
Design of a Socio-economic Processes Monitoring System Based on Network Analysis and Big Data

Submitted 14/12/18, 1st revision 4/2/19, accepted 13/2/19

A.G. Polyakova^{1,2}, M.P. Loginov³, A.I. Serebrennikova⁴, E.I. Thalassinos⁵

Abstract:

Socio-economic policy should satisfy the interests of the society as much as possible and contribute to improving the quality of life. This actualizes the role of developing the theoretical and methodological foundations for building an innovative information system for monitoring the socio-economic situation and population responses.

The study built and tested an algorithm for supporting management decisions based on the collection of large data sets of socio-economic information based on the principles of the digital economy and processing them through network analysis.

The algorithm is focused on building a monitoring system that presupposes a synergy of the authorities and the society, not only in its pensionary part, but also among the masses, which are diverse in their representativeness.

The result of the study was the formation of a theoretical and methodological framework for creating a system for making management decisions and assessing the effectiveness of the activities of government bodies, based on the principles of reflection of the final beneficiaries of economic policy.

Keywords: *Socio-economic development, socio-economic system, digital economy, monitoring, management decisions, big data, network analysis.*

JEL Classification: *E66, C55, D03, M15.*

¹*Plekhanov Russian University of Economics, Moscow, Russia, agpolyakova@mail.ru*

²*Financial University under the Government of the Russian Federation, Moscow, Russia*

³*Ural State University of Economics, Ekaterinburg, Russia, port-all@mail.ru*

⁴*Ural State University of Economics, Ekaterinburg, Russia, serebrennikova_ai@usue.ru*

⁵*University of Piraeus, Greece, thalassi@unipi.gr*

1. Introduction

Issues of socio-economic policy take a leading place in the socio-political discourse, but a consensus on their content, methods and mechanisms of implementation has not yet been reached. One of the reasons is that the property possession, material well-being and self-esteem of the population depend on a significant number of factors, which are multi-directional and are perceived differently by different social groups. Considering that social and economic policy should satisfy all or most citizens as much as possible, it becomes necessary to promptly measure this satisfaction, including not only post factum on the basis of statistical or sociological observation, but also tentatively before making appropriate decisions. The foregoing necessitates the development of tools for rapid assessment of the consequences and the reaction of the population to events and newsbreaks: the effects of intensified efforts in the sphere of social and economic policy need to be studied.

Today, there are no effective feedback mechanisms between the population, as the carrier of power, and the authorities, to which the population has delegated authority. In this regard, it often happens that the decisions taken by the authorities in the socio-economic sphere not only lack approval by the broad masses, but also provoke strong protest. Some of these problems could be solved by proper information and explanatory work, but the other, more substantial part is because even if one wants to get validation of some solutions through discussion with the population, the executive and legislative branches do not have effective opportunities for this.

The second facet of the problem, no less important, is the insufficient representation of small social groups that are identical by the criteria of place of residence, professional affiliation, etc., (for example, residents of small settlements), whose problems in the scale of the economy will never be top priorities, but they are significant in the community scale, since they significantly impair their quality of life. There are many such loci, but they are distributed throughout the country, therefore they are divided and their opinion may not be heard. At the same time, the authorities also have a request to hear these private pockets of problems, but an economically viable way of collecting timely and relevant information has not been developed. As a result, science and public administration are forced to operate with two data sources: official statistics and the results of various field studies. In the first case, there is a narrow coverage of statistical observation of socio-economic processes, low degree of efficiency, and in the second - an exponential increase in the cost of research as the sample grows and the coverage of indicators. Hence, the problems of monitoring design and implementation are frequently proposed to solve using the resources of digital economy and big data operationalization.

Thus, the well-known thesis that social relations reflect economic policy can be successfully translated in relation to the Internet space, which seems to be poorly understood. The connectedness of space in the "internet reality" is higher and it will

be the driver of the information cascade, whereas in the field of socio-economic policy the problem field in question may not be considered at all and not voiced. The circumstances and arguments listed above sound in favor of the need and relevance of modeling a system for monitoring socio-economic status and population responses based on big data.

