**Information systems. Computer sciences. Issues of information security**

**Информационные системы. Информатика. Проблемы информационной безопасности**

*UDC 004.652*
*https://doi.org/10.32362/2500-316X-2023-11-2-33-49*

REVIEW ARTICLE

# Models and methods for analyzing complex networks and social network structures

**Julia P. Perova @, Vitaly R. Grigoriev, Dmitry O. Zhukov**

*MIREA – Russian Technological University, Moscow, 119454 Russia*
*@ Corresponding author, e-mail: perova_yu@mirea.ru*

**Abstract**
**Objectives.** The study aimed to investigate contemporary models, methods, and tools used for analyzing complex social network structures, both on the basis of ready-made solutions in the form of services and software, as well as proprietary applications developed using the Python programming language. Such studies make it possible not only to predict the dynamics of social processes (changes in social attitudes), but also to identify trends in socioeconomic development by monitoring users' opinions on important economic and social issues, both at the level of individual territorial entities (for example, districts, settlements of small towns, etc.) and wider regions.
**Methods.** Dynamic models and stochastic dynamics analysis methods, which take into account the possibility of self-organization and the presence of memory, are used along with user deanonymization methods and recommendation systems, as well as statistical methods for analyzing profiles in social networks. Numerical modeling methods for analyzing complex networks and processes occurring in them are considered and described in detail. Special attention is paid to data processing in complex network structures using the Python language and its various available libraries.
**Results.** The specifics of the tasks to be solved in the study of complex network structures and their interdisciplinarity associated with the use of methods of system analysis are described in terms of the theory of complex networks, text analytics, and computational linguistics. In particular, the dynamic models of processes observed in complex social network systems, as well as the structural characteristics of such networks and their relationship with the observed dynamic processes including using the theory of constructing dynamic graphs are studied. The use of neural networks to predict the evolution of dynamic processes and structure of complex social systems is investigated. When creating models describing the observed processes, attention is focused on the use of computational linguistics methods to extract knowledge from text messages of users of social networks.
**Conclusions.** Network analysis can be used to structure models of interaction between social units: people, collectives, organizations, etc. Compared with other methods, the network approach has the undeniable advantage of operating with data at different levels of research to ensure its continuity. Since communication in social networks almost entirely consists of text messages and various publications, almost all relevant studies use textual analysis methods in conjunction with machine learning and artificial intelligence technologies. Of these, convolutional neural networks demonstrated the best results. However, the use of support vector and decision tree methods should also be mentioned, since these contributed considerably to accuracy. In addition, statistical methods are used to compile data samples and analyze obtained results.

**Keywords:** social networks, modeling of social processes, oriented graphs, multilayer convolutional neural network, computational linguistics, clustering

• **Submitted:** 07.12.2021 • **Revised:** 23.12.2022 • **Accepted:** 09.02.2023

**For citation:** Perova J.P., Grigoriev V.R., Zhukov D.O. Models and methods for analyzing complex networks and social network structures. *Russ. Technol. J.* 2023;11(2):33–49. https://doi.org/10.32362/2500-316X-2023-11-2-33-49

**Financial disclosure:** The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

ОБЗОР

# Модели и методы анализа сложных сетей и социальных сетевых структур

**Ю.П. Перова @, В.Р. Григорьев, Д.О. Жуков**

*МИРЭА – Российский технологический университет, Москва, 119454 Россия*
*@ Автор для переписки, e-mail: perova_yu@mirea.ru*

**Резюме**
**Цели.** Целью статьи является исследование современных моделей и методов анализа сложных социальных сетевых структур и применяемых для этого инструментов, как на основе готовых решений в виде сервисов и программного обеспечения, так и средств разработки собственных приложений с использованием языка программирования Python. Такие исследования позволяют прогнозировать не только динамику общественных процессов (изменение социальных настроений), но и тенденции социально-экономического развития за счет мониторинга мнений пользователей по важным экономическим и социальным вопросам на уровне отдельных территориальных образований (районов, поселений небольших городов и т.д.) и регионов.
**Методы.** Рассмотрены и подробно описаны динамические модели и методы анализа стохастической динамики изменения состояний, учитывающие процессы самоорганизации и наличие памяти; методы деанонимизации пользователей; рекомендательные системы; статистические исследования, использующие методы анализа профилей в социальных сетях; методы численного моделирования для анализа сложных сетей и протекающих в них процессов. Особое внимание уделено обработке данных в сложных сетевых структурах средствами языка Python и применению его библиотек.
**Результаты.** Описана специфика решаемых задач при исследовании сложных сетевых структур и их междисциплинарность, связанная с использованием методов системного анализа, теории сложных сетей, текстовой аналитики и компьютерной лингвистики. В частности, исследованы динамические модели процессов, наблюдаемых в сложных социальных сетевых системах, структурные характеристики таких сетей и их взаимосвязь с наблюдаемыми динамическими процессами, в т.ч., с использованием теории построения динамических графов. Исследовано применение нейронных сетей для прогнозирования эволюции динамических процессов, наблюдаемых в сложных социальных системах, и их структуры. Значительное внимание уделено применению методов компьютерной лингвистики, что необходимо для извлечения знаний из текстовых сообщений пользователей социальных сетей при создании моделей, описывающих наблюдаемые процессы.
**Выводы.** Сетевой анализ помогает структурировать модели взаимодействия между социальными единицами: людьми, коллективами, организациями и т.д. По сравнению с другими методами сетевой подход имеет одно неоспоримое преимущество: он позволяет оперировать данными на разных уровнях исследования – от микро- до макроуровня, обеспечивает преемственность этих данных. Установлено, что практически все исследования используют методы работы с текстом, т.к. общение в социальных сетях почти полностью состоит из текстовых сообщений и публикаций. В большинстве исследований используются технологии машинного обучения и искусственного интеллекта. Лучший результат показали сверточные нейронные сети. Из используемых методов также следует выделить метод опорных векторов и дерево решений, т.к. именно они показывали самую высокую точность. Для составления выборок данных и правильного анализа полученных результатов применялись статистические методы.

**Ключевые слова:** социальные сети, моделирование социальных процессов, ориентированные графы, многослойная сверточная нейронная сеть, компьютерная лингвистика, кластеризация

• **Поступила:** 07.12.2021 • **Доработана:** 23.12.2022 • **Принята к опубликованию:** 09.02.2023

**Для цитирования:** Перова Ю.П., Григорьев В.Р., Жуков Д.О. Модели и методы анализа сложных сетей и социальных сетевых структур. *Russ. Technol. J.* 2023;11(2):33–49. https://doi.org/10.32362/2500-316X-2023-11-2-33-49

**Прозрачность финансовой деятельности:** Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

## INTRODUCTION

The study of social networks and the modeling of the processes observed in them is a very important scientific and practical task, since it allows predicting the dynamics of changes in the sentiments of their users and thereby ensuring the management of social processes in the interests of stable economic development.

The present review sets out to characterize the specifics of the research area, formulate its main objectives, indicate any links with other sciences, and give a brief overview of the main approaches and resources used. Dynamic models of complex social systems and network structures are discussed along with the characteristics of complex networks and observed processes, including those based on graph construction and data analysis using the Python programming language. In addition, issues involved in the use of neural networks to make the necessary evolution forecasts processes observed in complex social systems and network structures are discussed. Considerable attention in the review is paid to methods of computational linguistics used for extracting knowledge from text messages of social network users when creating models that describe the observed processes.

