AUTOMATED ENTROPY ANALYSIS OF THE SOCIAL CONSEQUENCES OF URBAN MAN-MADE ACCIDENTS AND NATURAL CATASTROPHES

Sviatoslav Timashev, Elizaveta Malyutina

•

Science & Engineering Center Reliability and Safety of Large Systems and Machines
Ural Branch Russian Academy of Sciences, RUSSIA
timashevs@gmail.com
2malyutina2@mail.ru

Abstract

The article describes the software package (SP) developed by the authors, which allows analyzing the social response to incidents and accidents of urban infrastructures in real time. The analysis is based on publications on the VKontakte social network. The SP performs filtering and categorization of textual information, storing data in a SQLite database, as well as quantitative and qualitative analysis of the reaction of citizens to a dangerous event. The SP automates the process of monitoring social response to emergencies, which makes it useful for the rapid response of government organizations and emergency services. The SP allows us to study, from an interdisciplinary perspective (combining the theory of technogenic safety and sociology), the interdependence of measurable damage caused by an accident of urban infrastructure and the intangible socio-cognitive damage to the community of citizens due to this accident.

Keywords: social networks, entropy, data analysis, safety.

I. Introduction

In [1, 2], the concept and approach to the entropy analysis of the social consequences of major accidents of urban infrastructures and systems are formulated based on the collection of necessary information published on social networks reflecting the details of accidents, moods, reactions and demands of society in connection with the occurrence of emergency situations. For the full-scale use of this approach, an algorithm is needed for automatic selection, collection and comprehensive socio-cognitive analysis of information purposefully extracted from social networks. This is the subject of this article.

The currently existing official methods and methods of informing the public about accidents related to the functioning of network systems and urban life support infrastructures are presented in Table 1.

There is also unofficial information, in the form of public reaction to the events, which is published on social networks. In this article, it is proposed to use social networks to monitor the state of critical infrastructures and obtain additional information about the reaction of urban society to man-made incidents and accidents [3].

II. Methods

Currently, the *VKontakte* social network is the most accessible, popular and, most importantly, legitimate platform. Therefore, it can be used as an effective tool for rapid response and understanding of public perception of incidents. *VKontakte* provides access to an extensive

database, which is formed by various social groups and reflects various socio-cultural aspects of society. These multiple data are representative and can be used for a consistent analysis of public opinion and trends among different categories of the population.

Table 1: *Modern means and methods of informing society during and after natural and man-made accidents and catastrophes*

Title	Advantages	Disadvantages
Traditional media	Provide global coverage and have high credibility	Delay in the dissemination of information, limited coverage of points of view
Official reports and press releases	Information from primary sources, formal and accurate data	May be biased, delayed, do not include the public opinion
Polls and public opinion research	Allow getting structured data	May take time, limited coverage, depends on the wording of the questions
Expert assessments	High level of expertise	Can be subjective, due to a limited range of experts

VKontakte has a VK API that provides access to various data about users, communities, groups, posts, comments and other aspects of activity on the platform. Thanks to the VK API, researchers can collect statistics themselves and conduct in-depth analysis of user behavior data.

The disadvantage of *VKontakte* is data access restrictions due to user privacy settings, strict platform rules, and frequent API changes, which makes it difficult to collect data. This can lead to an incomplete analysis.

Nowadays, social networks (SN) are becoming an increasingly useful tool for improving public life. Many city administrations use SN (for example, *the Incident Management* system) as a channel through which citizens can promptly report their problem.

It seems that the SN can be used more effectively and in a multidimensional manner if regular monitoring of the SN is carried out in order to:

(1) promptly obtain information about incidents, accidents, natural disasters;

(2) monitor socio-cognitive changes in society during and after the occurrence of these incidents and accidents;

(3) promptly adopt measures to mitigate, prevent, protect and ensure the safety of citizens;

(4) carry out optimal governance in its territory, using the SN as a positive feedback in the management chain.

This work is aimed at quantitative, entropy-probabilistic analysis of social networks. The city of Kamyshlov, located in the Sverdlovsk region, was chosen to work out the methodology and algorithm for analyzing the reaction of the population to incidents in urban infrastructure. As of January 1, 2021, the city's population was 25,582. Data collection was carried out in the main community of the city in the SN *VKontakte*, where the number of participants exceeded 30,000 people.

