Вісник Національного технічного університету "ХПІ", 2026, № 1 (15)
ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

UDC 004 **DOI: 10.20998/2411-0558.2026.01.10**

D. UHRYN, Dr. Sci. Tech., Prof., Yuriy Fedkovych CHNU, Chernivtsi,
A. KALANCHA, PhD student, Yuriy Fedkovych CHNU, Chernivtsi

FORMATION AND ANALYSIS OF INFORMATION CASCADES BASED ON TIMED SEMANTIC INFLUENCE

This study is devoted to the analysis of information flows and the search for hidden connections between Telegram news channels. The main goal is to develop a method to help identify sources with the most significant influence and model how information spreads on the network. The work proposes the Timed Semantic Influence (TSI) algorithm – it allows you to assess the connections between messages, taking into account not only the content but also the time of publication. Thanks to this method, the structure of news distribution is built in the form of tree-like cascades from the initial message to the channels that picked it up. As a result, the channel that most often triggers news waves was identified. Among the topics that caused the most excellent resonance after eliminating general air alerts, news about significant geopolitical events and international support for Ukraine dominated. The proposed approach opens up the opportunity to identify key players in the media space and determine the topics that most influence public opinion. Tabl. 1. Ref.: 11 items.

Keywords: information source; text similarity; information cascade; influence tree, time.

Introduction. Information flows are becoming more complex and multifaceted in today's digital world, in addition to expanding in size. It is very challenging to visually track or intuitively interpret the hidden connections between various information sources due to the rapidity of dissemination and the variety of communication platforms. These issues have been exacerbated by the growth of social media and messaging apps, which have produced enormous ecosystems where information is continuously repackaged, reshaped, and disseminated. Traditional analysis techniques frequently fail in this context because they cannot account for the temporal dynamics and semantic changes that take place as messages propagate across networks.

The use of natural language processing (NLP) techniques opens up new avenues for identifying the patterns that underlie information flows and differentiating between sources that merely repeat preexisting narratives and those that actively influence public opinion. Our study lies at the nexus of social network analysis and computational linguistics, with a particular emphasis on

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Telegram news channels as a vital information-sharing platform. We seek to uncover latent structures through statistical analysis of channel relationships, which could indicate intentional dissemination strategies, shared content origins, or indications of coordinated activity [1 – 2].

Cases where various channels air similar stories or propaganda messages are given special attention, which calls into question their objectivity and possible ties to organized influence networks. A closer look frequently reveals consistent patterns in the timing of publications and the semantic overlap of their content, despite the initial impression that these channels are unrelated. This presents significant questions for comprehending the formation of public opinion, the initiation of information waves, and the key players in the dissemination process.

In the current media environment, where a single news item can set off a swift information storm spanning hundreds of channels, the issue we tackle is especially pertinent [3 - 4]. It is still very difficult to determine the initial cause of these waves and to explain why the same subject is amplified nearly simultaneously by several channels. With millions of posts every day and thousands of active channels, the volume of digital communication makes manual analysis nearly impossible. Furthermore, the intricacy is increased by the fact that channel writers hardly ever explicitly reference or provide direct citations to their sources. Rather, they frequently retell or reframe information in their own words, making it more audience-specific and hiding its source. Such transformations obscure visible connections, but advanced computational methods can reveal statistically significant semantic and temporal patterns, even when textual similarities are subtle.

Purpose of the article. Scientific and technical methods that can automatically identify and analyze relationships in vast amounts of textual data are becoming more and more necessary in light of these difficulties. Since the timing of publications is a critical component in determining influence and information flow, an effective tool must not only measure semantic similarity between messages but also take the temporal dimension into account.

Our primary focus is to distinguish between channels that act as true opinion leaders – initiating influential messages and shaping narratives – and those that primarily react to or replicate already distributed content. By implementing the proposed Timed Semantic Influence (TSI) algorithm, we seek

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

to combine semantic analysis with temporal modeling to quantify influence more accurately. Ultimately, the study aims to produce specific statistical evidence that demonstrates the effectiveness of the proposed approach on a test dataset. These results will not only validate the methodology but also create a foundation for applying it to larger and more diverse data collections in future research. In this way, the work contributes both a theoretical framework and a practical toolset for analyzing complex information flows in modern media ecosystems.

The goal of this study is to create a thorough methodology that models the dynamics of information dissemination, finds the most influential sources, and makes it possible to uncover hidden relationships between Telegram channels. In order to compare information sources methodically, evaluate their relative influence, and reconstruct the structure of the digital information environment being studied, it is necessary to establish additional characteristics of the sources.

