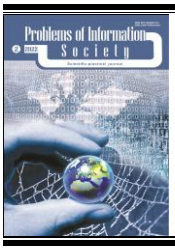

www.jpis.az

13 (2)
2022

Analysis of demographic characteristics based on data of social network users

Rana T. Gasimova[1], Rahim N. Abbasli[2]

[1]Azerbaijan National Academy of Sciences, Institute of Information Technology, B.Vahabzade str., 9A, AZ1141, Baku, Azerbaijan

[2]Driven, Montreal, Kvebek, 130 Adelaide St W, Suite 3100, M5H 3P5, Toronto, Ontario, Canada

[1]renakasumova@gmail.com, [2]rahim.abbasli@gmail.com

orcid.org/0000-0002-0480-9910[1], orcid.org/0000-0001-5965-447X[2]

ARTICLE INFO

http://doi.org/10.25045/jpis.v13.i2.09

Article history:
Received 10 January 2022
Received in revised form
15 Mart 2022
Accepted 2 June 2022

Keywords:
Demography
E-demography
Ddemographic characteristics
Ssocial network analysis
Big data
Statistical methods

ABSTRACT

The main goal of the state policy in the field of demography is to ensure the growth of population reproduction in accordance with the country's development strategy by eliminating negative trends in demographic processes. Demographic processes can be assessed by country, region and district. In this regard, demographic surveys can be conducted at the state, regional and individual levels. The implementation of an effective demographic policy in the country is an integral part of the e-government system. The article is devoted to the analysis of demographic characteristics based on the data of social network users. The spread of the Internet and digital technologies has created new opportunities for demographic research. To this end, the article analyzes demography as a field of multidisciplinary research and shows the importance of data collected in social networks for demographic research. This includes the use of data collected in the analytical systems of social networking services as a new source of information for demographic research. The article discusses foreign experience and current scientific and practical studies in the field of electronic demography, identifies current areas of research and analyzes their state-of-the-art. The paper explores the social network analysis systems and their classification by characteristics.

1. Introduction

In the modern world, the Internet has already penetrated everyone's daily life. It is increasingly penetrating the spheres of activity of the state, society and people in particular. Using the Internet is gaining new opportunities for development of all areas of human activity including science, education, politics, business, production, services, etc. Probably this is the reason why more than five billion people in the world are the Internet users [1, 2].

The International Data Corporation (IDC) statistics estimate the volume of digital information in the world to be 175 zetabytes by 2025 (Figure 1). Moreover, the number of devices connected to the Internet in 2020 is estimated to be 50 billion [3].

The report also highlights other trends in the Internet, i.e., the importance of increasing information for the life of society, the migration of large amounts of information to clouds, the growth of new intelligent devices generating large amounts of data without human intervention, and etc.

The global and broad proliferation of digital technologies, i.e., the Internet, social media, mobile phones, has created new opportunities for demographic research. Currently, a large part of Internet traffic refers to social networking sites. Simultaneously, the use of social media, the Internet and smartphones affect the people's daily activities, lives and demographic behavior.

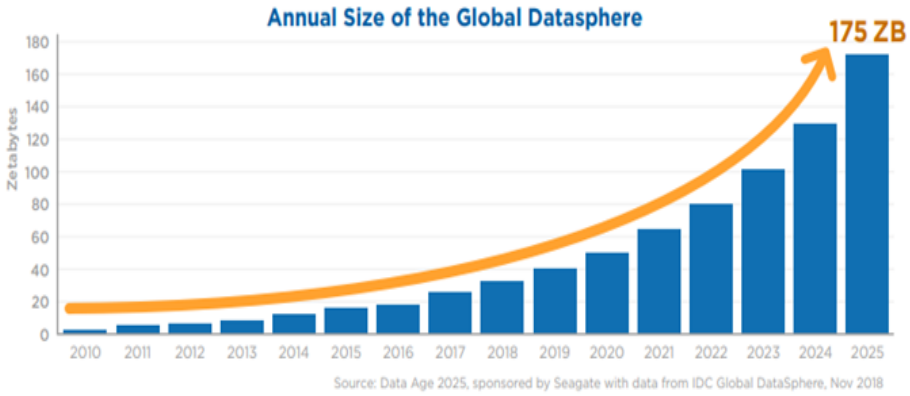

Fig. 1. Statistics on the growth of digital information in the world (https://www.idc.com)

The development of any state directly depends on the sustainability of demographic development. At the global and national levels, sustainable human development, along with economic, social and demographic development, forms a complex system of interactions. Prospects for demographic development are prepared in several variants based on a number of assumptions about the future dynamics of birth, death and population migration characteristics. The most realistic version of the expected reproduction from the developed perspectives is considered to be a forecast. Forecasting the demographic processes and population development is an important factor in the formation and implementation of effective socio-economic policy of the country, and ensuring sustainable development. Information on population, including the population's reproduction regime, growth rate, age-gender composition, population distribution throughout the country, and other additional characteristics and features are taken into account to forecast socio-economic development [5].