A major accident at the central pumping station occurred on March 30, 2019. Meltwater flooded the pump's engine room, which led to the failure of two sewage pumps and the shutdown of the central water supply. As part of the study, a chronology of events was compiled:

30.03.2019	13:39 – Accident at the central pumping station
	15:00 – Shutdown of the central water supply
	16:33 – The first message about the accident (VKontakte publication)
04.01.2019	02:38 – The mayor of the city addressed the residents through a social network

	11:00 – The end of the liquidation of the accident
	23:05 – End of water chlorination
04.03.2019	22:07 – The last post about the accident
05.15.2019	7:19 – The last mention of the accident on social media.

The mayor of the city informed residents about the measures taken to restore water supply, but many citizens expressed dissatisfaction due to the lack of information about: a) the time of restoration of water supply, b) contact services, c) the causes of the incident.

Visualization of social activity after water shutdown is shown in Fig. 1. Along the abscissa axis the dates of the main events that occurred during the liquidation of the accident are posted. The ordinate axis shows the number of likes, views (in thousands) and comments under the post.

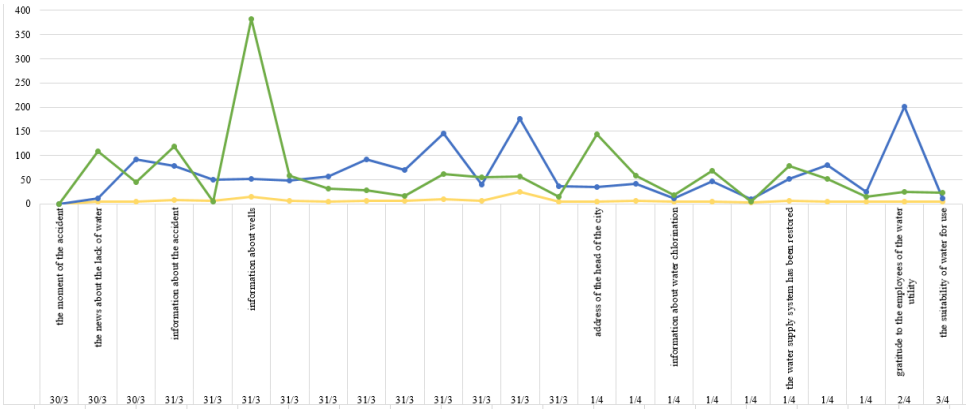

Figure 1: *User activity after the accident at the pumping station on 30.03.2019 (the number of views (in thousands) is given in yellow, the number of likes is given in blue, the number of comments is given in green)*

The greatest activity of citizens was observed on the second day after the water was turned off, when residents began to actively search for information about the possibilities of obtaining water and share information about available sources. The number of comments as a function of time is shown in Fig. 2.

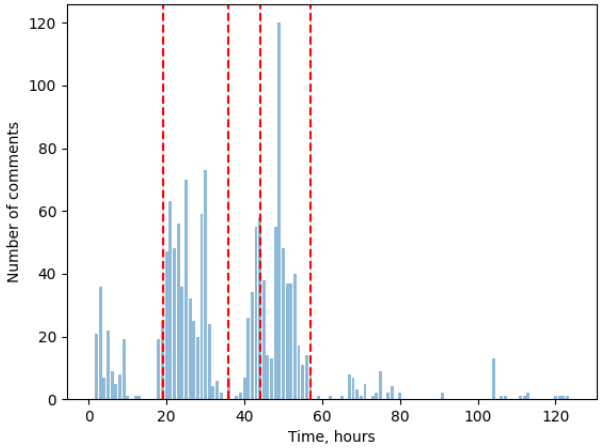

Figure 2: *The graph of the number of comments in time. The dashed lines indicate following events (from left to right): publication of information about available wells (19th hour), mayor addresses the citizens (6th hour), end of the accident liquidation (44th hour), end of water chlorination (57th hour).*

The graph shows a sharp increase in the number of comments, which peaked on the third day after the incident (April 1), when residents began to actively share their experiences, ask questions and support each other, after which the frequency of comments began to decrease, as the relevance of the topic decreases.

To improve the accuracy of the results, comments collected on social networks were classified

according to their information value. The entries were divided into four groups, depending on the content of the comment, and visualized as shown in Fig. 3.