Object and subject of the research. The phenomenon of information dissemination in digital communication networks – specifically, Ukrainian Telegram news channels – is the focus of the study. It includes the processes by which news messages are produced, copied, and disseminated through various media, creating intricate information flow structures. The processes through which these flows develop into extensive information cascades and the patterns of interaction among various sources that help to shape the media landscape are given particular attention.

Using the Timed Semantic Influence (TSI) algorithm, the study examines the temporal and semantic relationships between news messages posted on Telegram channels. Text similarity, time-sensitive dissemination, and structural dependencies that create cascade-like trees of information spread are among the quantifiable aspects of influence between channels that are the subject of this study. According to this framework, the study examines which media outlets serve as the originators of powerful messages, how information waves develop, and which subjects have the greatest resonance in the media, all of which influence public opinion and perception.

Timed Semantic Influence and Its Application for Detecting Information Cascades. To analyze the relationships between information sources, in particular Telegram channels, the Timed Semantic Influence (TSI)

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

algorithm was developed. This approach allows us to go beyond a simple comparison of textual similarity, taking into account key aspects - the temporal order of publications and the semantic proximity of the content [5 – 6].

The main idea is that the informational influence of one channel on another can be detected by analyzing how quickly similar content appears on other resources after the publication of the original message. TSI is designed to quantify the strength of such influence and determine its direction. The channel that first publishes a particular message is considered a potential source of influence, while another channel subsequently distributes a semantically similar message. To calculate this influence, a mathematical model is used that takes into account the exponential decay of the time factor (see Formula 1). This choice was made after comparison with the power-law model since the exponential function demonstrated greater stability and predictability of results when analyzing real data, while the power-law function turned out to be too sensitive to fluctuations at the time of publication.

$$TSI = sim(m_a, m_b) \times e^{-\alpha \frac{\Delta t}{60}} \quad (1)$$

Here, $sim(m_a, m_b)$ denotes the similarity value between the message texts ma and mb of the corresponding channels A and B, which ranges from 0 (i.e., complete difference) to 1 (complete semantic identity). The component Δt represents the time interval in minutes between the publications of the compared messages. A smaller value of this interval indicates a higher potential level of influence since it means more rapid dissemination of information. The coefficient α is an experimentally determined parameter that regulates the weight of the time factor in the formula. Its increase enhances the significance of the time component, while a decrease shifts the emphasis to the semantic similarity of the texts.

Thus, the TSI indicator comprehensively assesses both the textual proximity of information and the speed of its dissemination. A higher TSI value indicates a more substantial potential influence of one channel on the other. After calculating the TSI for each detected pair of similar messages, the direction of influence is established unambiguously: the source is considered to be the channel that published the corresponding message first.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

The TSI algorithm for analyzing the interaction between two selected Telegram channels, conventionally designated as channel A and channel B, is implemented in several stages. First of all, a specific message, M, is selected from channel A, which will serve as a benchmark with a precisely fixed publication time. For successful analysis, it is crucial that both channels demonstrate sufficient activity during the period under study because a small amount of data can lead to unrepresentative conclusions. Next, for this message M, an array N is formed from channel B messages that were published within the so-called "event horizon" – a time interval of ±6 hours from the moment of publication of M. Messages that do not fall into this range are considered irrelevant for this analysis. The next step is to find the optimal pair: message M is compared by cosine similarity with each message from the array N. The pair is selected for which the similarity value is maximum and at the same time exceeds the preset threshold S. If there is no such excess, message M is ignored, and the algorithm moves on to the following message of channel A. After successfully selecting the relevant message pair, the TSI indicator is calculated according to the formula above, and the direction of influence is determined.

Having previously calculated data on the mutual influence of messages between different information sources, we can proceed to the next stage – modeling of information dissemination cascades [7]. This allows us to visualize and structure how certain information waves diverge from one source to another in the media space.

The process of constructing such cascades begins with the identification of "root" messages. The root message is considered to be the one that initiated the information sequence and which, according to our data on the influence, was not influenced by any other message from the available analyzed set. That is, this message has never appeared as a target of influence. Each such identified root message becomes the beginning of a separate information cascade, or, in other words, the root of the tree (unidirectional graph) of information dissemination.

