
Study of regional innovation ecosystem based on the big data intellectual analysis

Leyla Gamidullaeva*, Alexey Finogeev,
Kirill Lychagin and Sergey Vasin

Penza State University,
RF, Penza, Krasnaya st., 40, Russia
Email: gamidullaeva@gmail.com
Email: alexeyfinogeev@gmail.com
Email: klychagin@ec-leasing.ru
Email: pspu-met@mail.ru
*Corresponding author

Abstract: The article is aimed at studying regional innovation ecosystems on the basis of intellectual analysis of data on firms' innovative development. The authors developed a methodology for modelling and exploring the innovative potential and investment appeal of regional firms. The goal of the proposed methodology is to reduce transaction costs on innovative actors' interaction and to increase the effectiveness of their innovation activities. To achieve the research objective, it is necessary to solve the tasks of collecting data on innovation activities of enterprises in the regions, to develop methods for assessing their innovative potential, to synthesise forecasting models of innovative development. The methodology utilises the tools of the informational and analytical platform with the flow architecture of the data collection and integration system and the convergent mechanism for big data processing. The objective is to support decision-making and management of innovative processes in the regions.

Keywords: innovation system; innovative activity; innovative potential; cluster analysis; benchmarking analysis; investment appeal; transaction costs; big data; cyber-social system; intellectual analysis.

Reference to this paper should be made as follows: Gamidullaeva, L., Finogeev, A., Lychagin, K. and Vasin, S. (2020) 'Study of regional innovation ecosystem based on the big data intellectual analysis', *Int. J. Business Innovation and Research*, Vol. 23, No. 3, pp.313–337.

Biographical notes: Leyla Gamidullaeva holds a PhD in Economics. She is a Doctoral candidate and Associate Professor at the Department of Management and Economic Security of Penza State University, Russia. She has authored over 200 refereed publications and over ten books in innovation management, regional economic growth and innovation development, networking and collaboration.

Alexey Finogeev received his Doctorate of Sciences from the Penza State University, Penza, Russia. During his research career, he worked as an Engineer at the Penza Polytechnic Institute, a researcher at Ottawa University, an Assistant Professor, an Associate Professor and a Professor at Penza State University. Since 2004, he is a Full Time Professor at the Department of Computer Aided Design of Penza State University. The area of his scientific research is connected with information technologies, network technologies,

wireless technologies, technologies for digital economy, big data, sensor networks, SCADA system, 3D modelling and virtual reality, and computer aided design.

Kirill Lychagin is a postgraduate student at the Department of CAD, Penza State University. Currently, he is the Head of the Sector for Creating and Maintaining Subsystems for Integrating Data of Analytical Systems at the company 'EC-leasing' Co.

Sergey Vasin is a Doctor of Economic Sciences, Professor and Vice-President for International Affairs at Penza State University, Russia. He has authored more than 250 publications in social and economic system transformation and regional economic development.

1 Introduction

The regional innovation system is a set of economic agents including companies, research and education enterprises in the regions. Modern information technologies make it possible to implement cyber-social tools to manage the innovation system by means of the elements of artificial intelligence. The tools must secure safe and reliable interaction of agents, adapt to their needs, analyse current situations and assist in making decisions on improvements of the innovative potential and competitiveness of companies and a region in whole.

The stimulation of innovation economy subjects' interaction determines the paths to improve the efficiency of their innovative development. Due to growing complexity of innovative processes and the interdisciplinary nature of applied technologies, the innovations more and more depend on the knowledge spread outside an individual enterprise. For this reason, the developed countries in recent years have put into practice the concept of open innovations (Chesbrough, 2003). According to the concept, the dynamic process of research and innovative development should be considered in open environment, where knowledge sources may be located both inside and outside an individual company (Vasin et al., 2019; Gamidullaeva et al., 2019; Gamidullaeva, 2018).

Owing to information exchange, companies may often obtain competitive advantages in the market conditions, when implementing innovation technologies created outside the said companies. The competitive advantages are the scientific, technological and educational potential, technologies and innovations at all stages of production and promotion of goods and services. That is why international companies strive for connecting corporative networks of their units and departments that take part in generation, storing and application of knowledge with the external internet. This helps to reinforce company's internal potential by accumulating new knowledge. The concept of open innovations makes researching a quick, commercially beneficial and diverse process. It integrates into corporative strategies of many firms requiring new business strategies of their development and opportunities to work in markets of new ideas and technologies.

The regional innovation ecosystem (Rothschild, 1990) as a highly organised network for interaction of companies, which is adapted to the regional social and institutional context. The regional innovation ecosystem represents an open dynamic system featuring

loops of feedback and response lagging. The concept is focused on such factors as characteristics of the interacting participants' network, trust relations between participants, socio-cultural features of the individuals, etc. The model of network collaborative innovations (Gloor, 2006) assumes that innovations are the results of interactions, while companies and enterprises that shape a regional innovation ecosystem act as growth sources for the economy (Tafti et al., 2019).

The building of regional innovation ecosystems determines a necessity to search and select methods, models and approaches to efficient management. The central aspect is extraction and accumulation of knowledge in the interaction between innovation system's participants. The basis for such interaction is open networks, charge-free distributed software, big data and the technologies of intellectual analysis and machine learning. Then, the regional innovation ecosystem may be represented as a cyber-social structure (Sheth et al., 2013). The information for modelling, researching and analysing in the cyber-social system is extracted in real time from social networks and other open internet sources, as well as from sensor cyber-physical systems of the internet of things (IoT). Data storing and processing utilise cloud technologies and structures. At the same time, individual knowledge, competencies, socio-cultural features are considered integral parts of the cyber-social system (Ley et al., 2015).

To model and research a cyber-social innovation ecosystem, it is offered to apply the neo-institutional approach utilising basic categories of formal and informal institutions, as well as transaction costs (North, 1991). Here, the efficiency of innovation processes is determined by such a criterion, as minimum transaction costs for economic agents (Gamidullaeva and Tolstykh, 2017). Reduction of transaction costs means optimisation within normal parameters, as the said costs are objectively needed for economic system functioning.

To lower transaction costs, innovation process participants must interact without hindrance when implementing joint projects focused on creation and commercialisation of innovations. Thus, the goal of regional innovation ecosystem management can be defined as an aspiration to lower transaction costs through increased efficiency and predictability of interactions between economic agents. This requires having a clear understanding of the features and characteristics of the regional innovation ecosystem development. The main problem in the study of regional innovation ecosystems is the complexity of data collection, as well as the presence of hidden dependencies and patterns, which are difficult to identify using traditional economic analysis tools.

Recently, the big data analytics has become a significant frontier for innovation, research and development (Vasin et al., 2018). The objective of the article is to develop methods and tools of decision-making support based on knowledge engineering and intellectual analysis of big real-time and retrospective data on the indicators of firms' development and regional innovation environment.

2 Theoretical background

The term 'big data' is a series of approaches, tools and methods for processing structured and unstructured data of very large volumes and considerable diversity for obtaining human-perceptible results (Dumbill, 2014). Big data now allows us to examine urban and

regional phenomena in ways that were previously not possible (Schintler and Chen, 2018). Rae and Singleton (2015) argue that:

“Huge volumes of data are generated within regions daily, such as through the use or management of public services (e.g. global positioning satellite (GPS) tracking of law enforcement officials and use of healthcare facilities), or captured from transport-related activities (e.g. road flow information gathered by networks of traffic cameras). Other data are also generated by the private sector, including transaction data associated with consumption or the use of social media, where content are georeferenced by mobile devices.”