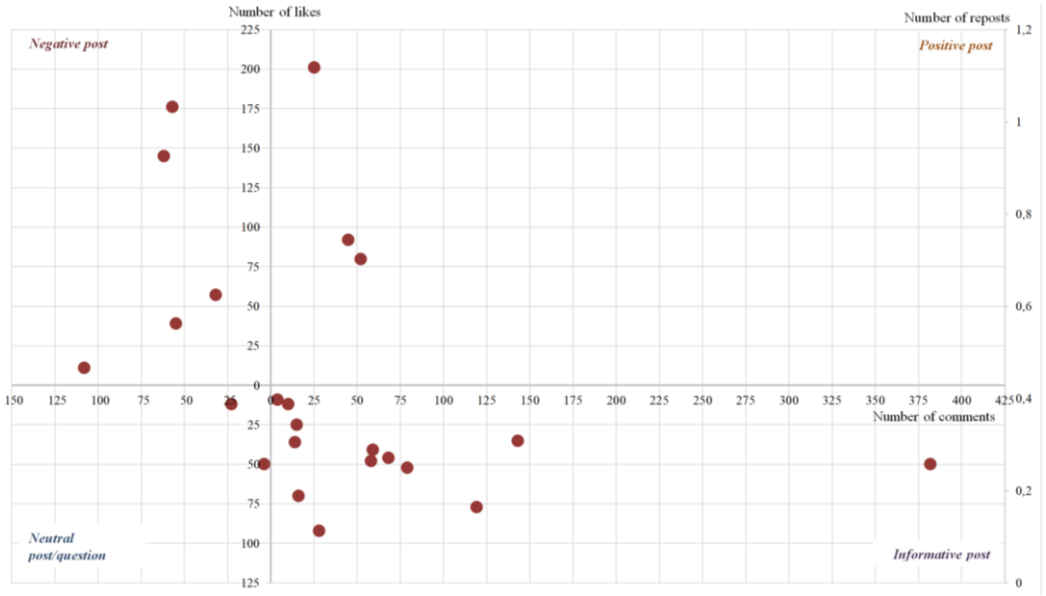

Figure 3: *A phase portrait of the accident*

Entries containing useful information are located in the lower right corner. Publications with a positive character and humorous messages are displayed in the upper right quadrant, and messages of a negative nature are displayed in the upper left quadrant. Neutral posts or messages containing a question are located in the lower left corner. The specific characteristic patterns in Fig.5 allow seeing at a glance what moods prevailed in the society during the crisis.

Fig. 4 shows the calculated entropy quanta (hourly entropy) (a) and its first derivative in time (b).

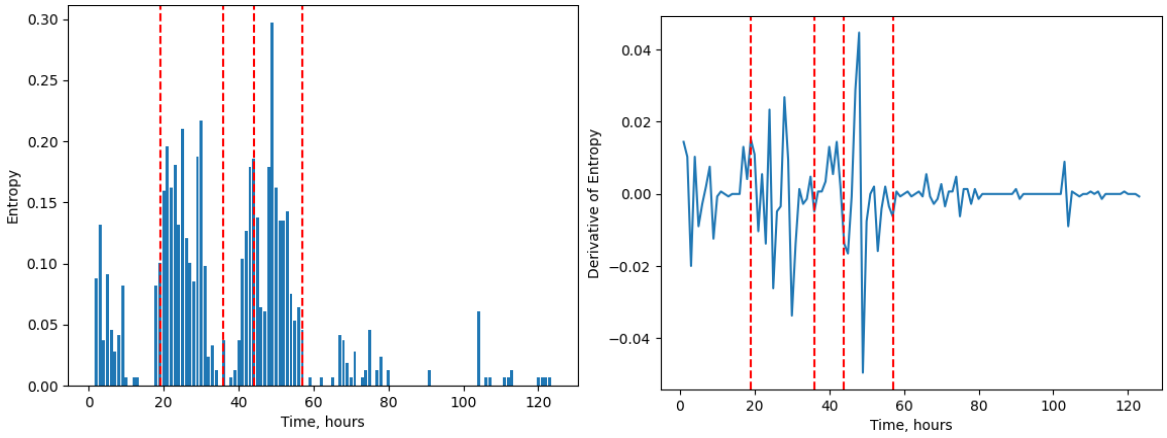

Figure 4: *Hourly values of entropy (a) and its derivative (b) in case of an accident at the central pumping station*

The initial stage of the study included: (1) manual selection of all messages related to the accident, which created a database for subsequent analysis; (2) chronological (hourly) systematization of comments about each message, and (3) their classification by content. Figure 5 shows the evolution of the probability density functions (PDF) and the entropy curve over time after an accident.