Due to the emergence of online social networking services, SixDegrees, LiveJournal, Facebook, Google, Twitter, MySpace, YouTube, etc. in the 1990s, the social network data analysis became popular all around the world. Determination of the demographic characteristics of countries using the data collected in the analytical systems of these projects became possible [6].

The socialization of personal information in the virtual environment is also related to this. Thus, biographical facts, correspondence, diaries, photos, videos, audio materials, travel notes and other facts become public through the virtual space. This has opened up unprecedented opportunities to address previously unresolved research and business problems due to a lack of information, as well as to create support services and applications for social network users. Moreover, companies and research centers are focusing on collecting and analyzing social media data more.

As the population grows, so does the volume of demographic data, and this large amount of data is a major source for demographic research. Today, social network data is also attracting special attention as a new data source for demographic research. The widespread use of the Internet and digital technologies, especially social networking platforms, results in the fact that the people are leaving more and more traces in the virtual environment [7, 8].

Consequently, the impact of the Internet, social networking services and smartphones on human life is considered to be a new unique data source for the study of demographic behavior. Social networking services affect not only the landscape of the Internet, they can also significantly change the users' behavior.

The development of these services in terms of content, services, accessibility, and the audience and territory covered has led to the emergence of new methods and tools to affect the society. They have created new opportunities for the interaction of people, organizations and state [9].

Being the core of e-government, e-demography allows the study of the impact of digital technologies on demographic behavior and the use of new data sources of for more in-depth study of demographic processes. E-demography is of potential importance for all areas of population research, from social demography to population geography. Therefore, digital demography is capable to solve some traditional demography problems using digital data sources and mining methods collected in social networking services on the Internet [4, 6, 7, 10, 11].

Given these important facts, this paper focuses on the use of social network data for demographic study. In this regard, the goals and objectives of demography, research methods, the relationship with other sciences, as well as the structural features of social networks, the acquisition and use of data from social networks for demographic study are studied.

2. Available studies in the field of demography

Demography as a social science has always been in the focus of politicians, sociologists and scientists. The object of demographic science is people and population, and the goal is the laws of reproduction of the population. The main field of demography includes the processes of natural movement of population, including birth and death processes, marriages and divorces, family and other demographic structures of population, identification of trends in the number and composition of population, the relationship of demographic processes and structures, and the identification and design of demographic forecasts. Moreover, the main task of demographic research is to determine the laws of natural movement of population [7, 12, 13]. The three main practical tasks of demography are as follows [12-15]:

- To study the trends and factors in demographic processes based on data collection and demographic analysis
- To develop the demographic forecasts
- To develop the measures and methods for collecting demographic statistics

The study of scientific and practical literature in the field of demography shows that a number of important studies have been conducted in recent years to identify the demographic characteristics of population and to explore new sources of information [4, 12, 16-20].

Various approaches to the use of digital data in the study of demographic events are available. For example, some studies report that the use of Internet data to study fertility in some countries is still in its infancy. Research in this area examines demographic behaviors by analyzing search engine queries and monitors them [4, 5, 15].

[7, 21] also highlight that combining traditional data sources with the Internet search capabilities can increase the predictive performance of demographic models. Furthermore, it is shown that when the nature of the relationship between searches, news, and behavior changes, it can, as a rule, lead to erroneous estimates.

Note that ongoing studies focus on the importance of statistical and mathematical methods in demography [4, 5, 12, 14]. Most of the information about the population is collected and checked with these models, errors are identified, standardization is carried out, and so on.

[4, 5] also include search queries in web browsers, social media data, data collected in state registers, e-service data, feedback on government-citizen relations, and so on. are shown as the main sources of demographic research. The analysis of large volumes of demographic data collected in these sources allows for the acquisition of new knowledge related to demographic behavior.

Demographic research based on social networking data is currently evolving significantly due to the proliferation and availability of various large-scale online social networking websites. Researchers analyze the dynamics of the population by analyzing the digital footprints left by Internet users on social networks. In addition, it was possible to learn the details of health and death through search queries, social network data or other archived digital data [7, 9]. This shows that social network data creates a great research opportunity for demographic research.

As part of statistics, it is no coincidence that demographic statistics appeared first. Most of the demographic data is acquired from population statistics and medical statistics. There are many social media outlets available in the medical segment today. This includes social networking platforms, blogs, microblogs, wikis, data exchange sites, virtual reality, and more. These tools serve to improve professional networks, develop public health programs, and educate patients [22].

Recent studies have looked at the use of data collected in population registers as a new source of information for demographic surveys and statistics. National statistical offices have implemented a number of initiatives for the transfer of demographic data from population registers to conduct demographic statistics [4, 17, 23].