2. Literature review

The prologue to the conceptualization of the digital economy were the works of Wiener (1948) and Turing (1950). The concept itself was made possible by summarizing and revisiting the achievements of Bell (1973), Toffler (1995) and Galbraith (1967), supplemented by the theory of information networks and network analysis methodology. The analysis of social networks as a methodology has been enriched and extended after Wasserman *et al.* (1985) and Granovetter *et al.* (2005).

Brandt *et al.* (2009) note that since the early 1990s, a lot of information about network analysis has begun to appear, but academic and political rhetoric has been accompanied by an extremely small amount of analytical research. The problem development can be found in O'Sullivan *et al.* (2014), where empirical data substantiates the possibility of applying key concepts and definitions from graph theory, using graph measures to study the local and global network structure. It also focuses on the structural features of the central level network, cohesive subgraphs and structural equivalence. Specific tasks like identifying leaders of opinion we proposed to solve in Sengupta *et al.* (2015), which describes a method of formalizing relations between actors, considering the characteristics of these relations. From earlier works Farrugia *et al.* (2010) stands out; it investigated several ways to build networks of interaction between actors and described approaches to identifying opinion leaders.

A significant number of works by American, European and Russian researchers are devoted to the formation of new forms of interaction between the state and society in the digital economy. However, among the numerous papers, one deserves a better attention is Taleb (2007) who has focused on a fundamentally new paradigm of ensuring the stability of systems based on dynamic disequilibrium. Testing of the ideas presented by Taleb will be possible using big-data driven monitoring systems.

Scientific works summarizing the practical experience in building monitoring systems published by international organizations are of considerable interest (Thalassinos *et al.*, 2012). For example, since there is no single indicator for assessing the socioeconomic situation of an individual, many are attempting a composite index to measure it. Notable is the experience of the European Union Statistics Office, which provides data for comparing European countries and regions using aggregates. They are an integral part of the infrastructure created and maintained by Eurostat, and are defined as a consolidated measure related to a key issue or phenomenon, derived from a series of observed facts. Summary indicators

can be used to display relative positions or to display positive / negative dynamics. In contrast to the data from which they are obtained, indicators are used for specific purposes, which are determined by the questions to which it is intended to answer, and by the parties asking these questions.

Practice shows that the experience of using big data is not so extensive, although it is becoming increasingly common. Big data are widely used in management of various objects, starting with corporations and ending with government agencies (Hanas *et al.*, 2007). China is among the leaders of big data employment at the level of public administration; the State Council of China recently announced guidelines for stimulating the development of big data technologies. Thus, at the end of 2017, the development of a platform for interagency data exchange was completed. In addition, by the end of 2018, a unified government data exchange base was supposed to deploy. It is expected that big data technologies will contribute to the efficient use of resources, enhance management, introduce industrial innovations, create new business models and support the economic restructuring.

Monitoring systems developed by companies, as a rule, are used for scoring purposes. Almost every individual is connected to social networks that provide a lot of data. Companies use platforms to sift through data gathered from Facebook, Google+, LinkedIn or Twitter. These platforms collect data (such as the keywords used by individuals in their posts) and turn them into accessible information. Such work is extremely diverse. Tinkoff credit systems, OTP Bank, Sociohub.ru, Double Data, Skorista analytical agency, IQ' men, Palantir, Social network agency are among the pioneers in working with social networking data. Several analytical reviews of these organizations, which summarize the experience of various organizations in line with the studied issues, are publicly available.

3. Methodology

The purpose of the study was to design and test an innovation-type of information system designed to monitor the socio-economic situation and people's reactions. The core components of the system's principles come from the digital economy, synergy of government and society – not only in its pensionary part, but also in its wide representation.

Standard statistics data processing tools (sampling and grouping, correlation and regression analysis of time series, etc.) are not suitable for analyzing big data, because data volumes are growing exponentially and opportunities to cover them with standard analysis and generalization tools are already limited by available capacity and a high degree of dependence on the degree of automation of the data collection process. Since it is also not possible to collect sociological data on demand online, the solution lies in the use of hidden and implicit data that are generated by the social and economic activity of individuals, including activity in the Internet space reflecting the processes that take place in social, economic,

political areas of interaction between individuals and institutions in the "real" world. At the same time, the Internet space is becoming more and more independent in terms of generating social interactions, as well more open for use in various spheres of life. Internet space appears to many as an alternative reality, but even in the "physical space" not all processes can be designated as "objective", since there is always a subjective distortion of information perception, as well as competition of interests.