Next, the algorithm begins a recursive search for child nodes (messages) for each such root message. Based on the pre-computed impact data, the system finds all messages in other channels that are directly affected by the current (root or already child) message. These detected affected messages are added to the tree as a child node.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

A key constraint when constructing these branches is to avoid cyclic propagation within a single channel on a single cascade path. That is, if a message from channel A affected a message from channel B, then when further searching for descendants for a message from channel B, the system will search for impacts only on those channels that have not yet been encountered on the path from the initial root message to the current message from channel B.

The process continues recursively for each child message: for it, affected messages are also searched, subject to the unique condition of the channels in the current branch. The recursion exits under two conditions: when there is no other message in it that would be affected by it or when all channels available for influence (that were not yet in this branch) have already been reviewed.

As a result of this algorithm, we obtain an array of trees. Each tree represents a separate information cascade. The structure of each node in such a tree is unified and looks as follows: a dictionary containing the message identifier, the name of the channel in which this message was published, and a list of child messages. If the message has no descendants in the cascade (i.e., it did not affect other messages), then the list of child messages will be empty. Thus, we obtain a visual and structured representation of the information distribution paths that arise from each independent information impulse.

1. Information Cascades Processing and Analysis. In this case, let's consider the results of the analysis based on the specific data under study. In total, 887 information cascades were identified. After applying the degree filter (as already mentioned - at least three distribution levels), 782 cascades remained for further detailed analysis. This indicates that the vast majority of the identified information waves had a fairly significant distribution.

2. A key indicator for identifying trends that set information trends is the frequency of their appearance as the initiator of these significant, wider cascades. An information source whose messages often become the beginning of a long chain of message distribution can be considered one that generates influential content that is picked up by other sources.

Looking at the data in Table 1, we can see that the source kievreal1 clearly stands out as the main driver of wide information cascades. It accounts for about a third of all such cases (33.12%), making it the strongest initiator among the listed sources. In other words, kievreal1 often acts as the spark that sets off discussions and sharing across other platforms, serving as either the original

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

point of publication or one of the most influential early nodes that others pick up from.

Table 1

Channel Leaderships in Information Cascades

Channel	Number of occurrences at the root of the cascade	Leadership percentage
kievreal1	259	33.12%
truexanewsua	138	17.65%
UaOnlii	112	14.32%
voynareal	83	10.61%
suspilnenews	80	10.23%
UkraineNow	56	7.16%
lachentyt	28	3.58%
ukr24_7	26	3.32%

The next most active sources are truexanewsua and UaOnlii, with 17.65% and 14.32% respectively. While they don’t reach the same level as kievreal1, they still play a major role in setting the information agenda. Content from these sources tends to catch on quickly, often sparking waves of conversation that spread further into the broader information space.

A bit further down, we find voynareal (10.61%) and suspilnenews (10.23%). Their figures are close to each other, which suggests they have a similar level of influence. While not as dominant as the top three, they still contribute significantly as sources whose material frequently gains traction and gets redistributed across other platforms.

Meanwhile, sources like UkraineNow (7.16%), lachentyt (3.58%), and ukr24_7 (3.32%) play a noticeably smaller role. They are far less likely to trigger wide information cascades, which might mean their original content doesn’t resonate as strongly with broader audiences, or simply that their visibility is lower compared to the more influential accounts.

Thus, analyzing root messages in broad cascades allows us to identify sources that most often set the tone in the information environment by publishing content that other participants in the environment as mentioned earlier actively distribute.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Analysis of the Most Influential Messages. After identifying the sources that most often initiate information cascades, the next logical step in our study is to analyze the messages themselves that began these disseminations. Of particular interest are those messages that were not simply picked up but became the impetus for the most extensive cascades. Such messages can be considered the most influential because they managed to penetrate a significant number of levels of dissemination and cover a wide range of information sources.

To do this, we select those cascades demonstrating the most remarkable dissemination breadth. The root messages of these widest cascades are the primary information units that caused the most excellent resonance and lived the longest in the information environment, passing from one source to another.

Based on a visual analysis of the content of such root messages that led to the widest cascades, we can state several key observations: a significant part of these messages concerns the topic of air alerts and warning the population about danger. This is quite logical since such warnings are of critical importance, and sources strive to convey them to their readers as quickly as possible, which naturally leads to the simultaneous or almost simultaneous dissemination of similar information and, as a result, to the formation of wider cascades. Nevertheless, such messages do not carry much significance because, firstly, they do not contain a thought load, and secondly, they do not include specific tokens (words). However, if we filter out the messages related to air alerts and analyze the remaining influential information impulses, another clear trend emerges. The remaining messages mainly concern topics of resonant geopolitical decisions and events that have a significant and strategic impact on Ukraine. A vivid example is news about the approval of aid packages for Ukraine by foreign states or other critical international statements and decisions of leaders of Western countries. This type of content, which carries high social significance and potentially influences the development of events, becomes a catalyst for the formation of broad and long-lasting information cascades, spreading through a large number of sources.