Hilbert (2013) has also pointed out that “the crux of the ‘big data’ paradigm is actually not the increasingly large amount of data itself, but its analysis for intelligent decision-making.”

The advances being made in data modelling capabilities using ‘big data’ such as data mining, large-scale simulation models, and agent-based techniques – offer considerable potential to: enhance our understanding of the complexities of urban and regional systems; and assist in finding solutions to pressing problems (Batty et al., 2012).

Much literature addresses the technical and methodical implementation, the transformative strength of big data, enhancement of firm performance by building analytics capability, or other managerial issues. Little work covers the transformation process from first ideas to ready analytics applications (Kayser et al., 2018).

Big data can be structured (data from sensors, log files, financial data, input data, data on visits to websites, etc.), unstructured (satellite imagery, scientific data, photo and video documents: recordings from surveillance cameras and video recorders, radar data: road, meteorological, oceanographic, seismological observation data, documentation of organisations: the texts of official documents, registers, survey results, e-mails, etc., social network data, mobile data) and semi-structured. To proceed these large amounts of data in an inexpensive and efficient way, using parallelism. Big data is data whose scale, diversity and complexity require a new architecture, methods, algorithms and analytics to manage it and extract value and hidden knowledge from it (Alamri and Qureshi, 2015).

The key principles for using analytics include responsibility, transparency and consent, privacy, validity, access, minimising adverse impacts, etc. (Martin and Thawabieh, 2017).

While definitions of big data vary, it is usually characterised by five properties, referred to as the 5V’s.

- 1 *Volume*: Typically, big data refers to massive volumes of data, usually in terabytes (TB), petabytes (PB), zettabytes (ZB) or more.
- 2 *Variety*: The data consists of a mixture of structured, unstructured and semi-structured information, drawn from such vastly heterogeneous sources as RFID, web searches, social media, mobile sensors like GPS and accelerometers, high fidelity industrial sensors, video streaming, etc.
- 3 *Velocity*: Big data arrives at varying speeds ranging from milliseconds to days to years, and has differing requirements on the speed with which it is to be processed.
- 4 *Value*: Some researchers consider value as a key characteristic of big data, with data being considered valuable if useful information (from a business or engineering perspective) can be extracted from large datasets where individual data points may not carry any value by themselves.

- 5 *Veracity*: This refers to the accuracy and trustworthiness of the data. This becomes increasingly relevant when large numbers of users may be reluctant to report truthful data due to privacy and security concerns (Saravanan et al., 2019).

The theoretical basis of the study is the convergence of a number of theories, including the theory of regional innovation ecosystem (Rothschild, 1990; Marshall, 1993, etc.), concepts of open and collaborative innovation (Chesbrough, 2003; Gloor, 2006), neo-institutional theory (North, 1991), etc. The emergence and application of digital technologies in regional settings significantly imply the changes in the research tools utilised in the modern regional innovation management. Hence, there is a need to enrich an open innovation and ecosystem approach to the regional development with big data and IoT tools to transform regional innovation system into a cyber-social innovation ecosystem.

3 Research methodology

The final goal of the study is to develop methods and tools of decision-making support based on expert knowledge engineering and intellectual analysis of big real-time and retrospective data on the indicators of firms' development and regional innovation environment. Such methods and tools provide informational and analytical support to the innovative development of individual companies and a region as a whole.

Table 1 Analytical research methods

<i>Market research methods</i>	<i>Data mining methods</i>
Strategic analysis methods	Classification without response
Boston Consulting Group matrix	Classification with response
Industrial analysis ('five forces')	Associations
Strategic groups	Event sequence analysis
SWOT	Time-to-event forecasting
External environment analysis	Artificial neural networks
Problem analysis	Decision trees, symbol rules
Stakeholders scenario	Clustering, methods of nearest neighbour and k-nearest neighbour
STEEP analysis	Support vector machine
Financial analysis	Bayesian networks
Cost effectiveness evaluation	Regression analysis (linear regression)
Sustained growth rates	Correlation-regression analysis
Assessment of competitors and competitiveness	Benchmarking, cluster analysis, mining association rules
Determination of 'dead zones'	Non-hierarchical methods of cluster analysis (k-means and k-median algorithms)
Customer value estimation	Limited search method
Segmentation of consumers	Evolutionary programming and genetic algorithms
Management profiling	Methods of data visualisation, analytical report generation

Table 1 demonstrates well-established analytical methods implemented when researching and analysing markets and applicable for researching innovative development processes.

Table 2 shows data mining methods used in business analytics and applicable for innovative development researching.

Table 2 Data mining methods in business analytics

<i>Operations with data to reveal and utilise formalised regularities</i>	
<i>Direct data usage</i>	<i>Template distilling (data templates are extracted from initial data and transformed into formal constructions)</i>
Cluster analysis	Logical induction methods
Nearest neighbour method	Fuzzy queries and analyses
K-nearest neighbour method	Symbol rules
Analogical reasoning	Decision trees
	Genetic algorithms
	Cross tabulation
	Agents, Bayesian (belief) networks
	Cross tabular visualisation
	Equation-based methods
<i>Machine learning</i>	
<i>Data mining statistical methods</i>	<i>Data mining cybernetic methods</i>
Analysis of the nature of statistical data (testing of hypotheses of stationarity, normality, independence, homogeneity, assessment of the distribution function and its parameters, etc.)	Artificial neural networks (identification, clustering, forecast)
	Evolutionary programming (including algorithms by the group method of data handling)
Identification of linkages and patterns (dispersion analysis, linear and nonlinear regression analysis, correlation analysis, etc.)	Genetic algorithms (optimisation)
	Associative memory (search for analogues, prototypes)
Multivariate statistical analysis (linear and nonlinear discriminant analysis, cluster analysis, component analysis, factor analysis, etc.)	Fuzzy logic
	Decision trees
Dynamic models and forecasting on the basis of time series	Expert knowledge processing systems
<i>Classification, clustering</i>	
<i>Descriptive methods (search for templates that describe interpretable data)</i>	<i>Forecasting methods (prognostic modelling of unknown and prospective values of target variables)</i>
K-means algorithm	Neural networks
K-median	Decision trees
Hierarchic methods of cluster analysis	Regression analysis
Self-organising Kohonen maps	Nearest neighbour method
Cross tabular visualisation methods	Support vector machines

Each method features strengths and weaknesses, and none of them claim to be universal. However, there is an opportunity to provide solution to the set problems of regional innovative development research using the data obtained by methods of intellectual analysis.

The regional innovation ecosystem management is focused on searching and selecting the optimal way of evolutionary development and further stability maintenance. According to the system-synergetic approach, the management methodology consists in synthesising and analysing multiple paths of development corresponding to management goals, determining domains of attraction (attractors) of the said paths, barriers on optimal trajectories and measures to overcome these barriers. This approach allows to take account of the dynamic aspect of innovation ecosystem development by researching such categories of the participants' economic activity, as complexity, openness, uncertainty, self-organisation, etc. The identification of attractors in the field of predictable development makes it possible to interpret system changes in the context of establishing relations between innovation agents in the regional environment. The determination of order parameters and impact factors defining the behaviour of ecosystem's components enables to synthesise models and methods of prognostic modelling of system's dynamics for the purpose of selecting strategies of its evolutionary development.

To achieve the management goal, it is necessary to elaborate tools, including models, methods and algorithms of intellectual analysis, forecasting and making decisions on innovative development strategy selection and improvement of companies and entire region's innovative potential. The tool complex is a cyber-social system of regional innovative development management that combines practical management with the software and hardware analytical platform intended for collecting and analysing big data on current conditions of economic agents, prognostic modelling, benchmarking and assessment of competitiveness development and improvement strategies. Benchmarking allows to transfer and adapt companies-leaders' experience of efficient functioning in the field of innovation development and implementation in a region at many other companies and enterprises.

