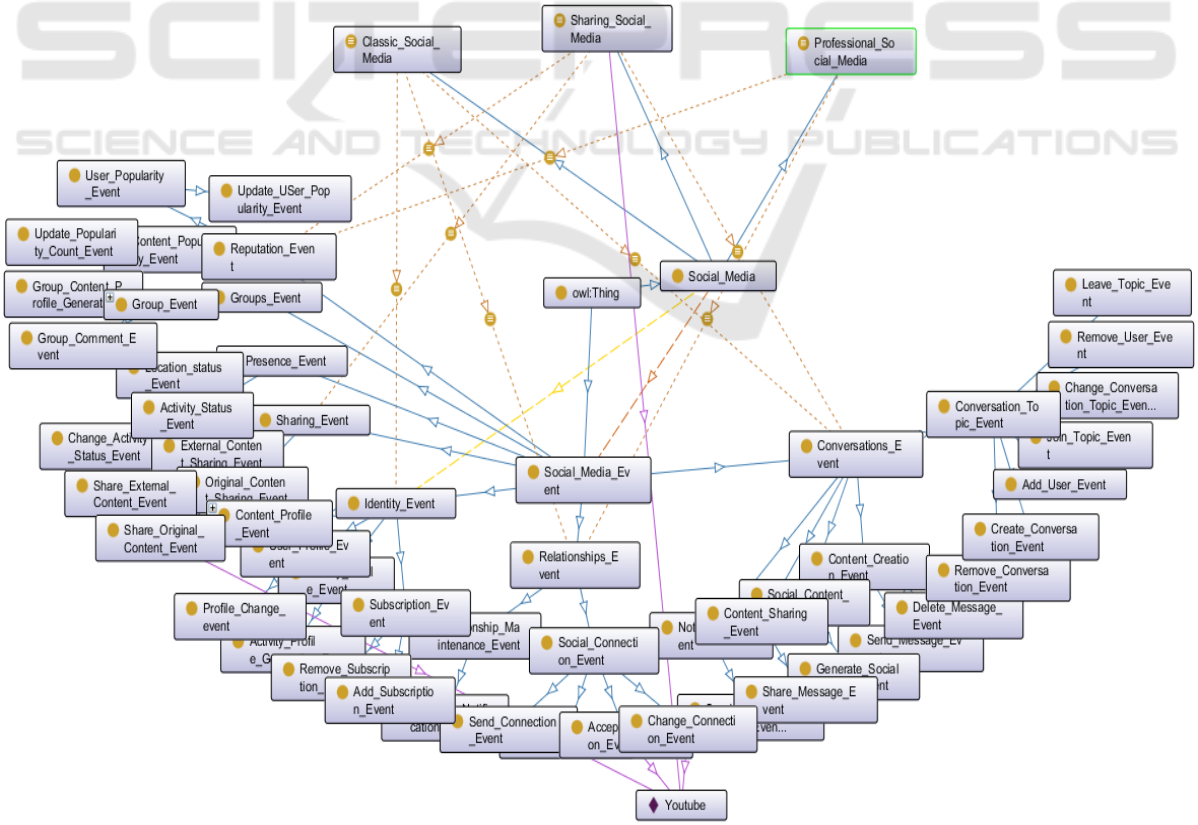

Figure 4: Social network ontology.

4.1 Event Logs Generation

To create an event log using the API, a request to the API to obtain data must contain the following information: event type (corresponds to events classified in ontology); event description; resource (event performer) and timestamp.

To achieve this goal, when working with the VK API, the functions of obtaining information on the content profile and "likes" on publications of the content profile ("identity" block), on notifications ("conversations"), on the distribution of content in the community ("distribution"), on obtaining user status ("presence"), on obtaining applications for friends ("relationship") were used.

After filling the ontology with the events received from a social network, the user should be able to unload event logs from the ontology using queries.

The generated event logs should also be stored in the ontology for further usage in social network mining with Process Mining tools (for example, event logs can be exported to ProM).