However, if "physical" space is the object of research in many areas of knowledge, then generally accepted and approved approaches to the study of Internet space have not yet been developed, which makes it necessary to find and adapt appropriate tools that meet the requirements of the scientific approach. There are various methods of large data analysis: data mining, machine learning, pattern recognition, building artificial neural networks, etc. One of the promising areas of analyzing big data is network analysis. Network analysis (social network analysis) is a dynamically developing area of quantitative research in the socio-economic sphere, which has both theoretical value and significant applied value. Network analysis is characterized by the presence of its own methods for assessing and reporting data on the relationship between economic actors.

Traditional approaches to data analysis have lost their due to an excess of information considered, which objectively or implicitly influences decision making or the choice of behavior patterns. The need to admit that the multiply increasing amount of information in the process of setting goals or planning activities seems to be logical, however, the Big Data toolkit available today is far from the desired operability, and therefore the niche of unmet demand continues to expand.

Social network analysis is a modern trend, which involves building a relationship graph in which nodes are represented by communication participants, and edges are links between them. Evaluation of schemes of discourse network organization allows not only to describe the nature of relations between participants of communication, but also to evaluate the strength of connections. In contemporary literature, including journalism and business journalism, the construction and visualization of networks is quite common, which, nevertheless, does not fully reveal the analytical and search capabilities of social network analysis. The attractiveness of the instrument lies in the possibility of its application to large amounts of data, the "manual" processing of which seems to be a very laborious process, and automated - through self-learning linguistic mechanisms - is costly in terms of the resources required. At the same time, social network analysis eliminates the need for the analyst to surf through all the records, drawing attention only to the key nodes and connections that form the subject of his research interest.

Social network analysis originates in graph theory, the roots of which in mathematical sciences were laid by L. Euler in the 18th century, but it became a stand-alone research methodology after Wasserman *et al.* (1985) and got promoted

by Granovetter *et al.* (2005) and Brandt *et al.* (2009). Transformation in understanding the methodological foundations of strategic development and planning has already happened: in modern studies, the multiparameter nature of economic space is increasingly noted, which makes it possible to move away from transportation costs as the main measure of the distance between its elements in favor of multidimensional assessments describing the intensity and nature of interactions between actors, usually not covered by statistical observation to the extent necessary for management.

In addition to this, series of publications studying the economic space relatedness appeared as in the works of Polyakova *et al.* (2018), Miheeva *et al.* (2018), Akhmetshin *et al.* (2018), Voronkova *et al.* (2018), Kolmakov *et al.* (2019). It is highly probable that in the future, the Big Data principles and network analysis will become the tools to eliminate these gaps and to create the basis for a new strategic planning and forecasting system. Software design implemented as a user interface enabling the algorithm for collecting and processing data, suggested the following solutions:

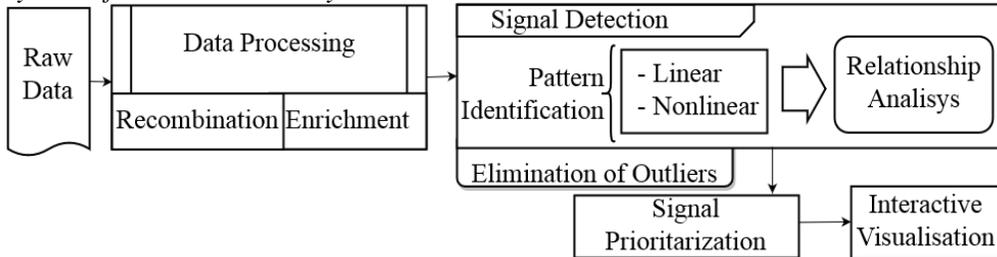
- ✓ Data collection algorithms for the following social networks: Twitter, Facebook, Instagram.
- ✓ Input data binding module (different sources of data to one database).
- ✓ Module to display and filter the retrieved database.
- ✓ Data extractor to upload entries to machine-readable format (for example, XML / JSON / CSV) for analysis.