The algorithmic method of innovation ecosystem research may be represented as a sequence of the following steps:

- 1 Stratification of settlements in regions to sort out areas of innovative development according to set criteria, and synthesis of innovation agents' universum in the strata. Collection of data on innovation agents from open internet sources to be added to the universum, determination of indicators and factors impacting innovative development. Determination of the optimal stratification method by the agents' innovative potential and investment appeal evaluation criteria.
- 2 Determination of a complex of indicators and integral criteria of company's innovative development, as well as impact factors decelerating or accelerating this process. Examples of effective criteria are regional innovation activity rating, industrial production rating, investment volume, population's living standard indicators, etc.
- 3 Sensitivity assessment and selection of significant indicators and factors impacting the results of agents' innovative development by the method of multivariate dispersion analysis. These results may subsequently be applied as order parameters and clustering attributes. Ranking of indicators and factors by the extent of influence on regional innovation system's effectiveness in strata.

- 4 Determination of the optimal number of innovation company clusters in strata. Hierarchic multilayer clustering of innovation activity participants. Ranking of companies in clusters to select innovation leaders that become innovative development attractors, as well as 'models' for benchmarking of other participants' innovative potential in clusters and strata.
- 5 Determination of 'model' sets of values of innovative development target indicators taking into account current conditions and factors impacting innovation activities and investment susceptibility in clusters. These sets characterise attractors and virtually represent the space of formal and informal institutions of the optimal regional innovation environment.
- 6 Benchmarking of companies' innovative potentials with ones of innovation leaders in clusters to forecast and estimate the influence of their interaction on transaction costs and impact factors, their innovative potentials and investment appeal.
- 7 Synthesis of reports on current conditions and the dynamics of probable improvement of companies and entire region's innovation climates. Synthesis of strategies for pessimistic, optimistic and optimal scenarios of companies' innovative development in clusters. Elaboration of a set of measures and recommendations to improve efficiency and effectiveness of innovation activities, to reduce transaction costs, to develop a road map of companies' innovative potentials and investment appeal improvement in clusters, strata and regions.
- 8 Comparison of innovative development zones in regions by impact factors to sort and identify regions with similar characteristics of the innovation environment. Comparative analysis of regions by integral criteria of innovative development and synthesis of a prognostic model of effective interaction of regional innovation systems.

Let us consider the stages of the proposed methodology in detail.

3.1 Collection of data on innovation agents and stratification of innovative development zones

Before stratification and clustering, there is a preliminary stage of searching and selecting potential innovation activity participants from many economic agents to evaluate their innovative potentials. We shall introduce the notion 'universum of innovation agents' meaning a multitude of potential innovation activity participants that we find out about as a result of the searching and selection. Hence, the formation of a multitude of innovation agents can be broken down into two stages:

- 1 Search for new economic subjects to be introduced to the universum.
- 2 Determination of a set of characteristics of the chosen subjects, by which we can consider the latter potential innovation agents.

The following indices may be considered as basic ones, upon which a decision on including an economic agent into the universum can be made:

- Agent's innovative potential value (internal potential). The indicator is calculated on the basis of company's internal data on intellectual property assets created and/or owned by another agent for a certain period of time.
- Agent's innovation investment value. The indicator is estimated as a volume of own funds invested by an agent to create and/or acquire innovations from the third party for a certain period of time.
- Agent's attracted investment value. The indicator is estimated as a volume of investments attracted to create and/or acquire innovations from the third party for a certain period of time.
- Forecasted agent's investment appeal value. The indicator is estimated as a forecasted volume of external investments that an agent can attract for innovation activities for a certain period of time.
- Indicator of investment appeal loss. The indicator is determined as a function having value 1, if the ratio of the integral level of investment risk for an agent (calculated by the expert method) for the last three months to the risk level for the last three months is below 50%, and having value 0, if on the contrary.

The main sources of data on economic agents for the research of regional institutional environment are the agents' websites, special databases, websites with official statistical data by the Russian Statistical Agency, Ministry of Economic Development, Ministry of Finances, Ministry of Education and Science of Russia, websites of core ministries and agencies and other resources associated with the innovation activity of Russian regions, statistical collections of 'Higher School of Economics', websites with published results of economic and sociological studies and expert surveys in Russian regions.

Information search and data extraction are realised through the multiagent approach. For this purpose, the analytical platform includes searching modules for regular tracking of websites and data extraction there from. The modules extract characteristics needed for innovative potential and investment appeal evaluation, as well as geolocation data, designation, intellectual property information and other data. The modules are designed in the form of searching robots (Shmid et al., 2016) by means of a crawler creating library in Python Scrappy language. The crawlers execute the 'curious user' algorithm that opens each link of interest on a social network page/site. They compile sequences of queries to service several parallel data loading flows.

As data on agents are taken from different sources, it is necessary to settle the issues of designation and address unification, double information exclusion, unreliable information elimination, etc. For example, an efficient way of address unification is direct (address-to-coordinates transformation) and reverse (coordinates-to-address transformation) geocoding using Yandex Geocoder. PostgreSQL database with address information clearing and geotagging is applied to integrate data from different sources. The system also searches for duplicates using coordinates and deletes double records.

Stratification is required to group townships and settlements, where innovation actors are located, into strata taking into account the amount of population and innovation process peculiarities in regions. The stratification method contains a list of operations intended to obtain the optimal number of strata according to the amount of population and regional peculiarities with innovation agents' geotagging. The outcome is the compilation of lists of regional townships and settlements belonging to similar strata by

certain criteria. In the general case, stratification may be performed by different methods, out of which it is necessary to choose the optimal one for the set objective. Such methods of mathematical statistics as dispersion and factor analysis are applied to assess stratification methods. Stratification also determines the degree of similarity and disassociation between strata and a preliminary set of attributes to subsequently search for the optimal clustering of innovation enterprises taking into account their geolocation. In this case, it is reasonable to apply two rules when evaluating similarities and differences of strata. The first rule proposes to reveal similarities/differences of strata by estimating the level of economic subjects' innovative potential. The second rule – by the amount of investments in strata directed to develop and implement innovations.

There are several ways of division into strata. The simplest one is the division by the amount of able-bodied population. Another way is to divide by the amount of population employed at innovation enterprises. The third way is to divide into strata according to the company's innovative potential index. The fourth way is the division in accordance with investment appeal and investment risk criteria. In the course of the research, all four ways of stratification were realised, and the preference was given to the third way. The adequacy of the given stratification model was assessed experimentally by calculating static difference of strata using mean values and medians of the chosen order parameters through the method of multivariate dispersion analysis. This method presupposes division into five strata:

- Stratum 1 Townships including enterprises with estimated total innovative potential over ten billion of roubles volume.
- Stratum 2 Townships including enterprises with estimated total innovative potential over one billion of roubles.
- Stratum 3 Townships including enterprises with estimated total innovative potential over ten million of roubles.
- Stratum 4 Townships including enterprises with estimated total innovative potential over 100 thousand of roubles.
- Stratum 5 Townships including enterprises with estimated total innovative potential less than 100 thousand of roubles.

3.2 Indicators of innovation ecosystem's development in the institutional context

It is required to elaborate a list of indicators (attributes) of innovation activity to analyse and assess innovation ecosystem' conditions, as well as to forecast probable innovative development. The task is complicated by the fact that such indicators may be quite numerous, including not only direct indicators, for example, a number of registered intellectual property assets, innovation financing volume, but also indirect ones displaying the level of regional development, such as transport infrastructure presence and condition, the level of company employers' income, the condition of informational and telecommunication infrastructure and internet connection costs among the regional service providers, the degree of competitive interactions between innovation agents, etc. Therefore, it is necessary at the first stage to compile a maximum possible list of indicators that characterise companies' innovative conditions directly and indirectly.