3. Data sources in demographic research and general characteristics

Sources of population data include sources of demographic data. [12, 24, 25] distinguish demographic data according to the following features:

- Coverage (population of the world, region, country or its part);
- Scale or nationality (national and international);
- Characteristics of data (number and composition of the population during the period of demographic events, mixed features of demographic processes);

- Time of reference (historical-demographic, retrospective, current, demographic projections);
- Method of obtaining information (official publications, special research materials);
- Nature of publication (special, general type) and etc.

Due to their completeness and reliability, the sources of population data are classified as primary and secondary. The main sources of demographic data include the direct results of the processing and comparison of the collected statistical data and the calculation indicators obtained on their basis. Secondary demographic data sources are the result of author's calculations and assessments based on primary sources of demographic data. Thus, primary demographic data is raw data, and secondary data is demographic data obtained from primary demographic data [12, 26].

The most accurate and reliable demographic data is provided by special state statistical agencies of the countries (in our country it is the State Statistical Committee). In terms of varying degrees of completeness and reliability, demographic data is also provided from other sources: documents, legislation, as well as information services in the field, etc. The main sources of information on population and demographic processes are population censuses, socio-economic data, as well as data obtained from current official statistical reports. The United Nations Demographic Yearbook is one of the most influential international demographic publications [12, 15].

All types of demographic work are based on primary demographic statistics. Processing and analysis of primary demographic statistics provides secondary demographic information such as, tables, publications, monographs, articles, websites, etc. In order to carry out various types of high-quality work and, first of all, work in the demographic field, it is necessary to have complete, detailed, multifaceted and reliable information. Population data should be systematically gathered, updated, summarized and analyzed.

The main types of primary demographic data are as follows [12]:

- Population census;
- Current accounting of demographic events;
- Special sample surveys;
- Population lists and registers (administrative records).

The current population estimation is based on the results of the last census, which annually adds the number of births and arrivals to the region and subtracts the number of deaths and departures. Based on the results of the current census, the population is calculated and clarified for the period after the prior census. According to the results of the census, information tables are compiled on the number, age, gender and ethnic composition, literacy and education level, marital status, households, as well as the grouping of urban and rural settlements by population.

Many countries perform a state population register. The practice of using them has extended with the application of information technology and software development. In the countries where the population is registered, each person is given a personal identification number (PIN) at birth or when entering the country. All information about them is updated here throughout their life, entered into a centralized database.

In the absence of a unified identification system, it is very difficult to coordinate and integrate various registers. The current technologies allow to connect all existing state registers in a short period of time using only PIN. When the work on the creation and maintenance of the population register is well organized, there is no need for a population census. Thus, in some countries (e.g., Finland, Denmark, the Netherlands, Sweden, etc.) the census is not traditionally conducted, all the necessary data is acquired through registers [12, 13].

Digitization helps improve census and registration, as well as access to existing data such as bibliometric databases. Consequently, new sources are emerging, such as online databases that can be used to study demographic processes. Thus, the digitization of traditional demographic data makes personal data about the population in censuses and other sources accessible to other researchers [7]. Digital demography also studies demographic processes by analyzing digital footprints on social networks. This allows the use of new data sources, such as data from social network users, for a better understanding and in-depth study of social and demographic processes taking place in the virtual world.

4. Data analysis, methods and applications in social networks

Population, births and deaths, migration, age and gender composition, national and ethnic composition, geographical location of the population, as well as the socio-economic situation affect the demographic processes, and these indicators need to be taken into

account in demographic research. Therefore, the data collected in the analytical systems of social networking services are used as a new source of information for demographic research and statistics.

With the advent of social networking web-sites and services, the analysis of social network data is spreading rapidly around the world. Nowadays, there are several online social networking surveys. The most important of them are scientific works that have an applied field and represent the trends in this area. Social networks are a unique source of information about people's personal lives and interests and have a strong impact on their demographic behavior. In the study of people's demographic behavior, the data of social network users are considered to be new unique sources of data [4, 6].

Apparently, social networks available on the Internet vary according to the purpose, location and age of users. Simultaneously, the essence of social networks dictates the existence of elements, such as user pages (profiles) and links (at least among users and often among other objects acting as a network hub) found in all networks. Digital footprints that people keep while using social networks create new opportunities for demographic research as new sources of information. This information becomes public with the consent of each individual. Here, user questionnaires (user profiles) on social networks are considered.

As already mentioned, the information in the questionnaire is filled in independently by the users, voluntarily, and is placed publicly with their consent. Thus, from a judicial point of view, an access to this information does not differ from the access to other information on the Internet. Obtaining and processing this information requires the use of certain technical tools, but it is also more accessible than the data on mobile networks.