## SOCIAL NETWORKS AND THEIR GENERAL PROPERTIES

A social network comprises a structure together with a set of objects and relations defined in relation to it. In terms of the number of users, the largest social networks include Facebook[1] (banned in Russia), VKontakte[2], Odnoklassniki.ru[3], YouTube[4], etc. The term "social network" refers to the concentration of social objects that can be considered in terms of a network (or graph), whose nodes are objects, and whose links are social relations [1]. Today, the term "social network" denotes a wider concept than that implied by its original purely social aspect: the term covers, for example, many information networks including the world-wide web itself. Formally, any complex social network structure can be represented by a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the set of graph vertices, and $\mathbf{E}$ is the set of graph edges. In a social network graph, the vertices are the participants (or actors), while the edges indicate the existence of relationships between them. Relations can be either directed (directed graph) or undirected. In the theory of complex networks, there are three main areas: the study of statistical properties that characterize networks; the creation of network models; predicting the behavior of networks and the processes observed in them when their structural properties change, including, as a result of destructive impacts on them.

Social network analysis (SNA) is widely used in a number of applications and disciplines. Some common applications of network analysis include data collection and accumulation, network propagation modeling, network and sample modeling, feature and user behavior analysis, community-provided resource support, location-based interaction analysis, social sharing and selection, recommendation systems development, as well as link prediction and object analysis. In the private sector, firms use social media analysis to support activities such as customer interaction and analysis, marketing, and business intelligence. The public sector's use of SNA includes the development of leadership participation strategies, analysis of individual and group participation, use of the media, and community-based problem solving.

SNA thus represents an efficient system for discovering and interpreting public online connections. These can be explored using a range of analytical techniques, ranging from simple centrality measures to multilevel modeling. If data collection formerly represented a task that required a lot of effort and time, today's electronic networks have somewhat simplified this task. This happened through the use of passive data (such as web pages and mail store data). However, due to the increase in efficiency leading to a limitation in data collection, it became necessary to determine criteria for determining relationship significance. Solving such problems requires high-level technical skills, in particular, knowledge of programming languages or related programs.

Given these issues and limitations, studies propose more efficient and reliable data collection methods in such networks. In addition, issues such as spoof node

[1] https://www.facebook.com/. Accessed December 07, 2021.
[2] https://vk.com/. Accessed September 20, 2022 (in Russ.).
[3] https://ok.ru/. Accessed September 20, 2022 (in Russ.).
[4] https://www.youtube.com/. Accessed September 20, 2022.

detection, as well as fake nodes and links, need to be studied.

In order to solve many applied problems currently in practice, sets of ready-made SNA tools are typically used. However, these have some limitations; in particular, they do not allow the development of new approaches and models for studying the observed processes. A detailed description of the tools and methods used in SNA can be found in the review [2].

When analyzing the structure of complex networks, as in graph theory, individual node-, entire network-, and network substructure parameters are studied. Nevertheless, some questions, such as the planarity of a graph for the theory of complex networks, are of little practical interest. Among topical problems in the study of complex networks, the following can be distinguished: determination of cliques in the network (cliques are subgroups or clusters in which nodes are more strongly interconnected than with members of other cliques); selection of components (parts of the network) that are not interconnected, but whose nodes are connected within these components; identifying blocks and jumpers (a node is called a jumper if, when it is removed, the network breaks up into unconnected parts); selection of groupings comprising clusters of equivalent nodes having the most similar link profiles.

## KEY AREAS OF RESEARCH IN COMPLEX NETWORKS AND APPLIED TOOLS

The theory of complex networks is a complex scientific area at the intersection of such sciences as discrete mathematics, graph theory, algorithm theory, nonlinear dynamics, the theory of phase transitions, the theory of percolation, and many others. Therefore, in order to successfully analyze and model complex networks, basic knowledge from all these areas is required. There are a large number of publications, including, for example, textbooks[5], which cover theoretical aspects of complex networks: characteristics, algorithms, models, search and ranking problems. The publications also provide information necessary for mathematical and computer modeling and analysis of complex networks.

The theory of complex networks covers the following problems:

1) study of standard characteristics of graphs of complex networks of different nature—random graphs, scale-free networks, small world networks, etc.;
2) determination and study of new characteristics of complex networks (for example, elasticity and survivability under destructive influences);
3) study of various "physical" processes on complex networks—diffusion, epidemic processes in society, the spread of various flows (for example, traffic in computer networks or vehicle flows in transport networks);
4) a very important direction in terms of application—methods for restoring, protecting and destroying networks, and solving issues of their optimization;
5) search for implicit or latent connections between participants, which can be very important for identifying members of criminal communities.

It should be noted that for the study of complex networks and identification of the main patterns of the processes occurring in them, methods are used that were first created for the study of natural science problems, in particular, methods of theoretical physics.

The theory of complex networks as a field of discrete mathematics studies the characteristics of networks, taking into account not only their topology, but also statistical phenomena, the distribution of weights of individual nodes and edges, the effects of leakage, percolation, and conductivity in such networks of current, liquid, information, etc. It turned out that the properties of many real networks differ significantly from the properties of classical random graphs.

The study of such parameters of complex networks as clustering, mediation, or vulnerability is directly related to the theory of survivability, since these are the properties that determine the ability of networks to maintain their operability in the event of a destructive effect on their individual nodes or edges (connections). Despite the fact that the theory of complex networks includes various networks—electrical, transport, information, the greatest contribution to the development of this theory was made by the study of social networks.

Due to a significant increase in the volume of textual information generated by Internet users and the need for automatic processing of texts in natural language in order to determine the state of the nodes of complex social networks (for example, opposition or loyalty), computational linguistics has now received a significant impetus for its development.

The task of computational linguistics can be formulated as the development of computer programs for automatic processing of texts in natural languages in order to extract knowledge, cluster texts into semantic groups, annotate, etc.

The source material for extracting the necessary linguistic information can be collections and corpora of texts. A corpus of texts is a collection of texts collected according to a certain principle of representativeness (by genre, authorship, etc.), in which all texts are marked up, i.e., are equipped with some linguistic markup (annotations)—morphological, accent,

[5] Snarsky A.A., Lande D.V. *Modeling of complex networks.* Textbook. Kyiv: NTUU KPI; 2015. 212 p. (in Russ.).

syntactic, etc.[6] Currently, there are at least a hundred different corpora—for different natural languages and with different markup. In Russia, the most famous is the National Corpus of the Russian Language[7]. Labeled corpora are created by linguists and are used both for linguistic research and for tuning (training) models and processors used in computational linguistics using well-known mathematical methods of machine learning.

Computational linguistics shows quite tangible results in various applications for automatic processing of texts in natural languages. Its further development depends on both the emergence of new applications and the independent development of various models of the language, in which many problems have not yet been solved. The most developed are the models of morphological analysis and synthesis. Syntax models have not yet been brought to the level of stable and efficient modules, despite the large number of proposed formalisms and methods. Even less studied and formalized are models of the level of semantics and pragmatics, although automatic processing of discourse is already required in a number of applications. Despite this, the already existing tools of computational linguistics itself, the use of machine learning and text corpora can significantly advance the solution of these problems.

It should be noted that computational linguistics is not only used to analyze information in complex social systems in order to determine the state of nodes, but also itself uses the achievements of the theory of complex networks. The first step in applying the theory of complex networks to text analysis is to represent this text as a collection of nodes and links, thereby building a language network. There are different ways of interpreting nodes and links, which leads, respectively, to different representations of the network of the language. Nodes can be connected to each other if the words corresponding to them are next to each other in the text, belong to the same sentence, are connected syntactically or semantically. The preservation of syntactic links between words leads to the image of the text in the form of a directed network, where the direction of the link corresponds to the subordination of the word.

The study of graph properties of complex networks is becoming increasingly popular due to the growing availability of scientific and social data presented in graph form. Because of this, many researchers have focused on developing improved graph neural network models. One of the main components of a graph neural network is the aggregation operator required to generate a graph-level representation from a set of node-level embeddings. The aggregation operator is of crucial importance, since it should, in principle, provide an isomorphism-invariant representation of the graph, that is, the representation of the graph must be a function of the nodes of the graph, considered as a set.