The monitoring system should contribute to the creation in the medium term of a new model of decision making and evaluation. During the study, the process of big data preparation and processing was considered as a part of socio-economic processes monitoring system. An important step in designing a big data system is finding the optimal structure, which involves choosing how to store and distribute data between its nodes, select equipment and data warehouses. At the same time, it is necessary to consider the specific features of the queries performed, as well as ways to optimize the structure of the system. The process of preparing analytics and big data processing in the system is schematically represented in Figure 1.

The development of algorithms for analyzing arrays of big data allows to build extremely large-scale networks of interaction of actors, for example, in social networks (Facebook, Twitter, Instagram, etc.). Currently, the social network analysis is becoming an interdisciplinary research tool, a key technique in many areas of knowledge. Network analysis is applicable for building diagnostic monitoring systems or identifying hidden dependencies, which is well illustrated on the theory of the “six handshakes” - S. Milgram's experiment, according to which any two people on Earth are separated by no more than five levels of common friends (and, accordingly, six connections). Social networks analysis can be used as a tool for

content analysis of media and blogs to assess popularity of certain ideas, concepts and images, as well as to identify the channels of their distribution.

Figure 1. The process of preparing and processing big data in the monitoring system of socio-economic dynamics



The problem to be solved within the empirical research was to test the tools of social network analysis and develop a research algorithm available for replication in solving a wide range of analytical and search problems. As a case study, a network was constructed and analyzed, reflecting public interest in the process of discussing the reduction of federal budget expenditures in the State Duma of the Russian Federation (the lower house of the Russian Parliament). The network was built for the keyword "budget". When testing the toolkit, practical implementation of the principles of big data was provided according to Twitter. The relevance of references during the last 5 days from the date of the event was considered.