4.2 Event Logs Examples

Event logs are generated in XML format. When generating, it is possible to select events that interest the user: you can select records by user ID, dates, or event types, for example.

The short fragment of the event log generated by user ID (only events for 330355947 ID are selected):

```
<entry>
  <content>null</content>
  <date>04/06/2021</date>
  <event>
  new_ontology.
       send_connection_event2
  </event>
  <user>330355947</user>
</entry>
```

The short fragment of the event log generated by selected event type (only event records of type *Send_Connection_Event* are uploaded):

```
<entry>
  <content>null</content>
  <date>04/05/2021</date>
  <event>
    new_ontology.
         send_connection_event1
  </event>
  <user>120818089</user>
</entry>
<entry>
  <content>null</content>
  <date>15/06/2021</date>
  <event>
    new_ontology.
         send_connection_event2
  </event>
  <user>330355947</user>
</entry>
```

Only single entries (fragments of logs that can contain many records when analyzing real networks) are shown above.

Users, by studying social networks, can combine data from different journals formed independently of each other, for example, containing information about events in several groups.

5 CONCLUSIONS

The main result of the study at this stage is the ontology of social media events. The presence of the described information in the event logs allows to analyze the connections between various events in social networks, to identify patterns of user behavior, to evaluate the "intensity of work" and "workload" of users. Event logs are prepared in a standard format for export to analyze with external applications (Process Mining tools, for example, ProM). However, it is possible to extend protocols with new attributes, which enable the development of analysis tools that take into account values of these attributes.

The next stage of the study involves experimenting with the information, extracted from social networks, with simulation tools (Dmitriev, 2020). This research method allows generating event logs based on models and comparing results with real data obtained with network mining.

REFERENCES

Abrosimova, P., Shalyaeva, I., Lyadova, L. (2018). The Ontology-Based Event Mining Tools for Monitoring Global Processes. In: *Proceedings of the IEEE 12th International Conference on Application of Information and Communication Technologies (AICT 2018)*. Almaty, 2018. Pp. 108-113.

Bindu, P.V., Thilagam, P.S., Ahuja, D. (2017). Discovering suspicious behavior in multilayer social networks. In: *Computers in Human Behavior*. Vol. 73, pp. 568–582.

Calvanese D., Montali M., Syamsiyah A., van der Aalst W.M.P. (2016). Ontology-Driven Extraction of Event Logs from Relational Databases. In: *Business Process Management Workshops. BPM 2016. Lecture Notes in Business Information Processing*, vol 256. Springer, Cham.

Carrasquel, J.C., Chuburov, S.A., Lomazova, I.A. (2021). Pre-processing Network Messages of Trading Systems into Event Logs for Process Mining. In: *Tools and Methods of Program Analysis. TMPA 2019. Communications in Computer and Information Science*, vol 1288. Springer, Cham. Pp. 88-100.

Dang-Pham, D., Pittayachawan, S., Bruno, V. (2020). Applications of social network analysis in behavioural information security research: Concepts and empirical analysis. In: *Computers & Security*, 2017. Vol. 68, pp. 1–15.

Dawot N.I.M. (2014). A review of features and functional building blocks of social media. In: *Proceedings of the 8th. Malaysian Software Engineering Conference (MySEC)*. IEEE, 2014. Pp. 177-182.

De Medeiros, A.K.A., Van der Aalst, W., Pedrinaci, C. (2008). Semantic process mining tools: Core building blocks. In: *Proceedings of the 16th European Conference on Information Systems*. Galway, Ireland, 2008. Pp. 15-23.

Dmitriev, I, Zamyatina, E. (2020). How to Prevent Harmful Information Spreading in Social Networks Using Simulation Tools. In: *Analysis of Images, Social Networks and Texts. AIST 2019. Communications in Computer and Information Science*. Vol. 1086, pp. 201–213. Springer, Cham.

Ilieva, D. (2018). Fake News, Telecommunications And Information Security. In: *International Journal "Information Theories and Applications"*. Vol. 25(2), pp. 174–181.