Conclusion. The modern information space, especially in such a dynamic segment as Telegram news channels, poses significant challenges for analyzing the dissemination of information and identifying true sources of influence. Traditional methods, which are mostly based on text similarity analysis, often prove insufficient, as they do not take into account key aspects such as the temporal order of publications and the direction of the information flow. This can lead to an incomplete understanding of the interaction between channels and complicates the identification of coordinated information campaigns. Therefore, the main goal of this study was to develop and test a comprehensive approach that would allow not only to compare information sources by the content of their messages but also to quantitatively assess their mutual influence and model the dynamics of the dissemination of information waves. To achieve this goal, an algorithm was developed and presented that operates in two main stages. The first stage consists of calculating the Timed Semantic Influence (TSI) indicator between messages from different channels. This indicator integrates the semantic similarity of texts (determined using cosine similarity) and the time interval between their publications using an exponential decay model of influence. The result of this module is a structured data set that records in detail the detected connections between pairs of messages, a quantitative assessment of the strength of this connection (TSI value), and a determined direction of influence. Thus, this stage allows us to quantitatively assess how quickly and how similar in content the content spreads between the analyzed sources.

The second module of the algorithm, based on the obtained data set on influences, is designed to build information cascades. An information cascade models the process of spreading a specific information impulse from one channel to another. These cascades are visualized in the form of hierarchical trees, where the root node represents the initial message that initiated the wave, and the subsequent levels and branches of the tree reflect the paths of its further spread through other channels, taking into account time delays and established influence relationships.

The developed approach was tested on the basis of data collected from eight popular Ukrainian Telegram news channels. The study successfully identified a significant number of information cascades, which confirmed the effectiveness of the proposed methodology. Further analysis of these cascades, in particular those with a distribution depth of three or more levels, allowed us to determine the relative leadership of individual sources in the information space. Channels were identified as much more likely than others to act as initiators (root nodes) of such deep information distributions, which indicates their significant role in shaping the information agenda. In addition to analyzing the role of channels, the topics of the most influential messages that began the deepest cascades were studied. The results showed that a significant part of such

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

messages concern air raid warnings, which is expected for news channels in today's conditions. At the same time, after excluding such operational warnings, messages about resonant geopolitical events and decisions, such as the approval of international aid to Ukraine, remain among the most influential, emphasizing their high social significance.

The conducted research and the obtained results open several promising directions for further work. First, it is relevant to expand the database of the studied data. This includes both increasing the number of analyzed Telegram channels and extending the time period of data collection, which will allow obtaining more representative, statistically significant, and generalized conclusions about the dynamics of the information space. Second, there is an urgent need to improve the method of comparing individual messages, especially those containing a small amount of text. Thirdly, an important task remains the further careful selection and optimization of the parameters of the developed algorithms to achieve the optimal balance between accuracy, completeness of connection detection and computational efficiency. Finally, a promising direction is the use of clustering of information sources not only on the basis of their lexical similarity [8 - 9], as was done in previous works, but also on the basis of the calculated TSI values and the characteristics of their participation in information cascades. This will allow clutering [10 – 11] channels according to similar patterns of information influence and behavior, which can provide a new understanding of the structure and dynamics of the information environment.

Refrences:

1. Buehling, K. (2023), “Message deletion on Telegram: Affected data types and implications for computational analysis” *Commun. Methods Meas.*, Vol. 18, No. 1, pp. 92–114.

2. Devang, S., Hriday, R. Hui Xian Ng, L., and Swapneel, M. (2021) “Bridging nodes and narrative flows: Identifying intervention targets for disinformation on Telegram” *arXiv preprint.*

3. Guille, A., Hacid, H., Favre, C., and Zighed, D.A. (2013), “Information diffusion in online social networks: A survey” *ACM SIGMOD Rec.*, Vol. 42, No. 2, pp. 17–28.

4. Cinelli, M., Cresci, S., Quattrociocchi, W., Tesconi, M., and Zola, P. (2022), “Coordinated inauthentic behavior and information spreading on Twitter” *Decis. Support Syst.*, Vol. 164, No. 113819.