The available data about users varies depending on the social networks. So that, if we consider the services of blogs, microblogs or photo blogs as social networks (for example, livejournal.com, twitter.com and instagram.com), they may have a fake name in the profile information about the user. However, compared to them, in other social networks such as vk.com, facebook.com and etc. user profiles contain more information.

Users are asked to enter their names, dates of birth and gender, as well as information about residence, education and workplace. Filling in all these fields is optional. However, considering that the main purpose of a social network as a service is to facilitate communication and networking, it is suitable for users to provide enough detailed information about themselves to make it easier for classmates, fellow students and colleagues to find them.

Furthermore, the profile pages of social network user pages contain personal information, as well as information about "friends", i.e., relationships with other users of the social network and communities (interest groups) of which the user is a member. Although all this information is incomplete, it can be quite regulated.

Administrators of social networks try not to differentiate the organizations mentioned in the questionnaires (for example, cities, schools, universities) on the profiles of different users. Nevertheless, this is not always successful. In addition to customized profile data, user profiles also contain unregulated user data, i.e., photos, text notes, audio and video files, and other documents. Unregulated data is also typical of the above-mentioned blog services, except that this information is practically not included in the questionnaire data. Analysis of unregulated data is also possible (at least theoretically), but it is much more difficult than simple processing of questionnaire data [27].

When registering on social networks, based on the questionnaire data, the followings are included:

- Last name, first name, patronymic;
- Gender;
- Date of birth;
- Country of residence;
- Place of birth;
- Marital status;
- Children;
- Education (school, university, etc.);
- Employment (workplace or unemployment);
- Sources of income;
- Hobbies;
- Language skills, etc.

Along with the above personal information, a user's social network profile contains other data that significantly complements the information about the user's migration and can be used to track their movements within the country (for example, the ability to list places where the user rests, etc.). Data from social networks can be considered an alternative to statistics from official sources. Table 1 compares the data from the surveys used in the census during sociological and demographic research with the questionnaire data from the social networks.

Table 1. Comparison of the survey data with the questionnaire data on social networks

Fields	In the census survey	On social networks
List of people living in the house	Yes	No
Relationship between the landlord and the people living in the house	Yes	An option to mention relatives
Gender	Yes	Yes
Date of birth	Yes	Yes
Birthplace	Yes	Yes
Marital status	Yes	Yes, but the answer options differ from the survey
Citizenship	Yes	No
Racial affiliation (citizenship)	Yes	No
Education	Yes	More details available
Language skills	Yes	Yes
Sources of income	Yes	No
Employment and unemployment	Yes	More details available
Migration	Yes	More details available
Number of children	Yes	Yes

The advantages of social network data are:

- Availability of data is not provided by statistical services;
- Details, data is not initially aggregated and can be collected in any way required;
- Retrospective, user questionnaires often contain not only information about the current status of the user, but also the biographical information about the user, which allows to obtain information about the past.

The experience of using data from social networks in research indicates that if the vast majority of those studied are presented on social media, it is possible to realize these advantages and avoid the lack of representativeness. Therefore, for the widespread use of data on social networks in demographic and sociological research, it is necessary to address the issue of developing and algorithmicizing criteria ensuring the reliability of filling in certain fields in the questionnaire. For example, advertising banners sent to a user's personal profile on Facebook are selected based on the user's age, income, interests, gender, employment and leisure time. These parameters are significant for in-depth demographic analysis. Increasing availability of geographic information on Internet resources has been of great benefit to migration research [4].

When creating profiles on social networks, users often intentionally or unintentionally do not enter data in certain areas, or misrepresent their biographies, interests and themselves. Moreover, on some social networks (Twitter, YouTube), the user profile is limited to a set of basic attributes that are not sufficient to address many issues associated with the individualization of results. In Internet marketing and recommendation systems, it is especially important to identify the demographic attributes of the user to purposefully promote goods and services in user groups with similar characteristics. In addition to Internet services, such demographic features are used in various areas, such as sociology, psychology, criminology, economics, medicine, personnel management, etc. [28].

As mentioned, the main place in demography as a source of information occupies giant Internet projects such as is Meta (Facebook), Google, Twitter, Yandex, YouTube, LiveJournal, etc. Demographic indicators can be conditionally divided into categories (gender, nationality, race, marital status, level of education, occupation, employment, religious and political views, etc.) and numbers (age, income level). The section is conditional because the values of a numeric attribute can be adjusted to categories, and in the future this attribute can be considered categorical. Particularly, age indicators can be divided into several age categories, and this division is often applied in practice [7, 5, 29, 30].

It is possible to determine the demographic characteristics of the country using the data collected on Twitter using the method proposed in [28, 31, 32]. The applied method has the following features:

- Creating a primary database;
- Preprocessing the text;
- Describing the signs;
- Selecting informative features;
- Teaching;
- Classifying.