In [3], the DeepSets aggregation operator based on self-organizing maps (SOM) is considered to transform a set of node-level representations into a single graph-level. The adoption of SOM allows computation of representations of nodes that embed information about their mutual similarity. Experimental results on several real datasets show that the proposed approach provides improved predictive performance compared to conventional summing aggregation and many modern graph neural network architectures presented in the literature.

In the framework of paper [4], the architecture of convolutional neural networks was considered, including the types of layers used and the principles of their operation, settings, and training features. The possibility of searching and preventing information leaks from corporate information systems on the Internet is described. The architecture of convolutional neural networks for the primary processing of information on Internet pages is proposed: the types of layers that make up the network, their purpose and mathematical representation, as well as the hyperparameters used are described. The paper [4] presents the architecture of the network and the model of its training. The possibility of using networks of this type for solving problems of detecting leaks of confidential data is described, as well as existing solutions and approaches are analyzed. Approaches are considered that enable using convolutional neural networks to solve the problems of classifying web pages containing news and information sources, navigation and information sources based on their text content.

As an example of a finished system, we can refer to Bidirectional Encoder Representation Transformers (BERT) [5]—language representation model, which is designed for preliminary training of deep bidirectional representations on simple unmarked texts by combining the left and right contexts in all layers. This allows you to tune a pre-trained BERT model with just one additional output layer and get the most up-to-date results for a wide range of tasks.

Standard language representation models that existed before BERT, such as the OpenAI GPT (Generative Pre-Trained Transformer)[8] were unidirectional. This limited the choice of architectures that could be used for pre-training. For example, in OpenAI GPT, each token could only serve the previous token (from left to right) in the internal attention layer of the model. Tokens are intended

[6] Boyarsky K.K. *Introduction to computational linguistics*. Textbook. St. Petersburg: NIU ITMO; 2013. 72 p. (in Russ.).

[7] https://ruscorpora.ru/. Accessed September 20, 2022 (in Russ.).

[8] https://openai.com/api/. Accessed September 20, 2022 (in Russ.).

for electronic identification, which are provided to the user after successful authorization. In a sense, a token is an electronic key to access something.

This approach creates a number of limitations, therefore, for pre-training BERT, a masked language model is used, in which a certain number of tokens in the input data are randomly masked. The model then has to predict the original meaning of the masked words based on the context. This provides the ability to combine left and right contexts, which in turn allows a bidirectional view model to be pre-trained.

There are two stages in using BERT.

1. Preliminary training. The model is trained on unlabeled data by performing various tasks.
2. Fine tune. The model is loaded with pre-trained parameters and trained on labeled data from subsequent tasks.

Besides theoretical methods, numerical simulation is often used to analyze complex networks and the processes occurring in them. In addition, one of the most powerful and widely used tools for analyzing complex networks is data processing using the Python language and the libraries available for it [7–12]. A special package (Python-networkX) has been developed for the Python programming language—a toolkit for creating, manipulating and studying complex networks, which allows you to determine many of their characteristics. Here we can also mention the NATASHA[9] tools—an open library for the Python programming language, which allows you to extract structured information from texts in Russian. NATASHA has a concise interface and includes extractors for names, addresses, amounts of money, dates, and some other entities.

The Python language and the libraries written for it can be used to effectively solve a wide range of tasks for analyzing various data:

- multidimensional lists (matrices);
- tabular data, when data in different columns can be of different types (strings, numbers, dates, etc.). This includes data that are typically stored in relational databases or in files with commas as separators;
- data presented in the form of several tables interconnected by key columns (what in SQL is called primary and foreign keys);
- equally spaced and not equally spaced time series.

This list is far from complete. A significant portion of datasets can be converted to a structured form more suitable for analysis and modeling. In cases where this fails, it is possible to extract a structured set of features from the data set. For example, a selection of news articles can be converted into a word frequency table, to which sentiment analysis can then be applied.

[9] https://pypi.org/project/natasha/. Accessed September 20, 2022.

## ANALYSIS OF NETWORK STRUCTURES AND FORECASTING THE DYNAMICS OF SOCIAL PROCESSES

A review of published papers shows that SNA methods are useful tools for creating a complete picture of public sentiment. These methods are cheaper to implement than population survey methods and provide more data, since in surveys not all people express their real point of view. Based on this, it is possible to study the behavior of modern society in the era of the spread of social networks.

Considering the dynamic approach, namely, the direction in the study of social networks, in which the objects of research are changes in the network structure over time, it can be noted that structural analysis and analysis of the behavior of connections in social networks is necessary in order to determine the most important peaks, connections, communities and emerging regions of the network. Such an analysis allows an overview of the global evolutionary behavior of the network.

Community discovery in dynamic networks no longer requires complex mathematical heuristics. Using a simple comparison of time slices, it is possible to determine dynamically changing temporary communities of users of social network structures. The study of these dynamic communities makes it possible to significantly simplify the analysis of the dynamics of a complex system of social interactions as it develops over time.

In [13], the authors present the fundamental structures of dynamic social networks based on a high-resolution dataset describing a densely connected population of 1000 first-year students at a large European university. The authors look at physically short interactions measured using Bluetooth, supplemented with information from telecommunications networks (phone calls and text messages), online social networks, and geolocation and demographic data.

Human social communities are overlapped by individuals participating in several communities (in complex network theory, such nodes are called jumpers). During the week, meetings of the subjects of the created structure are held (such structures are called kernels). It can be both a meeting of friends outside the university, and all students. In a network of short physical interactions, meetings require that all participants be present at the same time and that they be in physical contact.

The location of members of kernels can also be predicted. The object that helps to do this is the kernels themselves. By observing the usual routes of the people who make up the kernel and their behavioral habits, it is possible to predict the geographical location of a person in the next time interval with high accuracy (on average in 93% of cases). This high accuracy proves that human mobility patterns are regular. It is also worth noting that kernel members have fewer location states than other individuals, resulting in lower values of information entropy on average.

The fact that geospatial exploration occurs as part of a social group but limited to specific time frames shows the complex interplay between time, location, and social context, and thus supports the hypothesis that sometimes, when people are most unpredictable in the geospatial domain, they exhibit predictable social behavior.

Linking the results of paper [13] to the literature on dynamic community detection, it can be noted that there are many methods that would allow the discovery of collections in everyday life, but in paper [13] it is used a simple matching of graph components to emphasize that emerging social structures are so obvious that these complicated methods are not needed.

Thus, the authors of [13] give a quantitative assessment of long-term patterns encoded in the microdynamics of a large system of interacting individuals, characterized by a high degree of order and predictability.

Let us consider one more work devoted to dynamic models [14]. Recent developments in the field of social networks have shifted the focus from static to dynamic representations, requiring new methods for their analysis and modeling. While social networks are shaped by a variety of processes, two specific mechanisms have been found to play a central role in their emergence and evolution. The first is the strategy of activation of social ties, that is the selection process leading to the creation of a new connection or the activation of an old one. It is clear that the activation of social ties is not accidental. Empirical observations show that people tend to allocate most of their social activities towards pre-existing strong ties, while allocating fewer interactions to create new social relationships or maintain weak ties. In other words, over time, some connections are used frequently in repetitive interactions, while others are not. The second mechanism is a surge of activity, that is, the activity of separate individuals develops through heterogeneous distributions of time between events. In addition, the propensity of individuals to participate in a social act per unit of time is also heterogeneous. In fact, empirical measurements on real datasets capturing various types of social dynamics show that activity is not uniformly distributed among people. In other words, not only do individuals exhibit heterogeneous propensities for social activity, but their activation is also explosive, and this explosive activity can significantly affect the evolution of networks. Although the study of these mechanisms has been the focus of a number of publications, a general framework for modeling is still lacking. Such a structure would make it possible to give an analytical characterization of how the interaction of heterogeneous patterns of activity and the mechanisms of selection of connections shape the evolution of social networks and, in turn, the processes occurring in them. To do this, the authors introduce a model of time-varying networks, which allows you to simultaneously control the relative strength of the burst of activity and the strategy for activating the connection. The asymptotic behavior of the model is solved analytically and a non-trivial phase diagram is found that governs the interaction of two processes. In particular, one observes the regime in which the surge controls the evolution of the network, and another area where the dynamics is completely determined by the process of selection of links. If the reuse of previously activated connections is strong enough and people tend to preferentially contact the same social circle, the spike leads to an amplification mechanism even in the presence of divergent time intervals between events, without having any effect on the evolution of the network. Thus, the structure proposed by the authors can be used to classify the temporal features of real networks and can give a new idea of the influence of social mechanisms on the processes of distribution in social networks.