The analyzed real incident can be described as "standard", since it caused only one rapidly fading wave of social activity (five days). The main conclusions of this analysis are:

- The VKontakte social network plays an important role in the exchange of information and coordination of citizens' actions in times of crisis.

• The lack of information on the timing of the restoration of water supply and the causes of the accident contributed to an increase in social activity and discontent among the population.

• The short-term nature of the public reaction is due to the comparative speed of eliminating all the accident's consequences.

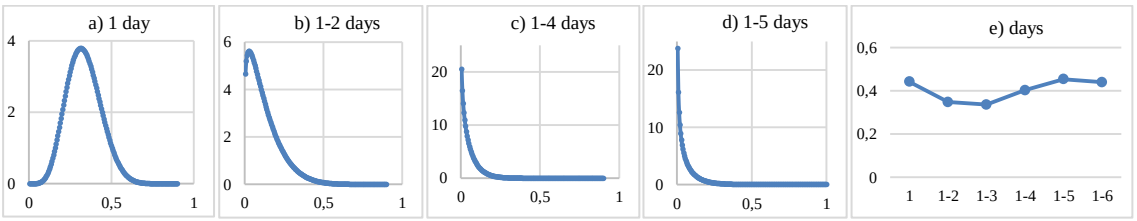

Figure 5: *a)-d) - Total PDF; e) - Change in entropy depending on the number of days that have passed since the accident*

The pilot study of the Kamyshlov population reaction has clearly shown that the use of modern technologies and communication platforms can significantly improve the quality of interaction between residents and local authorities, especially in times of crisis.

The main functions of the program for the automatic analysis of data on various incidents and events

The SP consists of the *following steps*:

1. Search in social networks for communities (public, groups, channels, etc.) by a given keyword (city name). The function is used to send a search request to the VKontakte API for groups by a given keyword. It returns a dictionary where the keys are the names of the groups and the values are their identifiers, which allows a quick find of groups of interest for further analysis

2. Filtering and analyzing publications (posts) according to a given *dictionary of keywords by category (such as "Fire", "Flood", "Terrorism", etc.)*. For this, the content of each group is parsed and checked for mentions of these words.

Additionally, it is possible to customize synonyms and variations of keywords. This makes the analysis more flexible, allowing to take into account different formulations and regional language peculiarities. For example, the expression "natural disaster" can be used instead of "catastrophe". This approach significantly increases the chances of finding important publications.

3. If a publication contains a keyword, information about it, including the text of the post, date, time, publication id, as well as the keyword mentioned in the text, is stored in the appropriate category. The program also collects and analyzes comments on these messages, which allows identifying the user reactions to specific events.

4. The program saves the result of the work to the SQLite database. Each category of data is stored in a separate table with an appropriate structure, which ensures organized storage of information. The SP contacts the SN daily and updates (replenishes) information about the incident, accident or catastrophe. SQLite provides compact storage and efficient access to data, which is critical for programs working with large amounts of information.

5. The quantitative analysis of the accumulated results includes the construction of graphs (Fig. 6) reflecting the dependence of the number of comments on time for each category of keywords. This function reads data from SQLite and visualizes the dynamics of discussions in VKontakte groups.

6. A qualitative analysis of texts is planned (identification of the mood of comments).

Each function of the program has a clearly defined task, which contributes to the modular structure of the code and facilitates the support and expansion of the program.

SP testing in the city of Yekaterinburg. ***Implementation***: a Python script using the requests, vk_api, matplotlib, numpy, pandas libraries to analyze messages from a social network.

Yekaterineburg, the capital of the Middle Urals, was chosen for SP ***testing.*** According to [4], 1,536,183 people lived in Yekaterinburg on January 1, 2024. Data collection was carried out in the groups of the city of the VKontakte social network. For the analysis, main types of incidents were selected - fire, flood, water outage, road accident, evacuation, pandemic, environmental crisis, terrorism, conflict, for each of which a dictionary of keywords was collected.

SP Advantages:

- Automation of the process of collecting and analyzing large amounts of textual information from social networks, which saves time and simplifies the real-time monitoring process.
- The ability to quickly identify and aggregate data on people's reactions to various emergency events.
- Creation of structured reports and analytics for subsequent use in making management decisions.
- Improving the efficiency and responsiveness of emergency response.
- Forecasting and prevention of emergencies.