With the exception of the first one, all steps are performed separately for each attribute. Messages

from social network users are an essential part of text content on the Internet. Furthermore, in many cases, social networks can play the role of an unofficial mass media that can broadcast any user's messages about current events. Altogether, the automatic receipt of a number of messages about any event becomes difficult due to the following factors [24, 25]:

- Large amount of input data;
- Large number of worthless messages;
- Users can describe the same event differently;
- Various events may overlap in terms of time;
- Difficulty in distinguishing an event and its sub-events (for example, international Olympiads on subject and a specific subject within this Olympiad).

The identification of the users on various social networks allows to get a full picture of their social behavior on the Internet. Finding accounts belonging to the same person on numerous social networks allows us to get a complete social graph. This can be beneficial in resolving issues such as information retrieval, online advertising, recommendation systems, etc. Since the search for user accounts on various networks generally requires the latest information about all users of these networks, it is recommended to limit the search area on the networks under investigation to the nearest neighbors of a user whose accounts are known.

Thus, the identification of users locally on different social networks involves comparing user accounts within the lists of contacts of any active user on those social networks. This is especially the case while working with user connections in social meta services to connect news feeds in supported social services or provide a unified messaging system. A similar problem emerges while using the function of automatic linking of connection from different sources (social networks, messengers, etc.), which is often found in modern mobile devices [28].

Analysis of the scientific and practical literature on the subject demonstrates that to resolve the problem of identifying different social network users, methods have been developed to search for different variants of virtual identities of the same user on several social networks. The methods applied are based on the characteristics of the profiles of users' virtual identities and the similarity of their relationships with other users [28, 31, 33].

The method suggested in [30] considers many of the shortcomings of existing methods of user identification, so the social connections of both social networks studied are compared with the original contact lists and used in combination with profile attribute data. Despite the attempts to establish a unified system to ensure a single interaction between different social platforms, these systems are not prevalent.

Data processing on social networks also requires the development of appropriate algorithmic and infrastructural solutions to consider their volume. For example, the annual report on the state of the digital world by We Are Social and Hootsuite, Global Digital 2021, demonstrates that there are currently 4.2 billion social media users in the world. This figure has increased by 490 million for the last 12 months. This means an annual increase of more than 13%. As of October 2021, 53.6% of the world's population are reported to use social networks [34, 35].

The availability of large amounts of data collected on social networks related to human behavior and advances in the analysis of big data sets are the main factors in the application of computational methods to study the dynamics of the human population [7]. Additionally, the storage of hundreds of terabytes and exabytes of Big data, real-time analytical processing, acquiring useful information from them has caused serious problems. Most of the existing algorithms that effectively resolve existing problems are not able to process large amounts of data in a timely manner (at the right time). All leading field analysts try to draw attention to the importance of the ability to work with big data to ensure competitiveness. This is another indication of the urgency of the problems in the field of big data, which are defined by the features such as greater volume, velocity and variety. Solving the problems, such as Big data maintenance, management, analytical processing, security, etc. requires novel technologies, management methods and more sophisticated intellectual analysis tools. In this regard, there is a need for new solutions that allow distributed processing and storage of data without significant loss of quality of results [36-38].

There is a necessity to develop mathematical tools for the analysis of large amounts of demographic data. [4, 6, 38] also propose a number of mathematical models for solving these problems.

Web interfaces of social networks are real-time data sources. Web interfaces are designed to view and interact with social networking pages in a web browser, as well as to use user data by specific applications. Scenarios for using social network interfaces do not involve the automatic collection of data by many users to build social graphs. This causes a number of problems [5, 25]:

- Data confidentiality. In general, the user data is limited to only registered users and authorized network members. This requires support for emulating a user session using special accounts;
- Partial data structuring. In many cases, social network software interfaces have limited functionality. This requires supporting the retrieval of static copies of HTML pages through the user web interface, the proper processing of their dynamic parts, including the execution of asynchronous queries, obtaining the necessary information using the appropriate algorithm or template, and building a structured representation appropriate for further processing;
- Access and blocking restrictions. Service owners often set explicit or implicit limits on the number of requests allowed from a single user account or IP address over a period of time to prevent an unauthorized collection of data and limit the load on the social network service infrastructure. This requires support for the dynamic circulation of the number of sent requests, as well as the collection of user accounts and IP addresses used to collect data.
- Volume of data. It implies the necessity for a parallel method of data collection, as well as the acquisition of representative selection methods for social network users.

Despite the numerous available tools and open databases for collecting data from social networks, the problem of creating random social graph models and tools with certain features for their generation is still relevant. They must be applied to a large number of data sets with various characteristics to reliably test data analysis methods on social networks. For example, search methods for user communities in social graphs can produce significantly different results depending on the size of the initial graph, the average peak, clustering coefficient, and other structural features. The collection of real data required for reliable testing is not limited to the time spent loading and processing partially structured large volumes of data. This is also due to the difficulty of managing the processes used to collect data sets with certain features [25, 28].