After that, it is required to estimate sensitivity of internal indices of innovative development to the indicators and to select a group of the most sensitive parameters that should be called order parameters in terms of the system-synergetic approach. The method of regional innovation ecosystem's characteristics analysis in the institutional context has been specifically developed to reliably and comprehensively select significant indices and factors impacting innovative development. The indicators of innovation activity are picked out as a result the formal and informal institutional context analysis. The groups of order parameters and clustering attributes are selected after sensitivity assessment and ranking of indices to solve such problems of intellectual analysis, as clustering and benchmarking.

All indicators that evaluate innovation ecosystem's institutions are divided as follows:

- a Elementary indicators – indicators measured with the help of one attribute (number of patents issued, GRP level, etc.).
- b Complex indicators – sets of qualitatively similar elementary indicators that can provide different indices, for example, innovation activity index, interpersonal trust index.
- c Composite indicators – sets of complex indicators.

Besides, all indicators may be divided into two major groups: quantitative (calculated by mathematical formulas) and qualitative (requiring expert evaluations). According to the institutional context scheme, all indicators may be divided into formal and informal institutions to evaluate the innovation environment as follows:

- 1 The institution of legal support of innovation activities including:
 - a a group of indicators demonstrating completeness of legal support of innovation activities
 - b a group of indicators displaying the level of law-enforcement mechanisms development
 - c a group of indicators reflecting conformity of federal and regional laws in the field of innovations.
- 2 The institution of regional development including:
 - a a group of indicators of financial security of innovation activities
 - b a group of indicators of human resource availability
 - c a group of indicators of material and technical resources availability.
- 3 The institution of strategic planning of innovation activities, which determines a group of indicators reflecting availability of strategic innovation planning elements.
- 4 The institution of professional education and science including:
 - a a group of indicators of educational system productivity
 - b a group of indicators of educational quality
 - c a group of indicators of financial security of scientific research and developments.
- 5 The institution of integrated communications, which determines a group of indicators of innovation agents' communicative activity.

- 6 The institution of innovation activity standardisation, which determines a group of indicators characterising standards that regulate innovation activities.
- 7 The institution for monitoring of changes in target markets and technological spheres, which determines a group of indicators characterising innovation activity monitoring techniques.
- 8 The institution of informational support, which determines a group of indicators characterising information resources and the degree of informational and telecommunicational technologies development.
- 9 The institution of competition, which determines a group of indicators characterising competitive interaction and interconnection of agents.
- 10 The institution of internal and external cooperation including:
 - a a group of indicators characterising cooperative interaction of agents with higher educational institutions and research centres
 - b a group of indicators characterising interaction and integration of innovation agents between each other
 - c a group of indicators characterising cooperative interaction of company's employees with the innovative potential.
- 11 The institution of intellectual property protection, which determines a group of indicators characterising the degree of company's intellectual property protection.
- 12 The institution of innovation activity funding including:
 - a a group of indicators characterising the degree of non-repayable funding and state funding of innovation enterprises
 - b a group of indicators of direct funding of research and development at innovation enterprises
 - c a group of indicators characterising availability and effectiveness of innovation enterprises.
- 13 The institution of reputation and trust including:
 - a a group of indicators characterising company's reputation
 - b a group of indicators characterising reputation and trust in regional administrative authorities in a stratum and a region.
- 14 The institution of shadow economic relations including:
 - a a group of indicators characterising company's 'opaque operations'
 - b a group of indicators of the corruption level.
- 15 The institution of innovation mediation, which determines a group of indicators characterising availability and effectiveness of innovation mediation (business incubators, business angels, venture capital funds, etc.).

Having collected initial institutional data, it is necessary to calculate integral indices to evaluate the institutions. The integral index calculation consists of rationing, aggregating and scaling. When rationing, the indicators are reduced to the dimensionless form and uniform range by the maximum-minimum method:

$$\bar{X} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where x – value of an indicator, $x_{\min}$ – minimum value of an indicator, and $x_{\max}$ – maximum value of an indicator.

The aggregation combines indicators to obtain an integral index taking into account weight coefficients assigned to indicators by the expert method. For this purpose, the values of subordinate levels of details are added to weight coefficients to obtain the values of a superior level (from indicator parameters to indicators, from indicators to directions, from directions to an integral index):

$$P_{\text{int}} = \sum_1^N \frac{x_i \times w_i}{N} \quad (2)$$

where x_i – value of i^{th} indicator of total number N , and w_i – indicator of i^{th} indicator's significance (weight).

The significance of certain indicators' parameters of innovation institutions is defined by the correlation analysis to avoid subjectivity of expert evaluations and to ensure credibility of integral indices. To resolve the multicollinearity problem, the indicators with low significance are excluded from the set of integral indices. For this purpose, it is required to assess sensitivity of integral indices to indicators by the method of multivariate dispersion analysis.

3.3 The method of indicator sensitivity evaluation for selection of order parameters and clustering attributes

In order to perform the intellectual analysis, the number of institutional indicators is too big and requires reduction of the phase space dimensions. For this purpose, it is necessary to single out the most significant indicators (order parameters), which have the greatest effect on effectiveness values of the innovative potential and investment appeal. The set of order parameters is a list of attributes (factors) that characterise company's innovation activity in strata and regions, potential investors' activity, optimality of existing infrastructure, etc.

The set of parameters will be used subsequently to the clustering problem as a set of attributes to synthesise a multidimensional clustering space. Clustering is required to group regional enterprises into innovation clusters and to identify innovative development leaders that surround themselves with other economic entities. According to the theory of synergy, innovation leaders represent centres of attraction (attractors).

The indicator's sensitivity is to be evaluated by a multidimensional model and the multivariate dispersion analysis. The point is to study the influence of selected indicators and factors on effectiveness indices for each problem to be solved, in particular, for the problem of stratification and clustering. The dispersion analysis is based on the analysis of deviations of units in the analysed aggregate from the arithmetical mean. The metric of deviations is dispersion. At the same time, deviations caused by the impact of a factor are determined and compared with deviations caused by random factors. In the outcome, the indicators with greater influence on the result than the influence of random factors are selected from a multitude of indicators. It is these indicators that we shall define as significant impact factors.

First of all, the method is applied to select the optimal way of stratification. For this purpose, it is necessary to evaluate four suggested ways of stratification by checking similarities and distinctions of strata in evaluating the effectiveness indices of companies' innovative potential and investment appeal in strata. The following are regarded as effectiveness indices of stratification:

- a Average number of created and attracted innovations at all enterprises in a stratum and their maximum number for the analysed period of time.
- b Median on the quantity scale of innovations at enterprises in a stratum with a half of measured values lying above it and another half below.
- c Percentiles (indicating what percentage of company's innovations is below a certain level), in particular, 5% percentile (the number of innovations below a set value from the sample are observed in 5% of companies), 25% percentile (the number of innovations below a set value from the whole sample are observed in 25% of companies), 75% percentile (the number of innovations below a set value of sales from the whole sample are observed in 75% of companies), 95% percentile (the number of innovations below a set value of sales from the whole sample are observed in 95% of companies), and 99% percentile (the number of innovations below a set value of sales from the whole sample are observed in 99% of companies).