The dynamics of user comments on the topic of incidents related to the "Fire" category from April 3, 2024 to July 31, 2024 is shown in Fig. 7. The vertical axis displays the number of comments left by users on each of the days related to the fire topic.

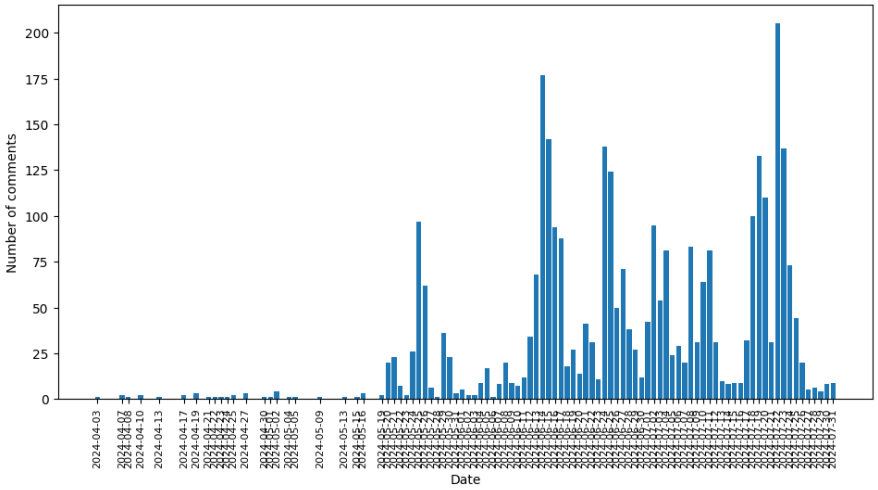

Figure 6: *A graph of the number of comments in time for incidents comments of the "Fire" category*

Judging by the graph (Fig. 6), bursts of discussion activity were observed on May 25, June 14, June 24 and July 22 (the highest peak) 2024, due to an increase in the frequency of fires. The latter is due to high air temperature, wind, dryness, active growth of grass and shrubs, and increased human activity in nature, which increases the likelihood of careless handling of fire.

The increase in user activity in the period from 11.06.24 to 17.06.24 is due to several large fires. At such moments, people tend to discuss what happened, express their feelings and get information about the situation. Peak days of activity are associated with the *size* of fires, *social unrest*; and *increased media coverage*, as users tend to follow the news and share their opinions.

User comments cover both informative and emotional aspects of the reaction to what happened. Analyzing these publications allows understanding which topics and content formats are causing an increased response from the audience that can be useful for further monitoring and analysis of incidents.