In the paper [15] recommender systems are discussed. There are decision-making situations in the context of Internet information overload, when people have an overwhelming number of choices available, such as products to buy on an e-commerce site or restaurants to visit in a big city. Recommender systems (reciprocal recommender systems, RRS) emerged as a data-driven personalized decision support tool. They are able to process user related data, filter and recommend items based on the user's preferences, needs and/or behavior. Unlike most traditional recommendation approaches, where items are inanimate objects recommended to users and success is determined solely by the end user's response to the recommendation received, in RRS users become objects recommended to other users. Therefore, both the end user and the recommended user must accept the compliance recommendation in order to ensure successful RRS performance. The operation of RRS not only makes it possible to predict accurate preference estimates based on user interaction data, but also makes it possible to calculate mutual compatibility between pairs of users by applying processes for combining one-way preference information of each user.

In [16], the assessment of public opinion and public sentiment is carried out using a method based on a lexicon inherited from the classical approach to the analysis of public sentiment. The neural network determines the keywords which are later verified by subject matter experts. The program first analyzes the articles and documents and finds how often different words appear in the articles. After that, the program highlights the most frequently occurring words and makes them keywords. Based on them, the program builds a lexicon that is characteristic of the public mood based on news articles.

The program described in [17] uses the method of analyzing topics from a social network, which, in addition to collecting, processing and sorting information, also measures the time elapsed between publications so that later, based on these data, to create a time scale. Thus, as a result of the work of the program, a graph is obtained, according to which one can trace the growth and decline in the popularity of certain topics discussed in social networks. You can also trace what moods in society accompany these events and what is the time period of active discussion of certain topics.

The authors of paper [18] discuss a method for studying political climate in society using SNA, which was carried out using the search for keywords in the text that were previously entered into the program database. The main purpose of creating this program is to trace what made certain political parties popular and what topics are discussed the most. Also, using the program, you can find out how many people support a particular political party.

The SNA technique using neural networks is presented in [19]. It was used during presidential campaigns in order to trace the mood in society. This technique can be used as a replacement for traditional methods of public emotions analysis, because it has the ability to find and analyze radical opinions that is impossible to do with traditional methods.

The authors of [20] use the method of collecting and processing data from Twitter[10] (banned in Russia) accounts to determine gender, age, political preferences, and approximate place of residence. Machine learning is used to process the data, with the help of which the authors of the paper were able to collect information from users of the Twitter (banned in Russia) social network based on their posts, subscriptions and account information.

In [21], the authors discuss a method for determining a user's political preferences by analyzing user records belonging to different political groups. Through voluntary surveys of people belonging to different political groups, the program analyzes the language that is inherent in each of the groups and highlights its keywords for each of the groups. Based on these words, the program will later analyze the user's profile on the Twitter (banned in Russia) social network and, on this basis, determine which of the political groups the user belongs to. Based on the survey, in addition to political preferences, the gender and age of a person are determined, so that later it would be possible to compile statistics by comparing the gender, age and political preferences of users. After the survey, accounts left by users are analyzed to remove accounts belonging to other people who did not take part in the survey. Then, the last 3200 records are analyzed from the users' page. Based on these records, databases of keywords specific to a particular political group are compiled. Based on these databases, charts are created that show how often they occur in the records of people belonging to this group. Also, when the program analyzes people's records, their mood is revealed.

The paper [22] is a comprehensive analysis of the trace that each of us leaves daily on the Internet. We make purchases, communicate; many familiar things have long gone online. Every action that takes place on social networks does not go unnoticed. Each of us has a so-called digital footprint—the actions that we perform on the Internet and which remain there. This can be both our public information, which we ourselves leave on our pages on social networks and our non-public actions, information about which still remains on the network and can be extracted from there. The authors argue that this digital footprint reveals a lot about the user. In total, they identify 14 different demographic characteristics that they were able to establish using such social tracking.

Data is easy to collect and use when the user posts it on their profile, but if they prefer to remain secretive, there are so many ways to find out this information. The authors cite a few of them: you can analyze smartphone log files, likes on a social network, browser search history, frequency of hashtags, and many more things.

The authors also inform that according to many studies, people tend to communicate with those with whom they are in the same social group, and all members of this group often have similar behavioral traits and manner of communication. This opens up a huge scope for study.

Social networks and other services that we use collect our personal information. In most cases, we agree to this, but can a social network collect any information about a user who is not registered with it? The paper [23] considers the shadow profile hypothesis, according to which a social network can collect, based on the public information of users of this network and data from their phone books (if, for example, the user himself provides access to the social network) information about those people who are not registered in this social network. The paper proves the fact that a shadow profile as a structure can be created, and the larger the social network, the more accurate the shadow profile data will be.

[10] https://twitter.com/. Accessed December 07, 2021.

In the study [24], the authors use a large amount of information about the user account: the time of publication of each of the entries and the frequency of their publication; the number of publications containing geodata; reposts of records of other users; the number of liked publications; the number of responses to publications of other users; the number of mentions of other users; the number of publications containing media data and the average amount of media data per publication; the date the account was created; the number of followings and followers and many other data and ratios related to account information. This is an example of a comprehensive and complete study of a profile in a social network, which does not include the analysis of the media data itself and the analysis of the user's friends/followers. It should most accurately predict the user's age based on a large amount of information. This paper also describes in more detail the algorithms for creating a sample, training a neural network, and directly analyzing user profiles.

Since, as the network grows, the search for similarity between nodes in the network is a time-consuming process for optimization, the researchers in [25] use swarm algorithms to solve the problems of link prediction and community detection. Swarm-based optimization techniques used in SNA are compared in this paper with community analysis and connection analysis. As a future area of application, swarm-based optimization methods can be extended to the use of deep learning neural networks, especially for updating gradients when creating models of such networks.

A social network is a social structure with a set of social actors and social interactions between them. The study of the dynamics of these structures can be used to explain local and global economic patterns that are important for development. In [26], the discussed topic is the development and analysis of automatic control systems for making decisions about oil projects. Before deciding to invest in an oil project, engineers describe the project by providing sufficient economic data. Based on this data, a decision can be made to conduct a professional analysis to determine if the project is feasible. To automate the manual process and overcome the shortcomings of traditional evaluation methods (for example, expert evaluation depends on the quality of the choice of experts themselves), a back propagation neural network is used in the economic evaluation of oil projects.

A huge amount of work was carried out using analysis from the social network Twitter (banned in Russia). In the paper [27], the authors discuss the analysis of a community with extremist views using SNA and neural networks. According to the results of the study, it is proved that it is possible to analyze the community and find people associated with it, and possibly to predict the plans of this community in order to prevent terrorist activities.

The paper [28] refers to the definition of the personal qualities of users of the social network Twitter (banned in Russia) based on the records they made, as well as on the basis of their subscriptions. The program takes into account gender, age, education and political preferences to obtain a more accurate result of the study. Thus, on the basis of users' records and subscriptions, it is possible to determine the area of his or her interests, after which it is possible to draw up a chart showing the dependence of gender, age, etc. on the area of interest, and track which user groups prevail in a particular group of interests.

The study [29] again uses the collection of user data from about 1500 sites and the comparison of these data with the data of their accounts on the Twitter (banned in Russia) social network. Based on these data, a more accurate demographic model of users is built. The program analyzes the interests of users, using not only the data of their social networks, but also the data of the Quantcast.com[11] service, which allows collecting more accurate information. As a result, a table of user demographics and communities of interest to which they belong is made up.