In the second case, the following are taken as effectiveness indices of investment appeal:

- a Average volume of investments in implementation of innovations in a stratum and the maximum volume for the analysed period of time.
- b Median on the measured investment value scale with a half of measured values lying above it and another half below.
- c Percentiles (indicating what percentage of company's investment volumes is below a certain level), in particular, 5% percentile (the volume of investments below a set value from the sample is observed in 5% of companies), 25% percentile (the volume of investments below a set value from the whole sample is observed in 25% of companies), 75% percentile (the volume of investments below a set value of sales from the whole sample is observed in 75% of companies), 95% percentile (the volume of investments below a set value of sales from the whole sample is observed in 95% of companies), and 99% percentile (the volume of investments below a set value of sales from the whole sample is observed in 99% of companies).

The null hypothesis is formulated during the analysis assuming that the studied factors have no impact on values of effectiveness indices and the obtained differences are accidental. The null hypothesis suggests as follows: mean values of the number of innovations and the volume of investments in strata will be similar for all strata regardless of stratification ways, and therefore, it does not matter which division method to choose. After that, provided the null hypothesis is fair, the probability of obtaining the observed differences is determined. If the probability is below the level of significance $\alpha = 0.05$, then the null hypothesis is rejected, and it is concluded that the research results are statistically significant. Although this fact does not precisely mean that the effect of the studied factors is proven, it is unlikely that integral indices are influenced by random

factors. Total dispersion D_t of indices includes factor dispersion, induced by set factors, and residual dispersion, defined by random factors:

$$D_t = D_f + D_p \quad (3)$$

where D_f – factor dispersion, which is characterised by distinction of mean values in each stratum and depends on the influence of the studied factor, by which a stratum is differentiated, and D_p – residual dispersion, which defines random changes caused by random factors.

A linear model of the multivariate model has the following form:

$$x_{i,j,k} = \mu_i + a_{i,j} + b_{i,k} + \dots + (ab)_{i,j,k} + e_{i,j,k} \quad (4)$$

where $x_{i,j,k}$ – result of i^{th} factor measuring, μ_i – arithmetical mean for i^{th} factor, $a_{i,j}$ – systematic error of i^{th} indicator measuring in j^{th} stratum by factor A , $b_{i,k}$ – systematic error of i^{th} indicator measuring in k^{th} stratum by factor B , $(ab)_{i,j,k}$ – systematic error of i^{th} indicator measuring in j, k^{th} stratum for a combination of factors A and B , and $e_{i,j,k}$ – accidental error of i^{th} factor measuring.

The dispersion analysis algorithm consists in determining the correlation of factor dispersion and residual dispersion in the measured data. The sum of squares of indicator's value deviation from the arithmetical mean is taken as the variability index. We shall designate this sum as S . The model of multivariate analysis includes sums of squares of deviations S_f for each factor individually, sums of squares of deviations for a set of factors and sum of squares of deviations S_p for random factors. For example, three-factor model S_t includes sum of squares of deviations S_f^A for factor A , sum of squares S_f^B of factor B , sum of squares S_f^C of factor C , sums of squares of interactions S_f^{AB} of factors A and B , S_f^{AC} of factors A and C , S_f^{BC} of factors B and C , interactions S_f^{ABC} of three factors A, B and C , and sums of squares of interactions S_p of random factors:

$$S_t = S_f + S_p = (S_f^A + S_f^B + S_f^C + S_f^{AB} + S_f^{BC} + S_f^{AC} + S_f^{ABC}) + S_p \quad (5)$$

Mathematical expectation M of the general sample is taken as a precise value of each indicator. In this case, an accidental measuring error is the difference between the result of i^{th} measurement of indicator $x_{i,j}$ in j^{th} stratum and the arithmetical mean for given indicator M_j . If the impact factor systematically influences the effectiveness index, the systematic error for the factor's impact will be the difference between the arithmetical mean in a stratum and the mathematical expectation of the general sample. Dispersions of mean-square deviations are determined as follows:

$$D_t = \frac{\sum_{i=1}^{n_j} (x_{i,j} - M)^2}{N - 1} \quad (6)$$

$$D_f = \frac{\sum_{i=1}^{n_j} (M_j - M)^2}{J - 1} \quad (7)$$

$$D_p = \frac{\sum_{i=1}^{n_j} (x_{i,j}^p - M)^2}{N - J} \quad (8)$$

where n_j – number of factor values in j^{th} stratum, N – volume of full sample of all factor volumes, and J – number of strata.

The correlation of factor and residual dispersion has Fisher distribution. It is determined using Fisher-Snedecor F -criterion:

$$F = \frac{D_f^i}{D_p} \geq F_{st}. \quad (9)$$

The given criterion is calculated for each i^{th} factor and its interactions and compared with the set standard value.

The impact power of factor η is determined by the share of factor dispersion (D_f) in the total dispersion (D_t):

$$\eta = \frac{D_f}{D_t} 100\%. \quad (10)$$

As a result of the dispersion analysis, the null hypothesis has been rejected and the optimal way of stratification has been selected. The same method is also applied when choosing a clustering model on the basis of effectiveness indices evaluation.

3.4 The model and method of innovation enterprises clustering in strata

The clustering problem is solved for all enterprises in a stratum and a region included in the universum both in industrial layers and in layers of small, medium and large enterprises in accordance with federal laws. This way, the method of multilayer hierarchic clustering is realised. As a result, for each stratum and region, there are clusters of similar innovation enterprises determined in general, by industrial affiliation and for different types of enterprises in accordance with federal laws. Regional affiliation and geolocation are also taken into account. Objects in one cluster should be similar in selected attributes, but differ from objects in other clusters. The decision-making criterion regarding the number of clusters is distances, over which the integration of companies takes place. Being selected earlier, the factors impacting effectiveness indices of innovative development are taken as clustering attributes. A modified k-means method is used as a clustering algorithm (Phillips, 2002). It requires setting the amount of clusters and the amount of initialisations of the iteration process of clustering (initial centroids), which depends on the time assigned for clustering. We shall designate a multitude of innovation agents in the universum of a stratum (region) as $\{x_i\}_{i=1}^Q$ where Q – number of given enterprises (general, for individual industries, by size) in a chosen stratum (region), Euclidean metric as d , a number of attributes as P . To determine the number of clusters, we shall use the k-means algorithm to set minimum and maximum numbers of sub-clusters from 2 to 75. We shall designate the obtained clustering models as M_K . The extent of intercluster difference for clusters K should be determined as follows:

$$W_K = \sum_{k=1}^K \sum_{i: M_K(x_i)=k} \sum_{j: M_K(x_j)=k} d^2(x_i, x_j). \quad (11)$$

The extent of complete distinction should be obtained according to the following formula:

$$T = \sum_{i=1}^N \sum_{j=1}^N d^2(x_i, x_j). \quad (12)$$

We shall interpret correlation $\frac{W_K}{T}$ as a share of unexplained distinction between agents within clusters. This correlation decreases as K grows. The optimal amount of clusters K^* shall be defined as:

$$K^* = \min \left\{ K \left| \frac{W_K}{T} \leq 0.2 \right. \right\}. \quad (13)$$

Thus, the minimum number of clusters is chosen in such a manner, so that the share of unexplained distinction will be under 20%. In the general case, it is possible to set any value depending on limitations on the number of clusters and the type of dependence of correlation $\frac{W_K}{T}$ on K . As a result, we will obtain a clustering model, including the following:

$\{\mu_{z,Type}^p\}_{p=1}^P$ mean attribute values for cluster z of certain *Type*

$\{\sigma_{z,Type}^p\}_{p=1}^P$ standard deviations of attributes for cluster z of certain *Type*

$K_{h,Type}^*$ optimal number of clusters for a stratum of certain *Type*

$M_{h,Type}^*$ clustering model with a number of clusters in a stratum h of certain *Type*.