III. Results

Mathematical processing of the results

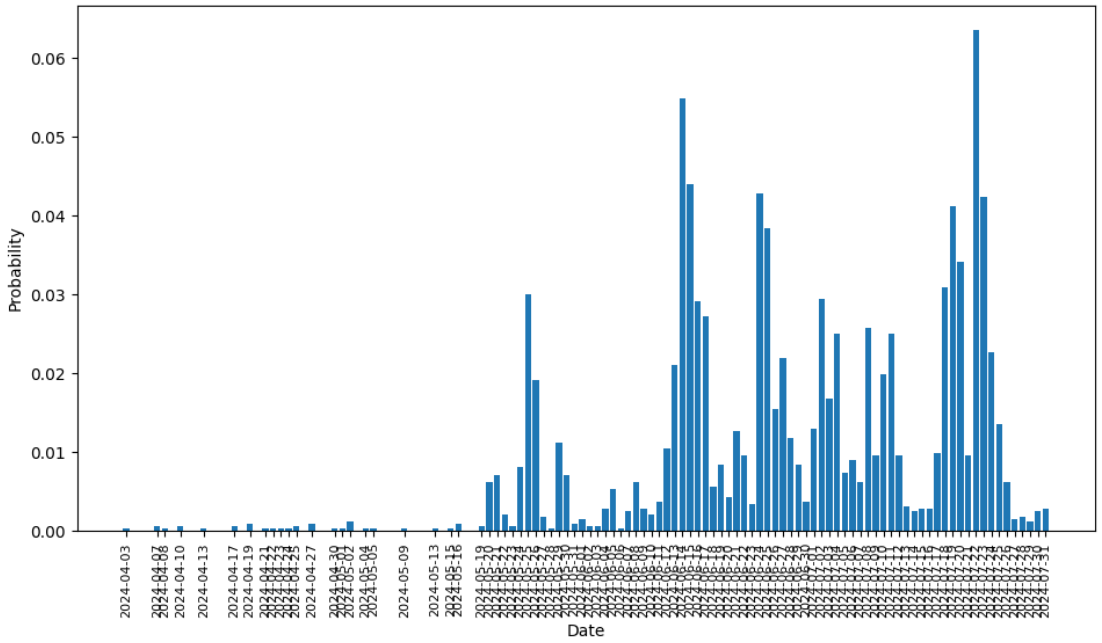

Figure 7: *A graph of the probability of comments for incidents of the "Fire" category*

The graph (Fig. 7) illustrates the probability of comments on fire publications depending on time, and shows peaks on the same dates as in Fig. 7, which indicates a direct relationship between the number of comments and public interest in incidents.

An increase in the likelihood of comments appearing at certain points confirms the importance of events taking place in a locality and their impact on social discussion.

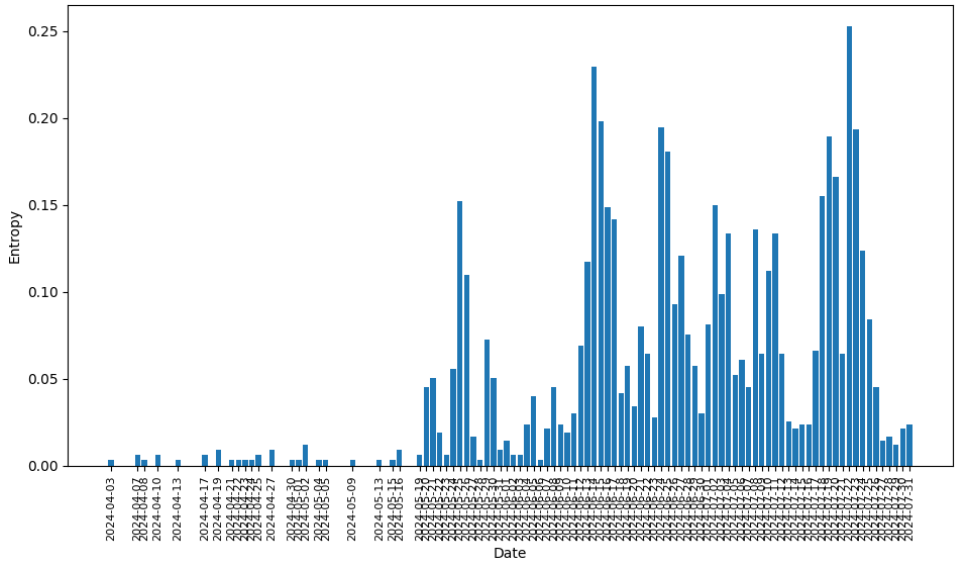

Figure 8: *A graph of the entropy of comments from time to time for the "Fire" category incidents*

The increased entropy during periods of maximum activity of citizens (Fig. 8) indicates a large number of unique and diverse opinions on the topic under discussion, which makes the audience's reaction more voluminous and saturated.