Kang, H., Munoz, D. (2015). A dynamic network analysis approach for evaluating knowledge dissemination in a multi-disciplinary collaboration network in obesity research. In: *2015 Winter Simulation Conference (WSC)*. Pp. 1319-1330.

Kietzmann J.H. (2011). Social media? Get serious! Understanding the functional building blocks of social media. In: *Business horizons*. 2011, vol. 54, №. 3. Pp. 241-251.

Mikov A., Zamyatina E., Kozlov A., Ermakov S. (2013). Some Problems of the Simulation Model Efficiency and Flexibility. In: *Proceedings of the 8-th EUROSIM Congress on Modelling and Simulation EUROSIM*, 2013, Cardiff, Wales, United Kingdom. Pp. 532–538.

Mukala, P., Buijs, J., Leemans, M., van der Aalst, W. (2015). Learning analytics on coursera event data: a process mining approach. In *Proceedings of the 5th International Symposium on Data-driven Process Discovery and Analysis* (SIMPDA 2015), Vienna, Austria, December 9-11, 2015. CEUR Workshop Proceedings; vol. 1527. Pp. 18-32.

Peña-Araya, V. (2015). Galean: Visualization of Geolocated News Events from Social Media. In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '15). ACM New York. 2015. Pp. 1041-1042.

Ritter A. (2012). Open domain event extraction from twitter In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2012. Pp. 1104-1112.

Shalyaeva, I., Lyadova, L., Lanin, V. (2016). Events Analysis Based on Internet Information Retrieval and Process Mining Tools. In: *Proceedings of 10th International Conference on Application of Information and Communication Technologies (AICT2016)*. Baku, 2016, pp. 168-172.

Shalyaeva, I., Lyadova, L., Lanin, V. (2017). Ontology-Driven System for Monitoring Global Processes on Basis of Internet News. In: *Proceedings of IEEE 11th International Conference on Application of Information and Communication Technologies (AICT2017)*. Moscow, 2017, pp. 385-389.

Tumbinskaya, M.V. (2017). Protection of information in social networks from social engineering attacks of the attacker. In: *Journal of Applied Informatics*. Vol. 12(3(69)), pp. 88–102.

Van der Aalst, W.M.P., Reijers, H. ., Song M. (2005). Discovering social networks from event logs. In: *Computer Supported Cooperative Work (CSCW)*. 2005, vol. 14, № 6. Pp. 549-593.

Vokhmintsev, A., Melnikov, A. (2013). The Knowledge on the Basis of Fact Analysis in Business Intelligence. In: *Digital Product and Process Development Systems. NEW PROLAMAT 2013. IFIP Advances in Information and Communication Technology*, vol 411. Springer, Berlin, Heidelberg. Pp.134-141.

Yang, D., Liao, X., Shen, H., Cheng, X., Chen, G. (2018). *Dynamic node immunization for restraint of harmful information diffusion in social networks*. In: Physica A: Statistical Mechanics and its Applications. Vol. 503, pp. 640–649. DOI: 10.1016/j.physa.2018.02.128.

Zamyatina, E., Dmitriev, I. (2020). How to Prevent Harmful Information Spreading in Social Networks Using Simulation Tools. In: *Analysis of Images, Social Networks and Texts. AIST 2019. Communications in Computer and Information Science*. Vol. 1086, pp. 201–213. Springer, Cham.

Zavarella V. (2014). An Ontology-Based Approach to Social Media Mining for Crisis Management. In: *SSA-SMILE@ ESWC*, 2014. Pp. 55-66.

Zhan, L., Jiang, X. (2019). Survey on Event Extraction Technology in Information Extraction Research Area. In: *Proceedings of the IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. Chengdu, China, 2019. Pp. 2121-2126.

Zhao, N., Cheng, X., Guo, X. (2018). Impact of information spread and investment behavior on the diffusion of internet investment products. In: *Physica A: Statistical Mechanics and its Applications*. Vol. 512, pp. 427–436.