Katnegi Mellon, Oxford, Stanford, Inria universities, as well as LiveJournal, Meta (Facebook), Google, Twitter, YouTube, Yahoo!, LinkedIn and other companies are actively involved in the study of social network data. Companies owning online social networking services are actively investing in the development of improved infrastructure and algorithmic solutions for processing large amounts of user data. New commercial companies are emerging and thriving, providing access to social media databases, collecting data on social networks based on given scenarios, social analytics, as well as expanding existing platforms using social network data [28].

Experts from research centers and analytical companies around the world use social networks data to model social, economic, political and other processes from the personal to the state level, to develop mechanisms to influence these processes, as well as to develop innovative analytical and business applications and services. Moreover, when working with data from social networks, the factors such as instability in the quality of user content, the problem of privacy of users' personal information during storage and processing, as well as frequent updates of user models and functionality must be considered. All this requires constant improvement of algorithms for resolving various analytical and business problems [39-41].

5. Social network analysis systems

The key tasks of information and analytical support while working with social networks in most research are as follows [13, 18, 20, 24, 31, 38]:

- Social networks monitoring and analysis;
- Forecasting;
- Management.

Monitoring includes obtaining and structuring raw data. In this case, the texts in the sent messages, contacts between users, references to external sources are gathered. The capabilities of the system are mainly determined by the richness of the data used and their processing mode. Real-time monitoring systems are more difficult to implement than the systems using retrospective data collection.

The analysis covers numerous stages of raw data processing. For example, initially, the calculation of key indicators to answer simple quantitative questions such as "How many messages did user A write?" Then, the identification of statistical and structural regularities in the data allows to understand the essence of the studied network.

Forecasting is possible after defining a mathematical model of the information process. Statistical models and the models of dynamic processes in graphs can be used here (for example, the spread of epidemics, etc.). Management consists of providing targeted effects on the social network to change information processes to the required state.

The tasks of analysis, forecasting and management may differ depending on the problem statement, i.e., an end user of the system. The end users interested in the analysis, forecasting and management of social networks are divided into the following groups [24, 31]:

- State authorities and local self-government bodies;
- Organizations of public and private sectors of the economy;
- Commercial organizations (primarily "brand");
- Research institutions;
- Mass media;
- Authoritative structures;
- Society, including political parties, individuals, etc.

Research indicates that currently the most developed systems are social network analysis systems for commercial organizations. In [6, 20, 32], regardless of the end user, social network analysis systems are classified based on several characteristics:

- Levels of social networks analysis. The systems can perform simple monitoring, social networks analysis, forecasting of processes in social networks, management of social networks.
- Models of social networks. The systems can implement certain structural models of social networks (models of random graphs, non-scale networks, models of large-scale networks, analysis of dynamic networks, etc.), information distribution models (Markov models, finite automation theory, diffusion models of innovation), etc. [5, 11].
- Methods of data analysis. Most statistical methods and graph theory are applied in social network analysis systems. Methods of semantic analysis and text tonality analysis can also be used to classify systems.
- Objects of social networks analysis. People, social groups, communities, organizations, parties, countries, etc. can perform as social objects. They are called actors or nodes of social networks. Relationships include not only mutual communication between the actors, but also the exchange of various resources, joint activities, including conflicts, and so on. Formally, the social network is described by a graph, the vertices of the graph are actors, and the edge is a set of connections [6]. Systems can focus on the analysis of sub-networks and communities, selected users, news, opinions, external nodes, which are mainly social network objects.
- Scope of information sources. The systems can collect and analyze data from classic online social networks (Facebook, VKontakte), blogs (LiveJournal), microblogs (Twitter), photo and video sharing services (YouTube, Flickr), forums, and etc.
- The amount of data processed. Systems can be designed for large amounts of data (Big Data software products) [33].

The study of the characteristics of social network analysis systems shows that the best systems for analyzing social networks shall:

- Work in different modes (real time and retrospective) at all analysis levels;
- Analyze the various objects of the social network and consider all the connections between the objects;
- Be based on mathematical models and methods of intellectual analysis (methods of analysis of statistics and graphs, etc.);
- Be integrated with various open-source subsystems (social networks, blog sites, news sources, etc.) to collect information;
- Process Big data, etc.

Currently, social networking research is increasing significantly due to the proliferation of social networking websites. Social networking services have shaped a large segment of people's lives.

6. Conclusion

The conduction of demographic research is of the urgent issues for the management of demographic processes and implementation of targeted demographic policy. The use of the Internet, social networking services, social media, and smartphones are new unique sources of data for the study of demographic behavior. The analysis of large amounts of data collected in these sources allows for demographic research and the emergence of new ideas and knowledge related to demographic behavior.