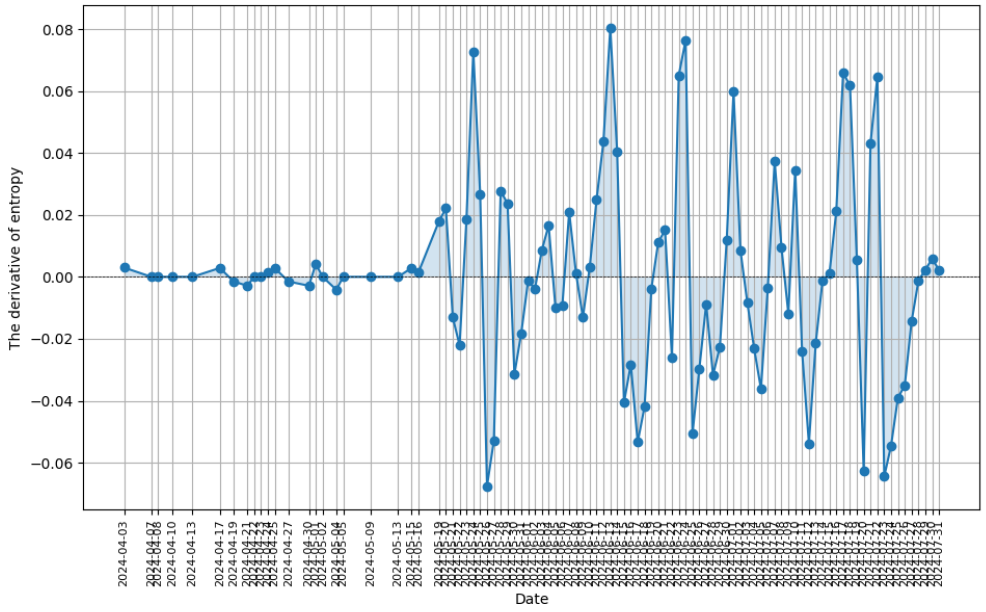

Figure 9: *Graph of the first derivative of the entropy of comments from time for the "Fire" category incidents*

The entropy derivative allows you to highlight the moments of abrupt changes in the volume and variety of comments. The points where abrupt changes in the derivative are observed may indicate significant events or changes in public opinion, which is important for crisis analysis and decision-making. These graphs provide a comprehensive understanding of the dynamics and reactions of users on social networks in response to incidents, allowing for more effective management of information flows in emergencies. The activity of users on social networks during crisis events has several important meanings:

A signal to the authorities: The high level of discussion indicates public concern, which may require government intervention.

Sentiment analysis: The number and tone of comments can help to understand the reaction of society to the incident, identify key problems and issues that need to be addressed.

Dissemination of information: Data analysis helps to track how information is disseminated among the population and which aspects of events are of the greatest interest.

The developed program is a multifunctional tool that significantly improves data processing in the context of big data flow, which makes it a valuable asset for all stakeholders. The results of testing the program showed that an automated approach to collecting, processing and analyzing data on events in the *VKontakte* SN allows:

(1) speeding up the process of monitoring incidents,

(2) better understanding the reaction of society to various emergencies,

(3) identifying key issues that concern the population, assess the emotional coloring of discussions,

(4) minimizing damage and save lives by ensuring more efficient use of resources,

(5) providing decision makers with the opportunity to make decisions that are more informed.

Combining the statistics accumulated by the SP with historical data on similar events will make it possible to create more competently a complete group of possible scenarios of incidents and accidents, which improves the quality of risk assessment of the operation of urban infrastructures.

In general, the use of this program can significantly improve the quality of real-time analysis of commercial, social and environmental information, which contributes to the creation of a more open dialogue between the authorities and society.

At this stage of SP development, the following stages of its upgrade are considered:

improving search accuracy by expanding the list of keywords; introducing machine-learning methods; expanding the functionality of the program by integrating with other social networks; improving data visualization; using sentiment analysis algorithms.

The increase in the volume of processed data leads to the problem of filtering information, and the use of data from social networks raises questions of confidentiality and ethics. It is important to find a balance between the need to respond quickly to emergencies and protecting the privacy of users. The establishment of transparent rules for working with data and compliance with ethical standards is an integral part of the effective use of SN monitoring technologies.

IV. Conclusion

The article describes the original principles and the approach to automated data analysis in social media to quantify the social consequences (unrest) that inevitably arise after each accident or natural disaster associated with the cessation/deterioration of the functioning of urban life support infrastructure. The article promotes (1) a more accurate assessment of the time of adaptation and recovery of a community at risk, (2) competent socio-economic elimination of the consequences of urban accidents by decision makers.

References

[1] Timashev S.A., Malyutina E.V. Entropic analysis of social activity during and after major accidents of urban infrastructure. — Yekaterinburg: Uralsky Rabochy, 2023.

[2] Timashev S., Malyutina E. Entropy Analysis of Social Unrest after Large Urban Infrastructure Accidents // ASCE INSPIRE. 2023. P. 631–639. DOI: 10.1061/9780784485163.074.

[3] Al-Quran M. W. M. Traditional media versus social media: challenges and opportunities. Technium: Romanian Journal of Applied Sciences and Technology, 4 (10), 145–160. – 2022.

[4] The number of permanent residents of the Russian Federation by municipality as of January 1, 2024 // Federal State Statistics Service: [Electronic resource]. URL: https://rosstat.gov.ru/compendium/document/13282