The authors of [30] compare the data of users of the social network Twitter (banned in Russia) with political preferences. The program, based on demographic data and political preferences of a person, compiles statistics that show which groups of the population support which parties. This can be judged by the combination of such variables as gender, age, income, race, etc.

In [31–34], SNA is carried out in order to track the political mood of the population. Thus, thanks to SNA, it is possible to map support for various political parties, track public mood and find out the rating of political parties in different periods before and after elections, while associating these levels of support with various events that took place around the party.

The authors of [35, 36] carry out analysis using machine learning and text sentiment, which are one of the main tools of SNA, especially for the restoration of demographic characteristics, which requires knowledge in the field of machine learning and computational linguistics. When studying computational linguistics, one can find many different methods of analyzing written text besides sentiment analysis. Many of them, perhaps, will expand the toolkit for SNA or improve existing algorithms. With the help of machine learning methods, it is possible to automate the analysis process and make it much more convenient.

[11] https://www.quantcast.com/. Accessed September 20, 2022.

The studies described in [37–41] are aimed at analyzing profiles in social networks using gender classification, which solved the problem of face recognition using neural networks, algorithms that work using emoji emoticons in the source text, and determining the age and gender from a photo.

In [42], the authors describe a method for analyzing SMS messages in order to classify senders by gender and age. For the study, the authors used several algorithms with different structures of neural networks, as well as various methods of working with natural language, trying to achieve the best result. Ultimately, the best result for determining the age was shown by the support vector machine—an accuracy of about 71%. The best result of gender determination accuracy—almost 80%, was shown by decision tree J48. It is worth noting that various methods and filters of natural language processing only slightly affected the results, practically not improving either the accuracy or the speed of the algorithms.

The subject of research [43] is microblogs. The authors used the keyword analysis method. This approach fits well with the environment, in which the analysis is carried out, because people in microblogs often discuss some events, news, or discuss a certain topic. Using this method and machine learning methods, they were able to divide the initial sample into six age groups and identify the topic that participants in each age group most often discuss and express their thoughts on most often. It turned out that teenagers under 18 most often discuss sports; young people aged 18–25 talk about entertainment the most; people aged 25 to 30 see other goals, they want to firmly settle in life, so they mainly discuss family and business; older people (31–36 years old) are most interested in technology; further, users aged 26–40 begin to worry about their health and speak out more about it, while those over 40 like to discuss politics the most. Thus, the most frequent topic for discussion was determined for each age group. This does not mean that every member of the group necessarily discusses this topic, but it is more likely that the person discussing this topic belongs to this age group.

We can consider several statistical studies [44–46] that have widely used the method of analyzing profiles in social networks. Their purpose is to identify the social mobility of people based on their publications together by geodata. The authors found a large number of such publications; based on them an approximate map of the user's locations was created, the main centers of activity were identified, and the person's place of residence was established. Based on the place of residence, the names of people were found out. Further, using a database of names distributed by gender, it was possible to determine the gender of more than half of all the studied accounts. Using the last names, the researchers tried to determine information about the race and age of users—successfully in 38% and 14% of cases, respectively. This study showed that it is possible to establish some demographic characteristics, only by tracking a person or knowing his or her first and last name.

The paper [47] describes a method for determining the gender and age using the voice message function. The results are given, according to which the authors managed to achieve an accuracy of 80% in determining the gender and age of the speaker. This technology can be quite successfully used to analyze voice messages in social networks, if such a need arises, but in reality, such an algorithm is unlikely to find wide application in this area, since voice messages are sent personally to the recipient, and social network analysis is usually carried out publicly based on available information. In other industries, the value of such technology is difficult to overestimate: it should be useful in forensics, biometrics and for designing a system for speech recognition or recreation.

In addition to a regular text and the previously discussed additions to it—reposts, pictures, emoticons and subscriptions to other users, links and hashtags are also very often used on social networks. Links are used to share some content, be it a picture or news, and hashtags are used to indicate the topic of the publication and make it easier for other users to find this publication in the search. The authors of [48] suggest that the content found on the links shared by users and the hashtags they use can tell a lot about the age of these users. The researchers decided to analyze the posts of various users to determine their age group, but unlike many similar studies, they used not only the posts themselves, but also the content located on the links shared by users as well as recent posts with the same hashtag that the user mentions.

The paper [49] proposes a new method for predicting changes in complex social network structures based on the application of percolation theory and approaches adopted in stochastic dynamics. New results of computer modeling of the influence of the density of a social network on the threshold of its penetration are discussed. Percolation thresholds are calculated for various network densities and can be used in models that describe the stochastic dynamics of a system's transition from one state to another. The stochastic model presented in this paper provides the possibility of an abrupt transition of moods (states) of people in a social network for a very short period of time without any external influences, which is determined by the features of the system's self-organization and the memory of its nodes about previous states.

The developed model allows you to create an algorithm for monitoring social conditions based on the theory of percolation and stochastic dynamics, which can be easily applied in practice. The essence of this algorithm is as follows.

1. With the help of sociological monitoring, the average number of connections per person in a given social network is determined; then the proportion of negatively inclined people at a given time is determined ($t = 0$). The average density makes it possible to calculate the percolation threshold of the network structure, i.e., the share of network participants who have certain views, which allow these views to be freely distributed in the network.
2. After a fixed unit of time (day, week, etc.), the share of participants with certain views that are currently being investigated is determined. The change in this share compared to the previous share allows you to determine the value of the upward and downward trends.
3. Further, it is possible to use the received information about the trends, the percolation threshold and the initial share of network participants with certain representations to track and manage network participants.

The results obtained can be applied to the management of social processes. In one of the approaches first the sentiment in a social network is analyzed. Further, modern methods of psycholinguistic analysis based on artificial network technologies make it possible to attribute each user to a specific target group in accordance with his or her moods and views. The network nodes that can be identified as users that violate certain laws, such as those that spread extremist views and sentiments, are blocked. However, there are groups of nodes that do not violate any laws, but can potentially move into a group with extremist views. Since they do not break the law, they cannot be blocked, but it is still possible to limit their communication abilities using technical approaches, for example, by reducing the data transfer rate and reducing the number of other nodes or connections available to them. At the same time, the penetration threshold for information that can be freely transmitted over the network increases.

The authors[12, 13] of [50–52] propose a method for evaluating media in several modalities (topics, evaluation criteria/properties, classes), combining thematic modeling of text corpora and making multi-criteria decisions. The evaluation is based on the analysis of corpora in the following way: the conditional probability distribution of carriers by topics, properties and classes is calculated after the formation of a thematic model of corpora. Several approaches are used to obtain weights that describe how each topic relates to each evaluation criterion and to each class described in the document, including manual tagging, multi-enterprise approach, and automatic approach. The proposed multi-corporate approach involves assessing the thematic asymmetry of the corpora to obtain weights that describe the relationship of each topic to a certain criterion. These weights, in combination with the topic model, can be applied to evaluate each document in the corpora according to each of the considered criteria and classes. The proposed method was applied to a corpus of 804829 news publications from 40 Kazakhstan sources published from January 01, 2018 to December 31, 2019. The BigARTM model (200 topics) was obtained. The experiments confirm the general capability of media estimation using the topical text corpora model, as the classification task achieved a receiver performance area under the curve (ROC AUC) estimate of 0.81, which is comparable to the results obtained for the same task using the BERT model.

The developed system, in which the proposed model is integrated, allows solving classical problems such as simple reports or sentiment analysis. In addition, it also has a number of unique use cases that distinguish it from existing solutions: automatic analysis by topic, significant event and object without the need to generate queries based on keywords; analysis according to an arbitrary list of criteria not limited to sentiment, but also including social significance, popularity, manipulativeness, propaganda content, attitude to a certain country, attitude to a certain area, etc.; analysis of the dynamic behavior of topics; predictive analysis at the topic level.