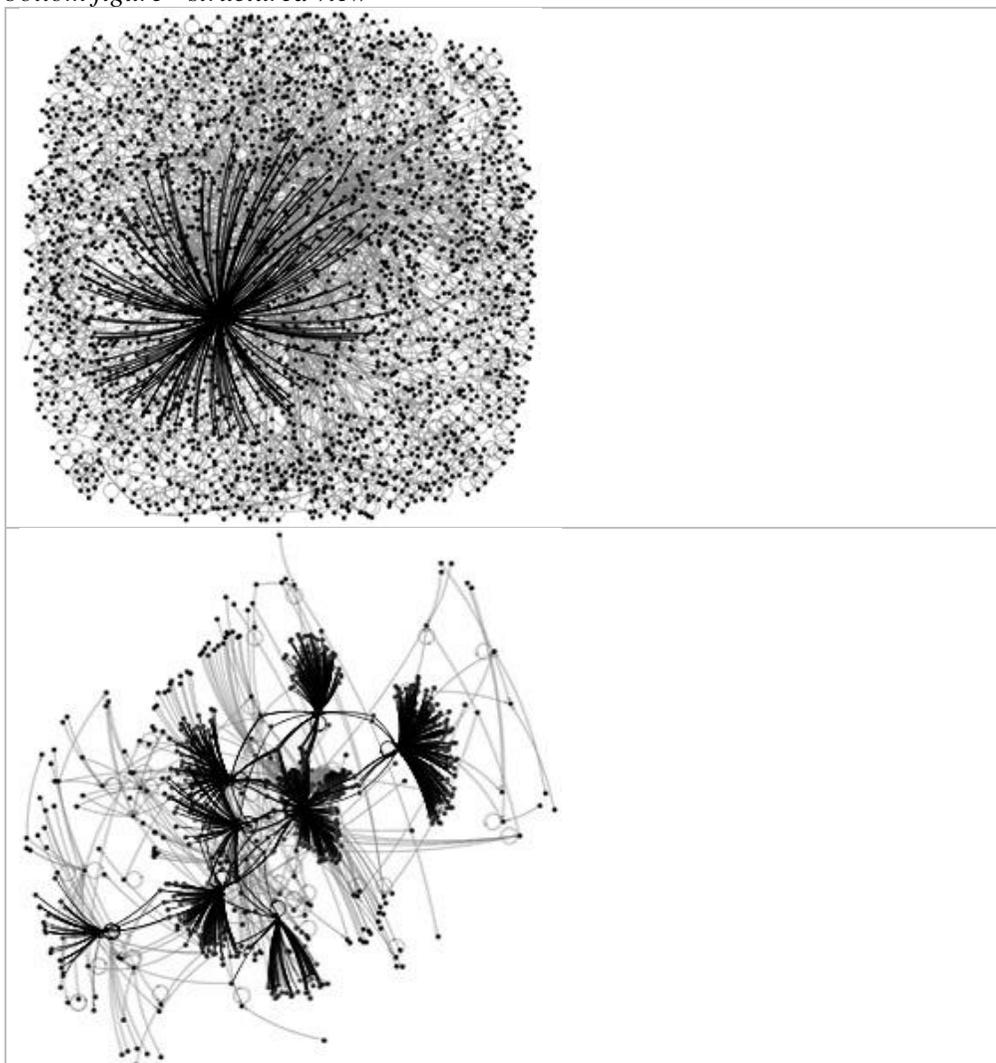
4. Results and discussion

The graph nodes are the users of Twitter, whose messages mentioned the key phrase "budget" in any context during the specified period (Figure 2). Thus, there are several discussion zones in the social space with their specific content. This allows to identify clusters based on a certain user identity and to presume the existence of social moods of a certain nature. Moreover, if sociological surveys are usually focused on identifying moods within the framework of the choices made by the respondent, the clusters in the network have natural personality traits that are inherent in groups.

As can be seen from the figure above, the problem in scope covers several centers of informational activity, as well as several separate user groups. In the illustration, these groups and clusters are highlighted in contrast. A qualitative analysis of the composition and content of the identified information activity centers allows to identify key actors that have the most significant impact on public opinion on the issue in focus. The nature of such an impact may consist either in establishing a trend that other actors support, transmitting information further through their network of contacts, or in initiating discussion on a designated question, within

which respondents' answers can be not only meaningful, but also emotional, that is, expressed in the spirit of approval or disapproval, sympathy or antipathy, etc.

Figure 2. Network graph, built on the "resonant" event in different visualizations (for the keyword "budget" at the date of the budget discussion in the State Duma), bottom figure - structured view



At the next stage, network metrics and quantitative parameters identifying the actors and their informational messages were analyzed. The choice of actors was based on two characteristics - in-degree and out-degree. The first is a measure of the number of incoming connections, that is, the absolute popularity of the actor, and the second is the number of outgoing connections, that is, a measure of the activity of the actor. The following data obtained by the results of network analysis are available (Figure 3).

Figure 3. Analyzed final data

Sources of newsbreaks, that provoked statements with the key phrase "budget" (addresses with newsbreaks)		Publication title	Number of references to messages
Most cited domains according to references to authors.			Number of references
Most frequently mentioned words in the sample	Number of references	Most frequently mentioned pairs of words in the sample	Frequency of references during the period under study
Most frequently mentioned actors by network users in the sample			Total number of references
Most active accounts by the number of messages in the population		Total number of messages for the period of the actor's existence	

Thus, pockets of public opinion were identified. It is noteworthy that the impersonal actor, whose mission is quite distant from public policy and information support of the management decision-making process, may have the greatest valence in the aggregate.

5. Conclusion

Designing and evaluating network connections based on big data can become a tool for creating systems (subsystems) of strategic planning. It is possible that the experience of building links can be the basis for spatial planning based on Big Data and the study of network effects in spatial economics. Successful practices in this area already exist: Zhong *et al.* (2014) empirically proved the possibility of applying network analysis to the study of spatial interactions in order to identify spatial centers, hubs and boundaries that are significant from the point of view of managing spatial interactions.