The algorithm of the obtained clustering model application consists as follows. Let us assume there is an innovation enterprise of certain *Type*, belonging to cluster z , defined by attribute vector x . Using attribute vector, we shall define vector with elements:

$$y^p = \frac{x^p - \mu_{z,Type}^p}{\sigma_{z,Type}^p}. \quad (14)$$

Let us apply clustering model $M_{h,Type}^*$ to obtained vector y . As a result, we will obtain a cluster number in model $M_{h,Type}^*(y)$, which contains three components: stratum number, cluster type and cluster number.

As a result of its functioning, the algorithm has generated several models of innovating agents clustering:

- a for the total number of all enterprises in the universum by strata and regions
- b for the total number of large enterprises in the universum by strata and regions
- c for the total number of medium enterprises in the universum by strata and regions
- d for the total number of small enterprises in the universum by strata and regions

- e for the total number of all enterprises belonging to a certain industry in the universum by strata and regions
- f for the total number of large enterprises belonging to a certain industry in the universum by strata and regions
- g for the total number of medium enterprises belonging to a certain industry in the universum by strata and regions
- h for the total number of small enterprises belonging to a certain industry in the universum by strata and regions.

3.5 *Selecting innovation leaders and formation a reference model of institutional indicators*

At the next stage, the enterprises in the clusters, obtained according to different clustering models, are subject to ranking by values of effectiveness indices of innovative development in order to select innovation leaders. Ranked enterprise samples are compared for various clustering models with a view to their intersecting and selecting of leaders. In most cases, it turns out that each model has its own list and leaders. It is assumed that innovation leaders make no more than 10% of enterprises in each list. The indices of innovative development leaders should be considered determinative for the estimation of cluster's innovative potential in whole. For example, such estimates include the following:

- 1 Potential volume of funds acquired from innovation product sales within a cluster. It is determined as the arithmetical mean of given funds w_i^d acquired by innovation leaders (number of leaders – L) in cluster d :

$$W^d = \frac{1}{|L|} \sum_{i \in L} w_i^d. \quad (15)$$

- 2 Potential amount of intellectual property assets in a cluster. It is determined as the average number of such assets s_i in leaders' possession:

$$S = \frac{1}{|L|} \sum_{i \in L} s_i. \quad (16)$$

Innovation leaders should be counted as attractors for other enterprises in the phase space of innovative development. They are the basis for determining the institutional indicator values characterising the reference model of an innovation enterprise. The values of indicators will differ for each stratum and region, industry and enterprises of different types, and will virtually take account of group features of innovative development. Therefore, the indicators that characterise the model innovation leader for a stratum and a region will represent a multitude of variables identified by the fuzzy method applying the corresponding membership functions. The said multitude is taken as the reference model of target indicators of attractors' innovation activity. Other enterprises in clusters should attain to the level of these attractors.

The sensitivity of clustering effectiveness indices to changes of indicators is also evaluated to decrease the number of indicators in the reference model by the method of

multivariate dispersion analysis. The following can be particularly taken as effectiveness indices to study the impact of indicators:

- 1 Index of company's intellectual property possession. It is determined as the amount of intellectual property assets created or acquired by a company in a cluster during a year.
- 2 Average proportion of company's products created via innovations by a company in a cluster during a year:

$$P_{cp_{i,j}} = \frac{P_{i,j}^{UH}}{P_{i,j}^{o\phi}} / 12 \quad (17)$$

where $P_{i,j}^{UH}$ – volume of innovation products at i^{th} company from j^{th} cluster, and $P_{i,j}^{o\phi}$ – total volume of output.

- 3 Average proportion of own and attracted investments expended on creation and attraction of intellectual property assets by a company in a cluster during a year:

$$S_{cp_{i,j}} = \frac{C_{i,j}}{B_{i,j}} / 12 \quad (18)$$

where $C_{i,j}$ – volume of own and attracted funds expended on creation and acquisition of intellectual property assets by i^{th} company from j^{th} cluster, and $B_{i,j}$ – total volume of own and attracted funds.

- 4 Average proportion of financial resources acquired from implementation and application of innovations by a company during a year:

$$W_{cp_{i,j}} = \frac{V_{i,j}}{F_{i,j}} / 12 \quad (19)$$

where $V_{i,j}$ – proceeds from selling of products and services provided by i^{th} company from j^{th} cluster using innovations, and $F_{i,j}$ – company's total revenue.

- 5 Averaged indicator of return on investments into innovations of a company during a year.

$$K_{cp_{i,j}} = \frac{D_{i,j}^{UH}}{R_{i,j}^{o\phi}} / 12 \quad (20)$$

where $D_{i,j}^{UH}$ – total net present value of income from innovation projects of i^{th} company from j^{th} cluster, and $R_{i,j}^{o\phi}$ – total present value of expenditures on innovation projects.

As a result, a model indicator matrix is formed for each cluster in strata and regions. It characterises the innovative potential and investment appeal of leaders. Individual indicator matrixes are formed for other enterprises in a cluster. They will be compared afterwards with the model one with a view to revealing lagging in concrete indicators and developing strategies of innovative development. The general scheme of clustering, innovation leader selection and model indicator matrix synthesis is displayed in Figure 1.

Figure 1 A scheme of clustering and model matrix synthesis

The indicator matrix of innovative potential characterises company's innovation activity in clusters, strata and regions. This way, the interconnection of sensitive indicator values is established in clusters with opportunities to increase the effectiveness of investments in regional innovative development.

3.6 Benchmarking of companies' innovative potential

Cluster's innovation leader selection and model indicator matrix synthesis are required to complete the next objective of the intellectual analysis – benchmarking of all enterprises in a cluster to determine factors of innovation lagging from leaders. Such analysis of indicator matrixes is supposed to give answers to the following questions:

- 1 How many times and by what means is it possible to increase company's innovative potential under a current indicator matrix?
- 2 How many times and by what means will the investment return increase, if a company's indicator matrix changes?
- 3 What combination of indicator values has got the greatest possible effect on the growth of revenues from implementation of innovations?

In general, benchmarking is a process of identifying, studying and adapting the best practices and expertise of other organisations to improve own company operations. Benchmarking has well proved itself in modelling and studying of competitiveness of trade, research and financial institutions (Mkrttchian et al., 2017). The method involves synthesising of a nonlinear forecasting model that enables to analyse company's possible development dynamics by retrospective values of indicators in short-term and long-term perspective, as well as to estimate the impact of external factors on effectiveness indices of the innovative potential and investment appeal. According to the model of multilayer clustering by types of enterprises and industries, enterprises from one cluster interact between each other within a unified organisational field. The process of company's innovative development is studied within such field. The field's system is also considered from the position of the system-synergetic approach. Therefore, the dynamics of companies' interaction is reasonable to be described by a mathematical model in the form of a coherent system of nonlinear equations. Pairwise comparison of all companies in the universum in a cluster with innovation leaders is required to research interconnections and interactions in the process of innovative development and identification of factors that impede or intensify this process. From the mathematical point of view, the modelling system for pairwise comparative analysis should be described as two coherent systems of differential equations with nonlinear feedback (Forrester, 2006). The application of such model is substantiated by their oscillatory nature of innovative development indicators' fluctuations. Since all companies in a cluster are autonomous, it is reasonable to use the

concept ‘self-sustaining system’ when describing dynamic processes. The Van der Pol equation system may serve as a formalised mathematical description for such systems (Teodorchie, 1952). Van der Pol coupled equations (oscillators) represent a simple Ginzburg-Landau chain, proposed by the authors to describe the dynamics of nonlinear coherent dynamic systems, which analyses order-chaos transition modes. Then, by analogy with the interaction of two self-sustaining systems, we shall describe the interaction between two innovation agents in the course of benchmarking as two Van der Pol equations with lagging. The lagging is introduced for system stabilisation in case of occasional alteration of indicators due to fluctuations (Benninga and Sarig, 1997):