In [53–55], the KroMFac method was proposed, which is used to detect a community by the regularized non-negative matrix factorization method based on the Kronecker graph model. KroMFac combines network analysis and community discovery techniques in a single unified framework.

The paper [56] is devoted to SNA and the development of methods for deanonymizing their users. Deanonymization refers to the identification of a user on the network or the true place of access to the network. After a comparative analysis of existing methods and models of user deanonymization [57–59], the authors propose a modified method based on the algorithm for combining selected vertices to defragment the deanonymization statement problem into smaller sub-problems that can

[12] Bushman B., Whitaker J. Media influence on behavior. Reference module. In: *Neuroscience and Biobehavioral Psychology*. 2017. http://scitechconnect.elsevier.com/neurorefmod/. Accessed November 24, 2020.

[13] Bandari R., Asur S., Huberman B.A. *The pulse of news in social media: Forecasting popularity*. https://arxiv.org/pdf/1202.0332.pdf. Accessed September 20, 2020.

be solved using existing methods. The main result of this work is the development of a new approach to optimization methods for identifying users of social networks based on the pairwise partitioning algorithm. The proposed algorithm improves the characteristics of existing deanonymization technologies, and is of theoretical and practical importance for the development of systems for modeling information actions in social networks.

The papers [60, 61] present the models developed by the authors for describing the stochastic dynamics of state changes in complex social systems, which take into account the processes of self-organization and the presence of memory. To create a model, graphical diagrams of the transition probabilities between the possible states of the described systems are considered, taking into account previous states that make it possible to take into account memory and describe not only Markovian, but also non-Markovian processes. Based on this approach, a nonlinear second-order differential equation is derived that allows one to formulate and solve boundary problems for determining the probability density function of the amplitude of parameter deviations that describe the observed processes of non-stationary time series depending on the values of the time interval of its determination and the depth of memory accounting. The resulting differential equation contains not only terms responsible for random change (diffusion) and ordered change (drift), but also contains a term responsible for the possibility of self-organization.

## CONCLUDING REMARKS

Network analysis can be used to structure models of interaction between social units: people, teams, organizations, etc. The network approach has an indisputable advantage as compared with other methods in terms of operating with data obtained at different research levels—from the micro to the macro level, to ensure data continuity. Network methods can also increase understanding by describing processes both theoretically and quantitatively. The relevance of network analysis is growing, since at the moment there is a globalization of the world processes, and above all, in the form of global networking.

Almost all studies use different methods of working with text, since communication in social networks consists almost entirely of text messages and publications. To apply these methods, knowledge in the field of computational linguistics is required. Here, sentiment analysis, lexical analysis, and keyword extraction are among the most commonly used approaches.

Having studied the results of research in this area, we can assert that almost all studies use various combinations of machine learning and artificial intelligence technologies. Although there many architectures and methods of neural networks, convolutional neural networks demonstrate the best results. Of the methods used, the support vector and decision tree approaches can be singled out as delivering the highest accuracy.

In order to work effectively with neural networks, it is necessary to compile a sample correctly. For this purpose, it is first necessary to determine what characteristics and initial data are required to classify users. Since results are predicted on the basis of statistical data, knowledge of statistics and their application is also necessary both for sampling and correctly analyzing the results.

## CONCLUSIONS

This review set out to familiarize a wide range of readers with contemporary models and methods for analyzing complex social network structures, as well as the tools used for this purpose, which include those based on ready-made solutions in the form of services and software, as well as research applications developed using the Python programming language. When setting up and conducting further investigations by a wide range of researchers, it is very important to consider the advantages and disadvantages of existing models and methods. It can be concluded that SNA methods may serve as a very useful tool for creating a complete picture of public mood.

The review describes the specifics of the main tasks solved in the study of complex network structures. The interdisciplinary discipline of system analysis includes the theory of complex networks, text analysis and computational linguistics, neural networks, and many related areas. In particular, the work explores dynamic models of processes observed in complex social network systems, as well as the structural characteristics of such networks and their relationship with observed dynamic processes, including using dynamic graph construction theory. The use of neural networks to predict the evolution of dynamic processes observed in complex social systems and their structure (for example, how activity, the number of users and the structure of their connections in social network communities change) is also analyzed. When creating models to describe the observed processes, considerable attention is focused on the use of computational linguistics methods to extract knowledge from text messages of social network users.

By monitoring the opinions of users on important economic and social issues, both at the level of individual territorial entities (for example, districts, settlements of small towns, etc.), and at the regional level, such studies can be used to help predict not only the dynamics of social processes (changes in social sentiment), but also socioeconomic development trends.

## ACKNOWLEDGMENTS

This research was supported by the Russian Science Foundation, grant No. 22-21-00109 "Development of the dynamics forecasting models of social moods based on the analysis of text content time series of social networks using the Fokker–Planck and nonlinear diffusion equations."

### Authors' contributions

**J.P. Perova**—collection and analysis of information for the review.

**V.R. Grigoriev**—processing of materials for the review.

**D.O. Zhukov**—conceptual idea and discussion of the obtained results.

## REFERENCES

1. Gubanov D.A., Novikov D.A., Chkhartishvili A.G. *Sotsial'nye seti: modeli informatsionnogo vliyaniya, upravleniya i protivoborstva* (*Social networks: models of informational influence, management and confrontation*). Moscow: MTsNMO; 2018. 223 p. ISBN 978-5-4439-1302-5 (in Russ.).
2. Batura T.V. Methods of social networks analysis. *Vestnik NGU. Seriya: Informatsionnye tekhnologii = Vestnik NSU. Series: Information Technologies*. 2012;10(4):13–28 (in Russ.). Available from URL: https://lib.nsu.ru/xmlui/handle/nsu/250
3. Pasa L., Navarin N., Sperdut A. SOM-based aggregation for graph convolutional neural networks. *Neural Comput. & Applic.* 2022;34(1):5–24. https://doi.org/10.1007/s00521-020-05484-4
4. Zhukov D.O., Akimov D.A., Red'kin O.K., Los' V.P. Application of convolutional neural networks for preventing information leakage in open internet resources. *Aut. Control Sci*. 2017;51(8):888–893. https://doi.org/10.3103/S0146411617080314
5. Zhang Z., Wu S., Jiang D., Chen G. BERT-JAM: Maximizing the utilization of BERT for neural machine translation. *Neurocomputing*. 2021;460:84–94. https://doi.org/10.1016/j.neucom.2021.07.002
6. McKinney W. *Python i analiz dannykh* (*Python and Data Analysis*): transl. from Eng. Moscow: DMK Press; 2020. 540 p. (in Russ.). ISBN 978-5-94074-590-5
[McKinney W. *Python for Data Analysis*: 2nd ed. US: O'Reilly Media, Inc.; 2017. 541 p. ISBN 978-1-491-95766-0. Available from URL: https://www.programmer-books.com/wp-content/uploads/2019/04/Python-for-Data-Analysis-2nd-Edition.pdf]
7. Nikolenko S., Kadurin A., Arkhangel'skaya E. *Glubokoe obuchenie. Pogruzhenie v mir neironnykh setei* (*Deep Learning. Immersion in the World of Neural Networks*). St. Petersburg: Piter; 2021. 476 p. (in Russ.). ISBN 978-5-4461-1537-2
8. Kan K. *Neironnye seti. Evolyutsiya* (*Neural Networks. Evolution*). LitRes; 2018. 380 p. (in Russ.).
9. Rashid T. *Sozdaem neironnuyu set'* (*Make Your Own Neural Network*): transl. from Eng. St. Petersburg: Al'fa-kniga; 2017. 272 p. (in Russ.). ISBN 978-5-9909445-7-2
[Rashid T. *Make Your Own Neural Network*. 1st ed. CreateSpace Independent Publishing Platform; 2016. 222 p. ISBN-13 ‎978-1530826605]
10. Galushkin A.I. *Neironnye seti: osnovy teorii* (*Neural Networks: Fundamentals of Theory*). Moscow: Goryachaya liniya-Telekom; 2012. 496 p. (in Russ.).
11. Savel'ev A.V. The philosophy of methodology of neuro-modeling: Sense and prospects. *Filosofiya nauki = Philosophy of Sciences*. 2003;1(16):46–59 (in Russ.).
12. Alekseev A.Yu., Kuznetsov V.G., Petrunin Yu.Yu., Savel'ev A.V., Yankovskaya E.A. Neurophilosophy as a conceptual basis for neurocomputing. *Neirokomp'yutery: razrabotka, primenenie = Neurocomputers: Development, Application*. 2015;5:69–77 (in Russ).
13. Sekara V., Stopczynski A., Lehmann S. Fundamental structures of dynamic social networks. *Proc. Natl Acad. Sci. USA*. 2016;113(36):9977–9982. https://doi.org/10.1073/pnas.1602803113
14. Ubaldi E., Vezzani A., Karsai M., Perra N., Burioni R. Burstiness and tie activation strategies in time-varying social networks. *Sci. Rep*. 2017;7:46225. https://doi.org/10.1038/srep46225
15. Palomares I., Porcel C., Pizzato L., Guy I., Herrera-Viedma E. Reciprocal recommender systems: analysis of state-of-art literature, challenges and opportunities towards social recommendation. *Information Fusion*. 2021;69(16):103–127. https://doi.org/10.1016/j.inffus.2020.12.001
16. Yatim Md.A.F., Wardhana Y., Kamal A., Soroinda A.A.R., Rachim F., Wonggo M.I. A corpus-based lexicon building in Indonesian political context through Indonesian online news media. In: *2016 International Conference on Advanced Computer Science and Information Systems* (*ICACSIS*). IEEE. https://doi.org/10.1109/ICACSIS.2016.7872794