Knowing the characteristics of the age structure of the population during this or that period, it is possible to make a well-founded hypothesis about the future trend of births and deaths, as well as other demographic processes and reproduction of the population as a whole. Knowing these features, it is possible to assess the probability of a number of problems in the economic and social spheres, to predict the demand for this or that commodity or service, and so on. The impact of demographic catastrophes (such as loss of life as a result of wars)

also affects declining birth rates and average annual population growth, urban and rural migration, and other demographic processes.

Thus, the analysis demonstrated that the analysis of large amounts of data collected on social networking services on the Internet allowed making decisions and solving some problems of traditional demography.

References

1. Gasimova, R.T. Abbasli, R.N. (2020). Analysis of search algorithms in large volumes of digital data. Problems of Information Technology, 1, 98-108. (Azerbaijani)
2. Internet World Stats. https://internetworldstats.com/stats.htm
3. Reinsel, D., Gantz, J., Rydning, J. (2018). Data Age 2025: The Digitization of the World–From Edge to Core, IDC White Pape, Framingham, MA, USA. https://seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf
4. Alguliyev, R.M., Yusifov, F.F. (2021). Architectural principles of national e-demographic system establishment. Problems of the Information Society, 2, 3-17. (Azerbaijani)
5. United Nations Development Program, "Human Development", textbook, 2014. https://az.undp.org/content/azerbaijan/en/home/library/human_development/HDtextbook.html (Azerbaijani)
6. Alguliyev, R.M., Aliguliyev, R.M., Yusifov, F.F., Alakbarova, I.Y. (2019). Formation of electronic demography as an effective tool for social research and monitoring of population data. State and municipal management issues. Higher School of Economics (NRU HSE), 4, 61–86. (Russian)
7. Hajirahimova, M.Sh., Aliyeva, A.S. (2021). Characteristics of the digital formation demography in the Big data era. Problems of Information Technology, 2, 74–88. (Azerbaijani)
8. Letouze, E., Jutting, J. (2015). Official statistics, big data and human development. Data-Pop Alliance White Paper Series. https://paris21.org/
9. Imamverdiyev, Y.N. (2010). Social networks analysis: concepts, models and research problems. Problems of the Information Society, 2, 9–20. (Azerbaijani)
10. Alguliyev, R.M., Yusifov, F.F. (2009). Social networks as a tool to improve the efficiency of public administration mechanisms. Telecommunications, 9, 25-30. (Russian)
11. Abdullayeva, F.D. (2009). About a method of building relationships between personal data in social networks. Problems of Control and Informatics, 1, 118-123. (Russian)
12. Demography: textbook. (2004). Ed. prof. V.G. Glushkova. Moscow: KnoRus. https://azstat.org/Kitweb/zipfiles/11030.pdf (Russian)
13. Kachagina, O.V. (2016). Fundamentals of Demography: Fundamentals of Theory and Practical Tasks: Textbook. Ulyanovsk: UlGU. (Russian)
14. Andrichenko, L.V., Meshcheryakova, M.A. (2012). Information registers as an effective means of collecting and monitoring population data. Journal of Russian Law, 8, 16–40. (Russian)
15. Borisov, V.A. (2001). Demography. Moscow: NOTABENE Publishing House. http://sociologos.ru/upload/File/Methods/Demography_Borisov.pdf (Russian)
16. Billari, F., D'Amuri, F., Marcucci, J. (2013). Forecasting births using google. Annual Meeting of the Population Association of America, New Orleans, USA.
17. Poulain, M., Herm, A. (2013). Central populatıon registers as a source of demographic statistics in Europe. Population, 68(2), 183–212.
18. Yildiz, D., Munson, J., Vitali, A. and et al. (2017). Using Twitter data for demographic research. Demographic Research, 37(46), 1477–1514.
19. Cesare, N., Lee, H., McCormick, T. et al. (2018). Promises and Pitfalls of Using Digital Traces for Demographic Research. Demography, 55(5), 1979–1999.
20. Gil-Clavel, S., Zagheni, E. (2019). Demographic Differentials in Facebook Usage around the World. İn Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019), Munich, Germany, 11-14 June 2019 (pp. 647-650).
21. Fire, M., Elovici, Y. (2015). Data Mining of Online Genealogy Datasets for Revealing Lifespan Patterns in Human Population. ACM Transactions on Intelligent Systems and Technology, 6(2), 1-24.
22. Mammadova M.H., Jabrayilova Z.Q., Isayeva A.M. (2020) Development of informative parameters for decision-making based on doctor-patient relations in social media resources. Problems of Information Technology, 1, 49–61. (Azerbaijani)
23. Careja, R., Bevelander, P. (2018). Using population registers for migration and integration research: examples from Denmark and Sweden. Comparative Migration Studies, 6(1), 6-19.
24. Gubanov, D.A., Novikov, D.A., Chkhartishvili, A.G. (2010). Social networks: models of information influence, control and confrontation. Moscow: Publishing House of Physical and Mathematical Literature. (Russian)
25. Kashepov, A.V., Volgin, N.A., Veselkova, I.N., Zvereva, N.V., Khorev, B.S., Khoreva, O.B., Shcherbakov, A.I. (2007). Demography. Moscow: Publishing house RAGS. (Russian)
26. Demography. Textbook for universities. (2005). Ed. Volgin, N.A., Rybakovsky, L.L., Kalmykova, N.M., Arkhangelsky, V.N., Ivanova, E.I., Zakharova, O.D., Ivanova, A.E., Denisenko , M.B., Tikhomirova, N.P., Tikhomirova, T.M. Moscow: Logos Publishing House. (Russian)
27. Chekmyshev, O.A., Yashunsky, A.D. (2014). Extraction and use of data from electronic social networks. IPM preprints im. M.V. Keldysh. Moscow. (Russian)
28. Korshunov, A., Beloborodov, I., Buzun, N., Avanesov, V., Pastukhov, R., Chikhradze, K., Kozlov, I., Gomzin, A., Andrianov, I., Sysoev, A., Ipatov, S., Filonenko, I., Chuprina, K., Turdakov, D., Kuznetsov, S. (2014). Social Network Analysis: Methods and Applications. Proceedings of the Institute for System Programming RAS, 26(1), 439-456. https://doi.org/10.15514/ISPRAS-2014-26(1)-19 (Russian)
29. Filippova, K. (2012). User demographics and language in an implicit social network. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea, 12-14 July 2012 (pp. 1478-1488). https://dl.acm.org/doi/pdf/10.5555/2390948.2391117
30. Yuxiao, D., Yang, Y., Jie, T., Nitesh, V. Chawla. (2014). Inferring user demographics and social strategies in mobile social networks. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining–KDD '14. August 2014 (pp. 15-24). https://doi.org/10.1145/2623330.2623703
31. Bartunov, S., Korshunov, A. (2012). Identification of users of social networks on the Internet based on social connections. Reports of the All-Russian scientific conference "Analysis of images, networks and texts" - AIST'2012. Ekaterinburg, Russia, March 16-18, 2012 (p.52-67). https://publications.hse.ru/mirror/pubs/share/folder/jspp0u781q/direct/69355023.pdf (Russian)