According to the results of the work, methodological developments were created for conducting classes on big data analysis as part of the training program for the Higher Reserve of Managerial Personnel of the Russian Federation. Methodical developments are based on the practical application of network analysis tools for the needs of strategic spatial development management. Successful testing of social network analysis tools for analyzing big data and developing a research algorithm made it possible to identify sources of spatial consolidation available for replication in solving a wide range of analytical and search problems, and also to use them in the educational process.

References:

Akhmetshin, E.M., Dzhavatov, D.K., Sverdlikova, E.A., Sokolov, M.S., Avdeeva, O.A. & Yavkin, G.P. 2018. The influence of innovation on social and economic

- development of the Russian regions. *European Research Studies Journal*, 21(Special Issue 2), 767-776.
- Brandt, A., Hahn, C., Krätke, S. & Kiese, M. 2009. Metropolitan regions in the knowledge economy: Network analysis as a strategic information tool. *Tijdschrift Voor Economische En Sociale Geografie*, 100(2), 236-249.
- Farrugia, M. & Quigley, A. 2010. Actor Identification in Implicit Relational Data Sources. *Studies in Computational Intelligence*, doi:10.1007/978-3-642-13422-7_5.
- Fienberg, S.E., Meyer, M.M. & Wasserman, S.S. 1985. Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80(389), 51-67, doi:10.1080/01621459.1985.10477129.
- Hanias, P.M., Curtis, G.P. and Thalassinos, E.J. 2007. Non-linear dynamics and chaos: The case of the price indicator at the Athens Stock Exchange. *International Research Journal of Finance and Economics*, 11(1), 154-163.
- Granovetter, M. 2005. The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, 19(1), 33-50, doi:10.1257/0895330053147958.
- Miheeva, N.M., Yu, V.O., Magsumov, T.A., Bulanakova, M.A., Kulchytskiy, A.V., Polyakova, A.G. 2018. Public diplomacy development stages through the black sea economic cooperation organization. *International Journal of Mechanical Engineering and Technology*, 9(11), 1382-1391.
- O'Sullivan, D. 2014. Spatial network analysis. In M.M. Fischer & P. Nijkamp (Eds.) *Handbook of regional science*, 1253-1273. Berlin: Springer Verlag.
- Polyakova, A.G., Akhmetshin, E.M., Goloshchapova, L.V., Rakhmeeva, I.I., Noeva, E.E. & Rakovskiy, V.I. 2018. A model of regional economic space modernization. *European Research Studies Journal*, 21(Special Issue 2), 624-634.
- Sengupta, S., Das, I., Sarkar, R. & Kolya, A.K. 2016. Event & event actor identification with event-sentiment relation. Paper presented at the Proceedings of 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2015, 423-428.
- Voronkova, O.Y., Zadimidcenko, A.M., Goloshchapova, L.V., Polyakova, A.G., Kamolov, S.G. & Akhmetshin, E.M. 2018. Economic and mathematical modeling of regional industrial processes. *European Research Studies Journal*, 21(4), 268-279.
- Zhong C., Arisona S.M., Huang X., Batty M., Schmitt G. 2014. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, 28(11), 2178-2199.
- Kolmakov, V.V., Polyakova, A.G., Karpova, S.V., Golovina, A.N. 2019. Methodological approach to cluster development based on competitive specialization of regions. *Economy of region*, 15(1).
- Wiener, N. 1948. *Cybernetics: Or Control and Communication in the Animal and the Machine*. Paris, Hermann & Cie, MIT Press.
- Turing, A. 1950. Computing Machinery and Intelligence. *Mind*, 49, 433-460.
- Toffler, A. & Toffler, H. 1995. *Creating a New Civilization: The Politics of the Third Wave*. Atlanta GA: Turner Publishing.
- Galbraith, J.K. 1967. *The New Industrial State*. Princeton University Press.
- Bell, D. 1973. *The coming of post-industrial society: A venture of social forecasting*. New York, Basic Books.
- Taleb, N.N. 2007. Black Swan and Domains of Statistics, *The American Statistician*, 61(3).
- Thalassinos, I.E., Hanias, P.M. and Curtis, G.P. 2012. Time series prediction with neural networks for the Athens Stock Exchange indicator. *European Research Studies Journal*, 15(2), 23-31.