17. Kirn S.L., Hinders M.K. Dynamic wavelet fingerprint for differentiation of tweet storm types. *Soc. Netw. Anal. Min*. 2020;10(1):4. https://doi.org/10.1007/s13278-019-0617-3
18. Karami A., Elkouri A. Political Popularity Analysis in Social Media. In: Taylor N., Christian-Lamb C., Martin M., Nardi B. (Eds.). *Information in Contemporary Society*. Part of: *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*). 2019. V. 11420. P. 456–465. https://doi.org/10.1007/978-3-030-15742-5_44
19. Belcastro L., Cantini R., Marozzo F., Talia D., Trunfi P. Learning political polarization on social media using neural networks. *IEEE Access*. 2020;8:47177–47187. https://doi.org/10.1109/ACCESS.2020.2978950
20. Vijayaraghavan P., Vosoughi S., Roy D. Twitter demographic classification using deep multi-modal multi-task learning. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. 2017;2(Short Papers):478–483. https://doi.org/10.18653/v1/P17-2076
21. Preoţiuc-Pietro D., Liu Y., Hopkins D., Ungar L. Beyond binary labels: political ideology prediction of Twitter users. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. 2017;1(Long Papers):729–740. https://doi.org/10.18653/v1/P17-1068
22. Hinds J., Joinson A.N. What demographic attributes do our digital footprints reveal? A systematic review. *PLoS One*. 2018;13(11):e0207112. https://doi.org/10.1371/journal.pone.0207112
23. García D. Leaking privacy and shadow profiles in online social networks. *Sci. Adv*. 2017;3(8):e1701172. https://doi.org/10.1126/sciadv.1701172
24. Pandya A., Oussalah M., Monachesi P., Kostakos P. On the use of distributed semantics of tweet metadata for user age prediction. *Future Generation Computer Systems*. 2020;102(5915):437–452. https://doi.org/10.1016/j.future.2019.08.018
25. Pulipati S., Somula R., Parvathala B.R. Nature inspired link prediction and community detection algorithms for social networks: a survey. *Int. J. Syst. Assur. Eng. Manag*. 2021. https://doi.org/10.1007/s13198-021-01125-8
26. Li H., Mao X., Wu C., Yang F. Design and analysis of a general data evaluation system based on social networks. *EURASIP J. Wireless Com. Network*. 2018;1:109. https://doi.org/10.1186/s13638-018-1095-4
27. Xu F., Sun D., Li Z., Li B. Research on online supporting community of extreme organization by AI-SNA based method. In: *Proceedings of the 8th IEEE International Conference on Software Engineering and Service Sciences* (*ICSESS*). 2018. V. 2017. P. 546–551. https://doi.org/10.1109/ICSESS.2017.8342974
28. Volkova S., Bachrach Y., Van Durme B. Mining user interests to predict perceived psycho-demographic traits on Twitter. In: *2016 IEEE Second International Conference on Big Data Computing Service and Applications* (*BigDataService*). IEEE. 2016. P. 36–43. https://doi.org/10.1109/BigDataService.2016.28
29. Culotta A., Ravi N.K., Cutler J. Predicting Twitter user demographics using distant supervision from website traffic data. *J. Artif. Intell. Res*. 2016;55:389–408. https://doi.org/10.1613/jair.4935
30. Barberá P. Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data. *Working Paper*. Available from URL: http://pablobarbera.com/static/less-is-more.pdf
31. Ardehaly E.M., Culotta A. Learning from noisy label proportions for classifying online social data. *Soc. Netw. Anal. Min.* 2018;8:2. https://doi.org/10.1007/s13278-017-0478-6
32. Franco-Riquelme J.N., Bello-Garcia A., Ordieres-Meré J. Indicator proposal for measuring regional political support for the electoral process on Twitter: The case of Spain's 2015 and 2016 general elections. *IEEE Access*. 2019;7:62545–62560. https://doi.org/10.1109/ACCESS.2019.2917398
33. Jungherr A., Schoen H., Posegga O., Jürgens P. Digital trace data in the study of public opinion: an indicator of attention toward politics rather than political support. *Soc. Sci. Comput. Rev*. 2016;35(3):336–356. https://doi.org/10.1177/0894439316631043
34. Mwanza S., Suleman H. Measuring network structure metrics as a proxy for socio-political activity in social media. In: *IEEE International Conference on Data Mining Workshops* (*ICDMW*). IEEE. 2017. P. 878–883. https://doi.org/10.1109/ICDMW.2017.120
35. Al-Agha I., Abu-Dahrooj O. Multi-level analysis of political sentiments using Twitter data: A case study of the Palestinian-Israeli conflict. *Jordanian Journal of Computers and Information Technology*. 2019;5(3):195–215. https://doi.org/10.5455/jjcit.71-1562700251
36. Basil M., Gaikwad S., Salim A.S. Deep learning approach based dominant age group based classification for social network. In: Khalaf M., Al-Jumeily D., Lisitsa A. (Eds.). *Applied Computing to Support Industry*: *Innovation and Technology. ACRIT 2019*. *Communications in Computer and Information Science.* 2020;1174:148–156. https://doi.org/10.1007/978-3-030-38752-5_12
37. Guimaraes R., Renata R., De Gaetano D., Rodriguez D.Z., Bressan G. Age groups classification in social network using deep learning. *IEEE Access*. 2017;5:10805–10816. https://doi.org/10.1109/ACCESS.2017.2706674
38. Bhat S.F., Lone A.W., Dar T.A. Gender prediction from images using deep learning techniques. *In: 2019 International Artificial Intelligence and Data Processing Symposium* (*IDAP*). IEEE. 2019. https://doi.org/10.1109/IDAP.2019.8875934