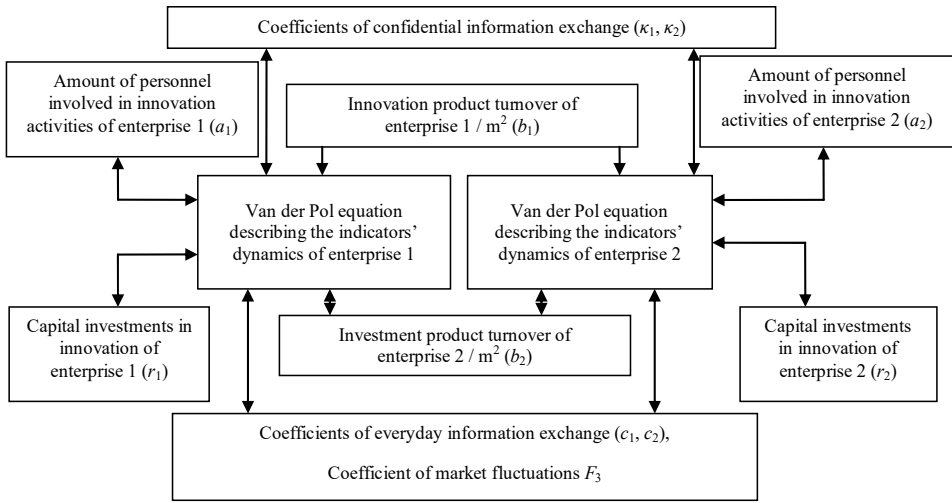
$$\begin{aligned}
 & \frac{d^2 X_1}{dt^2} - a_1(1 - Z_1) \frac{dX_1}{dt} + \omega_1^2(1 + k_1 X_2) X_1 = c_1 \frac{d^2 X_2}{dt^2} + r_1 F_1 + d_1 F_3 \\
 & b_1 Z_1 + T_1 \frac{dZ_1}{dt} = X_1^2 \\
 & \frac{d^2 X_2}{dt^2} - a_2(1 - Z_2) \frac{dX_2}{dt} + \omega_2^2(1 + k_2 X_1) X_2 = c_2 \frac{d^2 X_1}{dt^2} + r_2 F_2 + d_1 F_3 \\
 & b_2 Z_2 + T_2 \frac{dZ_2}{dt} = X_2^2
 \end{aligned} \tag{21}$$

Here X_1, X_2 – functions of innovative potential indices in time (amount of intellectual property assets, proportion of innovations, proportion of innovation investments, innovation investment return, volume of capital investments in innovations, etc.). Values a_1, a_2 are values proportional to the standard specific quantity of company’s personnel involved in innovation activities in annual average. Variables Z_1, Z_2 reflect the innovation research and testing efficiency in time. The efficiency is determined as a function proportional to the average monthly amount of intellectual property assets. The equations containing variables Z_1, Z_2 describe a function of performance of company departments in charge of innovations as a compensator of innovation leaps of main functions X_1, X_2 . The dependence of functions X_1, X_2 on functions Z_1, Z_2 is of nonlinear nature. Values T_1, T_2 are constants proportional to the average rate of innovation investment volume changes in the range from 0.1 to 2, which are determined, when the model is being adjusted for a specific company, and depend on the intensity of work of its innovation departments. They also depend on whether a company is an innovation leader. Values b_1, b_2 are indices proportional to the standard innovation product turnover per company’s unit area. Parameters ω_1, ω_2 are values opposite to the innovation product turnover time being one-year or one-month periods. Coefficients r_1, r_2 are constants proportional to capital expenditures on company’s innovative development (average annual index of company’s total capital expenditures is taken as 1, and company’s expenditures on innovations as a share of 1). Functions F_1, F_2 represent an exponent and display the resonance impact of the value of capital expenditures on innovations on the system’s solution. Coefficient d_1 equals to a thousandth from the value of average annual income from innovation product sales. The mathematical order of the given coefficient was defined in the work by Benzi et al. (1981) as the minimum noise amplitude required for stochastic resonance in bi-stable systems. Random function F_3 represents an exponent and characterises market fluctuations as external factors impacting innovating development. Coefficients k_1, k_2 and c_1, c_2 define company’s awareness on each other’s innovations, as all companies in the

organisational field are connected by the existence of two types of information. The first type – confidential information on innovations. The second type – open-access information available on open internet sources. Coefficients k_1, k_2 determine the degree of confidential information possession. Coefficients c_1, c_2 display the degree of open-access information awareness.

The model shows that the dynamics of company's innovation indicators in a market cluster depends on similar parameters of an innovation leader. Within the suggested model, the innovative potential may be defined as a dynamic form of changes of the value proportional to the volume of innovation product sales, the volume of revenues from innovation product sales or other indicators obtained for a certain period of time, in relation to an innovation leader. The benchmarking model is set out in Figure 2.

Figure 2 A simplified benchmarking model



When comparing an innovation company with a leader, the given system of equations enables to identify and calculate current and forecasted discrepancy between innovative development indices. Thus, the system performs prognostic modelling of possible development paths according to three main scenarios: pessimistic, optimistic and optimal. The modelling allows to prepare recommendations on altering the indicators and excluding the factors that decrease the innovative potential and investment appeal of certain companies. The mechanisms of regional innovative development management are developed on the basis of the said recommendations.

Software, algorithmic and informational support is provided in the form of an informational and analytical platform intended to analyse the regional innovative potential and to forecast the innovative development. The platform represents a set of tools for searching, extracting, integrating and storing of data from open sources, prognostic modelling and benchmarking of innovation companies' interactions. All tools remain in informational and technical interaction and provide automation of big data processing in the system of convergent processing of data using cloud technologies. The operation of the informational and analytical platform requires implementation of certain

procedures of working with decision-making support tools in the course of companies and regions' innovative development management, which include the following tasks:

- a Analysis of current situation in a regional innovation system.
- b Search and collection of data on institutional indicators of innovation activity.
- c Selection of indicators and analysis of their sensitivity to effectiveness indices of innovative development.
- d Intellectual analysis of big data.
- e Stratification and clustering of innovation enterprises.
- f Selection of innovation leaders and synthesis of model indicator matrixes for clusters, strata and regions.
- g Benchmarking of innovation companies and leaders' indicator matrixes.
- h Prognostic modelling of the dynamics of changes of indicator and effectiveness indices.
- i Estimation of factors impacting innovative development processes in regions.
- j Synthesis of companies and regions' innovative development scenarios.
- k Comparative analysis of the degree of regional innovative development.
- l Selection of recommendations to achieve optimal values of innovation indices for companies in a cluster and a region in whole.