32. Culotta, A., Kumar Ravi, N., Cutler, J. (2016). Predicting twitter user demographics using distant supervision from website traffic data. Journal of Artificial Intelligence Research, 55(1), 389-408.
33. Bouquet, P., Bortoli, S. (2010). Entity-centric Social Profile Integration. In Proceedings of the International Workshop on Linking of User Profiles and Applications in the Social Semantic Web (LUPAS 2010), Heraklion, Greece, 30 May-3 June 2010 (pp.52-57).
34. All statistics of the Internet and social networks for 2021 - numbers and trends in the world and in Russia. https://web-canape.ru/business/vsya-statistika-interneta-i-socsetej-na-2021-god-cifry-i-trendy-v-mire-i-v-rossii/ (Russian)
35. Social media statistics around the world. https://gs.statcounter.com/social-media-stats (Russian)
36. Gasimova, R. (2016). Big data analytics: available approaches, problems and solutions. Problems of Information Technology, 1, 62-78. https://jpit.az/uploads/article/en/BIG_DATA_ANALYTICS_AVAILABLE_APPROACHES,_PROBLEMS_AND_SOLUTIONS.pdf (Azerbaijani)
37. Rasim M. Alguliyev, Rena T. Gasimova, Rahim N. Abbasli (2017). The obstacles in Big Data process. International Journal of Modern Education and Computer Science (JMECS), 9(3), 28-35. http://mecs-press.org/ijmecs/ijmecs-v9-n3/IJMECS-V9-N3-4.pdf
38. Madden, S. (2012). From Databases to Big Data. IEEE Internet Computing, 16(3), 4-6.
39. Boyd, D.M., Ellison, N.B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1), 210-230, https://doi.org/10.1111/j.1083-6101.2007.00393.x
40. George Pallis, Demetrios Zeinalipour-Yazti, Marios D. Dikaiakos. (2011). Online Social Networks: Status and Trends. New Directions in Web Data Management 1, Studies in Computational Intelligence, 331, 213-234.
1. Baden, R., Bender, A., Spring, N., Bhattacharjee, B., Starin, D. (2009). Persona: an Online Social Network with User-defined Privacy. In Proceedings of the ACM SIGCOMM 2009 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Barcelona, Spain, August 2009 (pp.135-146).