39. Bulut İ., Erdoğan M., Gönülal B., Baş R., Kılıç Ö. Using short texts and emojis to predict the gender of a texter in Turkish. In: *2019 4th International Conference on Computer Science and Engineering* (*UBMK*). IEEE. 2019. P. 435–438. https://doi.org/10.1109/UBMK.2019.8907198
40. Dileep M.R., Danti A. Multiple hierarchical decision on neural network to predict human age and gender. In: *2016 International Conference on Emerging Trends in Engineering*, *Technology and Science* (*ICETETS*). IEEE. 2016. https://doi.org/10.1109/ICETETS.2016.7603026
41. Gupta R., Kumar S., Yadav P., Shrivastava S. Identification of age, gender, & race SMT (scare, marks, tattoos) from unconstrained facial images using statistical techniques. In: *2018 International Conference on Smart Computing and Electronic Enterprise* (*ICSCEE*). IEEE. 2018. https://doi.org/10.1109/ICSCEE.2018.8538423
42. Khdr J., Varol C. Age and gender identification by SMS text messages. In: *2018 International Conference on Artificial Intelligence and Data Processing* (*IDAP*). IEEE. 2018. https://doi.org/10.1109/IDAP.2018.8620780
43. Koti P., Pothula S., Dhavachelvan P. Age forecasting analysis – over microblogs. In: *2017 Second International Conference on Recent Trends and Challenges in Computational Models* (*ICRTCCM*). IEEE. 2017. P. 83–86. https://doi.org/10.1109/ICRTCCM.2017.38
44. López-Santamaría L.-M., Almanza-Ojeda D.-L., Gomez J.C., Ibarra-Manzano M. Age and gender identification in unbalanced social media. In: *2019 International Conference on Electronics, Communications and Computers* (*CONIELECOMP*). IEEE. 2019. https://doi.org/10.1109/CONIELECOMP.2019.8673125
45. Luo F., Cao G., Mulligan K., Li X. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*. 2015;70(3):11–25. https://doi.org/10.1016/j.apgeog.2016.03.001
46. Sánchez-Hevia H.A., Gil-Pita R., Utrilla-Manso M., Rosa-Zurera M. Convolutional-recurrent neural network for age and gender prediction from speech. In: *2019 Signal Processing Symposium* (*SPSympo*). IEEE. 2019. P. 242–245. https://doi.org/10.1109/SPS.2019.8881961
47. Wang Y., Song W., Liu L. Age prediction based on feature selection. In: *2017 2nd IEEE International Conference on Computational Intelligence and Applications* (*ICCIA*). IEEE. 2017. P. 359–363. https://doi.org/10.1109/CIAPP.2017.8167239
48. Pandya A., Oussalah M., Monachesi P., Kostakos P., Lovén L. On the use of URLs and hashtags in age prediction of Twitter users. In: *2018 IEEE International Conference on Information Reuse and Integration* (*IRI*). IEEE. 2018. P. 62–69. https://doi.org/10.1109/IRI.2018.00017
49. Zhukov D.O., Zaltcman A.D., Khvatova T.Yu. Forecasting changes in states in social networks and sentiment security using the principles of percolation theory and stochastic dynamics. In: *Proceedings of the 2019 IEEE International Conference* “*Quality Management, Transport and Information Security, Information Technologies”* (*IT&QM&IS*). IEEE. 2019. Article number 8928295. P. 149–153. https://doi.org/10.1109/ITQMIS.2019.8928295
50. Mukhamediev R.I., Yakunin K., Mussabayev R., Buldybayev T., Kuchin Y., Murzakhmetov S., Yelis M. Classification of negative information on socially significant topics in mass media. *Symmetry*. 2020;12(12):1945. https://doi.org/10.3390/sym12121945
51. Ko H., Jong Y., Sangheon K., Libor M. Human-machine interaction: A case study on fake news detection using a backtracking based on a cognitive system. *Cogn. Syst. Res*. 2019;55:77–81. https://doi.org/10.1016/j.cogsys.2018.12.018
52. Willaert T., Van Eecke P., Beuls K., Steels L. Building social media observatories for monitoring online opinion dynamics. *Soc. Media Soc*. 2020;6(2):205630511989877.
53. Tran C., Shin W.-Y., Spitz A. Community detection in partially observable social networks. *ACM Transactions on Knowledge Discovery from Data.* 2022;16(2):1–24. https://doi.org/10.1145/3461339
54. Chen Z., Li L., Bruna J. Supervised community detection with line graph neural networks. In *Proceedings of the 7th International Conference on Learning Representations* (*ICLR*). ACM. 2019. https://doi.org/10.48550/arXiv.1705.08415
55. Hoffmann T., Peel L., Lambiotte R., Jones N.S. Community detection in networks without observing edges. *Sci. Adv*. 2020;6(4):eaav1478. https://doi.org/10.1126/sciadv.aav1478
56. Bashuev Ya., Grigorjev V. Social nets deanonimization methods. *Vestnik RGGU. Seriya Dokumentovedenie i arkhivovedenie. Informatika. Zashchita informatsii i informatsionnaya bezopasnost’ = RGGU BULLETIN. Series: Records Management and Archival Studies. Computer Science. Data Protection and Information Security*. 2016;3(5):125–146 (in Russ.). Available from URL: https://www.rsuh.ru/upload/main/vestnik/pmorv/Vestnik_daizi3(5)-16.pdf#page=125
57. Wondracek G., Holz T., Kirda E., Kruegel C. A practical attack to de-anonymize social network users. *Technical Report TR-iSecLab-0110-001*. 2013. Available from URL: https://anonymous-proxy-servers.net/paper/sonda-tr.pdf
58. Simon B., Gulyás G., Imre S. Analysis of grasshopper, a novel social network de-anonymization algotithm. *Periodica Polytechnica: Electrical Engineering and Computer Science.* 2014;58(4):161–173. https://doi.org/10.3311/PPee.7878

59. Peng W., Li F., Zou X., Wu J. A two-stage deanonymization attack against anonymized social networks. *IEEE Trans. Comp*. 2014;63(2):290–303. https://doi.org/10.1109/TC.2012.202
60. Khvatova T., Zaltsman A., Zhukov D. Information processes in social networks: Percolation and stochastic dynamics. In: *CEUR Workshop. Proceedings 2nd International Scientific Conference "Convergent Cognitive Information Technologies*." 2017;1–2064:277–288.
61. Zhukov D., Khvatova T., Zaltsman A. Stochastic dynamics of influence expansion in social networks and managing users' transitions from one state to another. In: *Proceedings of the 11th European Conference on Information Systems Management* (*ECISM*). 2017. P. 322–329. Available from URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-85039839600&partnerID=MN8TOARS

## About the authors

**Julia P. Perova,** Senior Lecturer, Department of Telecommunications, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: perova_yu@mirea.ru. Scopus Author ID 57431908700, https://orcid.org/0000-0003-4028-2842

**Vitaly P. Grigoriev,** Cand. Sci. (Eng.), Associate Professor, Head of the Department of Information Warfare, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: grigorev@mirea.ru. RSCI SPIN-code 4088-0403

**Dmitry O. Zhukov,** Dr. Sci. (Eng.), Professor, Department of Information Warfare, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: zhukov_do@mirea.ru. Scopus Author ID 57189660218, RSCI SPIN-code 1798-8891

## Об авторах

**Перова Юлия Петровна,** старший преподаватель кафедры телекоммуникаций Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: perova_yu@mirea.ru. Scopus Author ID 57431908700, https://orcid.org/0000-0003-4028-2842

**Григорьев Виталий Робертович,** к.т.н., доцент, заведующий кафедрой «Информационное противоборство» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: grigorev@mirea.ru. SPIN-код РИНЦ 4088-0403.

**Жуков Дмитрий Олегович,** д.т.н., профессор, профессор кафедры «Информационное противоборство» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: zhukov_do@mirea.ru. Scopus Author ID 57189660218, SPIN-код РИНЦ 1798-8891.

*Translated from Russian into English by Evgenii I. Shklovskii*
*Edited for English language and spelling by Thomas A. Beavitt*