4 Conclusions

Models and methods of intellectual analysis of data for evaluation and prognostic modelling of the dynamics of companies and regions' innovative development are developed and implemented on the basis of the institutional concept, the multi-agent approach and the convergent technology of big data processing. The main result of this study is the development of the methodology for regional innovation development modelling and analysis by evaluating the innovative potential and investment appeal of companies. In the course of the research, there has been developed a new methodological approach involving an aggregate of models, methods and algorithms of searching, intellectual and comparative analysis of big data on innovation enterprises in regions. This approach makes it possible to obtain reliable information on potential capacities of regional innovation systems, as well as to develop real recommendations concerning innovative development of economic subjects and a region in whole. The introduction of the research results into the practice of regional management will interconnect innovative development system developers and users, identify their interaction, accumulate the experience of joint corporate projects, reduce transaction costs for economic subjects and potential investors. The practical importance is determined by the fact that the said models and methods are implemented in the form of an informational and analytical platform including tools of big data flow collection and processing, indicator indices analysis and evaluation in the institutional context. The platform is the main mechanism

of the cyber-social system of innovative development management with elements of artificial intelligence. It is capable of organising interaction of users and adapting to their needs, analysing current situations and elaborating decisions to increase effectiveness of interactions between innovation agents and to reduce transaction costs. The universal nature of the models and methods makes it possible to apply them to solve other problems, for example, to analyse company's competitiveness, trade capacity, appeal, productive potential, etc.

Thus, the authors in this article have developed the classical theoretical approaches to regional innovation development through elaboration of effective digital methods and tools for intellectual analysis. This will allow obtaining accurate data on innovation processes in the region in real time, to reveal hidden patterns and features in the development of regional innovation systems.

Acknowledgements

The reported study was funded by RFBR according to the research project no. 18-010-00204_a. The results of part 3 were obtained within the RSF grant (project no. 20-71-10087).

References

- Alamri, B.H. and Qureshi, M.R.J. (2015) 'Usability of cloud computing to improve higher education', *I.J. Information Technology and Computer Science*, Vol. 9, No. 1, pp.59–65.
- Batty, M., Axhausen, K.W., Giannotti, F., Pozdnoukhov, A., Wachowicz, M., Ouzounis, G. and Portuglai, Y. (2012) 'Smart cities for the future', *Innovation: The European Journal of Social Science Research*, Vol. 214, No. 1, pp.481–518.
- Benninga, S.Z. and Sarig, O.H. (1997) *Corporate Finance: A Valuation Approach*, McGraw-Hill, New York, 322pp.
- Benzi, R., Sutera, A. and Vulpiani, A. (1981) 'The mechanism of stochastic resonance', *Journal of Physics A: Mathematical and General*, Vol. 14, No. 11, p.453.
- Chesbrough, H. (2003) *Open Innovation: The New Imperative for Creating and Profiting from Technology*, Harvard Business School Press, Boston, MA.
- Dumbill, E. (2014) 'Defining big data', *Forbes* [online] <http://www.forbes.com/sites/edddumbill/2014/05/07/definingbig-data/> (accessed 14 February 2019).
- Forrester, D. (2006) *The World Dynamics*, Wright-Allen Press, Moscow, 284pp.
- Gamidullaeva, L. (2018) 'Towards combining the innovation ecosystem concept with intermediary approach to regional innovation development', *International Journal of Economics & Business Administration*, Vol. 6, No. 1, pp.39–53.
- Gamidullaeva, L., Vasin, S. and Surovitskaya, G. (2019) 'Towards combining the triple helix concept with competence-based approach of educational management theory', *Global Business and Economics Review*, Vol. 21, Nos. 3/4, DOI: 10.1504/GBER.2019.099393.
- Gamidullaeva, L.A. and Tolstykh, T.O. (2017) 'Transaction costs, institutions and regional innovation development: the case of Russia', *Proceedings of the 30th International Business Information Management Association Conference (IBIMA), Vision 2020: Sustainable Economic development, Innovation Management, and Global Growth*, Madrid, Spain, 8–9 November.
- Gloor, P. (2006) *Swarm Creativity: Competitive Advantage through Collaborative Innovation Networks*, Oxford University Press, Oxford.

- Hilbert, M. (2013) *Big Data for Development: From Information-to Knowledge Societies* [online] <https://irevolutions.org/2013/02/04/big-data-for-development-2/> (accessed 15 February 2019).
- Kayser, V., Nehrke, B. and Zubovic, D. (2018) 'Data science as an innovation challenge: from big data to value proposition', *Technology Innovation Management Review*, Vol. 8, No. 3, pp.16–25 [online] <http://doi.org/10.22215/timreview/1143>.
- Ley, B., Ogonowski, C., Mu, M., Hess, J., Race, N., Randall, D., Rouncefield, M. and Wulf, V. (2015) 'At home with users: a comparative view of living labs', *Interact. Comput.*, Vol. 27, No. 1, pp.21–35.
- Marshall, A. (1993) *Principles of Political Economy*, Progress, Moscow.
- Martin, A.L. and Thawabieh, F.A. (2017) 'The role of big data management and analytics in higher education', *Business, Management and Economics Research*, Vol. 3, No. 7, pp.85–91, Academic Research Publishing Group.
- Mkrttchian, V., Bershadsky, A., Finogeev, A., Berezin, A. and Potapova, I. (2017) 'Digital model of bench-marking for development of competitive advantage', in Isaías, P. and Carvalho, L. (Eds.): *User Innovation and the Entrepreneurship Phenomenon in the Digital Economy*, Chapter 14, pp.288–303, IGI Global, Hershey, PA, USA, DOI: 10.4018/978-1-5225-2826-5.
- North, D.C. (1991) 'Institutions', *The Journal of Economic Perspectives*, Vol. 5, No. 1, pp.97–112.
- Phillips, S.J. (2002) 'Acceleration of k-means and related clustering algorithms', *Workshop on Algorithm Engineering and Experimentation*, Springer, Berlin, Heidelberg, pp.166–177.
- Rae, A. and Singleton, A. (2015) 'Editorial – putting big data in its place: a regional studies and regional science perspective', *Regional Studies, Regional Science*, Vol. 2, No. 1, pp.1–5.
- Rothschild, M. (1990) *Bionomics. Economy as Ecosystem*, Henry Holt and Company Inc., NY.
- Saravanan, V., Hussain, F. and Kshirasagar, N. (2019) 'Role of big data in internet of things networks', in Kaur, G. and Tomar, P. (Eds.): *Handbook of Research on Big Data and the IoT*, pp.273–299, IGI Global, Hershey, PA, DOI: 10.4018/978-1-5225-7432-3.ch016.
- Schintler, L. and Chen, Z. (2018) *Big Data for Regional Science*, Schintler, L.A. and Chen, Z. (Eds.), Taylor & Francis Group.
- Sheth, A., Anantharam, P. and Henson, C. (2013) 'Physical-cyber-social computing: an early 21st century approach', *IEEE Intelligent Systems*, Vol. 28, No. 1, pp.78–82.
- Shmid, A. et al. (2016) *New Methods to Work with Big Data: Winning Strategies of Management in Business Analytics*, Palmir, 528pp, Moscow.
- Tafti, F., Abdolvand, N. and Harandi, S. (2019) 'A strategic alignment model for collaborative open innovation networks', *International Journal of Business Innovation and Research*, Vol. 19, No. 1, pp.1–28.
- Teodorich, K.F. (1952) *Self-oscillating Systems*, GosTechIzdat, Moscow, 272pp.
- Vasin, S., Gamidullaeva, L., Shkarupeta, E., Palatkin, I. and Vasina, T. (2018) 'Emerging trends and opportunities for Industry 4.0 development in Russia', *European Research Studies Journal*, Vol. 21, No. 3, pp.63–76 [online] https://www.ersj.eu/index.php?option=com_content&task=view&id=1336.
- Vasin, S.M., Gamidullaeva, L.A., Wise, N. et al. (2019) 'Knowledge exchange and the trust institution: a new look at the problem', *J. Knowl. Econ.* [online] <https://doi.org/10.1007/s13132-019-00588-2>.