5. Zhao, L., Wang, J., Huang, R., and Xiao, G. (2016), “Effects of time-dependent diffusion behaviors on the rumor spreading in social networks” *Physica A*, Vol. 452, pp. 1–11.

6. Ducci, F., Kraus, M., and Feuerriegel, S. (2020), “Cascade-LSTM,” *in Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*

7. Yanchenko, E., Murata, T., and Holme, P. (2024), “Influence maximization on temporal networks: A review” *Appl. Netw. Sci.*, Vol. 9, No. 1, 2024.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

8. Dasgupta, J., Mishra, P.K., Karuppasamy, S., and Mahajan, A.D. (2023), "A survey of numerous text similarity approach" *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, pp. 184–194.
9. Gahman, N. and Elangovan, V. (2023), "A comparison of document similarity algorithms" *arxiv preprint.*
10. Guan, R., Shi, X., Marchese, M., Yang, C., and Liang, Y. (2011), "Text clustering with seeds affinity propagation" *IEEE Trans. Knowl. Data Eng.*, Vol. 23, No. 4, pp. 627–637.
11. Janani, R. and Vijayarani, S. (2019), "Text document clustering using spectral clustering algorithm with particle swarm optimization," *Expert Syst. Appl.*, Vol. 134, pp. 192–200.

Статтю представив д-р техн. наук, проф. НТУ "ХПІ" Носков В.І.

Надійшла (received) 10.12.2025
Стаття прийнята до друку 23.01.2026
Опублікована 27.02.2026

Dmytro Uhryn, Dr.Sci.Tech, Professor
Yuriy Fedkovych Chernivtsi National University
2 Kotsyubinsky Street, Chernivtsi, Ukraine, 58012,
Tel: +38 (050) 989-15-46 , e-mail: d.ugryn@chnu.edu.ua
ORCID ID: 0000-0003-4858-4511

Kalancha Artem, PhD student
Yuriy Fedkovych Chernivtsi National University
2 Kotsyubinsky Street, Chernivtsi, Ukraine, 58012
Tel: +38 (067) 965-97-89, e-mail: kalancha.artem@chnu.edu.ua
ORCID ID: 0009-0004-1451-7470

УДК 004

Формування та аналіз інформаційних каскадів на основі часового семантичного впливу / Угрин Д.І., Каланча А.Д. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2026. – № 1 (15). – С. 155 – 166.

Це дослідження присвячене аналізу інформаційних потоків та пошуку прихованих зв'язків між новинними каналами Telegram. Головною метою є розробка методу, який допоможе визначити джерела з найбільш значним впливом та змоделювати поширення інформації в мережі. У роботі запропоновано алгоритм часового семантичного впливу (TSI) – він дозволяє оцінювати зв'язки між повідомленнями, враховуючи не лише зміст, але й час публікації. Завдяки цьому методу структура поширення новин будується у вигляді деревоподібних каскадів від початкового повідомлення до каналів, які його підхопили. В результаті було визначено канал, який найчастіше запускає новинні хвилі. Серед тем, які викликали найбільший резонанс після усунення загальних ефірних оповіщень, домінували новини про значні геополітичні події та міжнародну підтримку України. Запропонований підхід відкриває можливість визначити ключових гравців у медіапросторі та визначити теми, які найбільше впливають на громадську думку. Табл.: 1. Бібліогр.: 11 назв.

Ключові слова: джерело інформації; текстова схожість; інформаційний каскад; дерево впливу, час.

UDC 004

Formation and analysis of information cascades based on timed semantic influence / Uhryn D.I., Kalancha A.D. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkiv: NTU "KhPI". – 2026. – № 1 (15). – P. 155 – 166.

This study is devoted to the analysis of information flows and the search for hidden connections between Telegram news channels. The main goal is to develop a method to help identify sources with the most significant influence and model how information spreads on the network. The work proposes the Timed Semantic Influence (TSI) algorithm - it allows you to assess the connections between messages, taking into account not only the content but also the time of publication. Thanks to this method, the structure of news distribution is built in the form of tree-like cascades from the initial message to the channels that picked it up. As a result, the channel that most often triggers news waves was identified. Among the topics that caused the most excellent resonance after eliminating general air alerts, news about significant geopolitical events and international support for Ukraine dominated. The proposed approach opens up the opportunity to identify key players in the media space and determine the topics that most influence public opinion. Tabl.: 1. Bibliogr.: 11 items.

Keywords: information source; text similarity; information cascade; influence tree, time.