CLUSTERING ELECTRIC POWER SYSTEMS INTO ADEQUACY ASSESSMENT AREAS

Dmitry Krupenev[1], Denis Boyarkin[1], Nikolay Belyaev[2]

•

[1] Melentiev Energy Systems Institute, Russia.
[2] JSC «STC UPS», Russia.
krupenev@isem.irk.ru
denisboyarkin@isem.irk.ru
belyaev.na@yandex.ru

Abstract

The article discusses the issue of clustering electric power systems (EPS) into adequacy assessment areas to form equivalent models (EM) of power systems, with such models designed to calculate adequacy metrics. The level of detail of equivalent models is backed by the consideration of two properties of power systems being reliability and economic feasibility. To solve the problem posed above, the paper proposes to use a number of engineering approaches that take into account reliability and economic feasibility of power systems. The engineering approaches are supplemented with analytical methods. We present an analysis of analytical clustering methods as applied to the task of forming power system EMs. We review well-established clustering methods, such as k-means and the shortest path clustering, as well as methods of community detection in graphs by specified features, which prove the most suitable for solving the problem. Based on our analysis, we have found that for clustering of power system structure graphs, the most plausible result can be obtained by applying the Leiden method. The Experimental Study section presents the results of applying the methods we analyze here to form an EM of the test electric power system.

Keywords: electric power system, adequacy, assessment area, equivalent model, clustering, Leiden algorithm

I. Introduction

Electric power systems (EPS) are complex structures consisting of hundreds of thousands of different elements [1]. To analyze the balance reliability of such systems [2-4], as a rule, a technique based on the Monte Carlo method [5] is used, which is the most computationally efficient compared to others [6]. In adequacy assessment, due to the high dimensionality of the analyzed EPS and the resulting computational and analytical difficulties, EPS is clustered into reliability zones (RZ) and energy calculation models (ECM) are formed. In the practice of solving system analysis problems, various clustering methods have been developed [7,8], but there are no formalized clustering methods for adequacy assessment. When clustering for certain parts of the EPS, network restrictions that have little effect on the adequacy indicators are excluded. ECM formation is usually carried out before adequacy assessment, and the formed ECM is accepted as unchanged for the entire adequacy assessment period, but if necessary and computationally possible, EPS clustering into reliability zones can be carried out multiple times in a adequacy assessment cycle. The main reason for multiple clustering may be significant changes in the EPS parameters during the calculation period and, accordingly, the impact of these parameters on the

grid's capacity to transmit power. This may also be due to the fact that modern EPS have a complex structure with a wide variety of power equipment, in addition, EPS tend to enlarge and merge, which further complicates the structure and uncertainty in possible power flows.

In the practice of adequacy assessment, there are various approaches to EPS clustering. They are mainly based either on expert knowledge or on a simplified analysis of the balance situation in the EPS. The purpose of this article is to present a methodology for EPS clustering on the RZ, which has a mathematical formalization. The substantive problem to be solved can be formulated as follows: for a known EPS structure, characteristics of generating capacities and power consumption schedules, limitations on the capacity of transmission lines and electrical grid sections, organizational and economic conditions for the functioning of the EPS, it is necessary to determine the RZ boundaries for the formation of the ECM and further adequacy assessment.

II. Statement of the problem of clustering of EPS and formation of ECM based on methods of detecting communities in graphs

For the most efficient (fast and accurate) solution to the problem of clustering the EPS on the ZN and forming the ECM, it is necessary to develop an approach that will most fully reflect the influencing criteria and restrictions on the level of the EPS adequacy and will allow for the sufficiently fast and adaptive clustering of the EPS in the process of conducting research.

When building an algorithm of power system clustering into assessment areas, we need to obtain the function a:X→Z that assigns each node of the power system ($x \in X$) into an assessment area $z \in Z$. In the considered case, the set Z, i.e., the number of assessment areas, is unknown. The optimal number of assessment areas needs to be determined as part of the calculation process. It is worth noting that, as is always the case with clustering, Kleinberg's impossibility theorem holds for the problem at hand, which states that there is no optimal clustering algorithm. Each algorithm applied will introduce its own bias into the final solution. This will depend on the selected criterion and clustering metric.

The solution to the problem of clustering the EPS will be its division, satisfying the given optimality criterion. In the case under consideration, the optimality criterion is the minimization of the number of reliability zones:

$$\mathrm{Z} \underset{\rho}{\rightarrow} min. \tag{1}$$

The desire to minimize the number of reliability zones is natural, since this will lead to a decrease in the dimension of the ERM in the further analysis of balance reliability.

A limitation in solving the problem will be a limitation specified for the metric, which is formed from various features. In the case under consideration, the clustering metric should include both technical and economic features. Technical features include: installed capacity of generating equipment in EPS nodes; average value of the number of hours of use of installed capacity in EPS nodes; seasonal maximums and minimums of power consumption in EPS nodes; accident rate of power equipment; standards for scheduled repairs of power equipment; lengths of power transmission lines (PTL) between EPS nodes; transmission capacities of PTLs [9]. Economic features include: cost of commissioning generating and network facilities. The result of clustering will ultimately depend on the choice of metric. The metric should maximally reflect the specifics of the EPS operation and the influence of the taken into account features on clustering. One of the following metrics can be used as a measure of distance (degree of similarity): Euclidean distance or its square, Manhattan distance, Chebyshev distance, power distance, etc.

III. Overview of clustering algorithms for systems

As it stands now, quite a large number of algorithms of dataset clustering have been developed [10-12], each of which has its own unique features, advantages and shortcomings. The methods and adaptations of methods that allow working in a network of interconnected components, in other words graphs, stand out from the rest. The presence of relationships between the objects being clustered significantly affects the obtained result and imposes certain constraints. There are many graph-based clustering methods that can be used as part of network analysis. Some of them are modifications of established clustering methods such as k-means and spectral clustering and have adaptations that make them work with graphs.

Classification of clustering methods can be done based on the way the graph partition boundary is detected. The following classes of these methods can be distinguished:

1. Similarity-based methods. These methods detect clusters based on the degree of similarity between graph vertices. Such methods may use different similarity metrics, such as cosine similarity or Euclidean distance.

2. Distance-based methods. These methods detect clusters based on the distance between vertices of the graph. We may use different distance metrics for different methods for example Manhattan distance or Euclidean distance

3. Model-based methods. These methods detect clusters based on a probabilistic graph model. Such methods can be used to cluster graphs with a large number of vertices and/or edges.

4. Graph structure-based methods. These methods detect clusters based on the graph structure such as graph density, vertex centrality, etc.

Eeach class of the methods presented above has its own advantages and downsides, and the choice of a particular algorithm depends on the problem to be solved and the properties of the graph to be clustered.

Also, clustering methods can be classified on the basis of the approaches underpinning the clustering algorithm. Some of the most common methods for clustering on graphs include:

1. Similarity-based methods. These methods detect clusters based on the degree of similarity between graph vertices. Such methods may use different similarity metrics, such as cosine similarity or Euclidean distance.

1. The class of partitioning methods.

1.1. The k-means algorithm [11] is one of the most popular clustering methods belonging to the class of partitioning methods. It is based on dividing the vertices of the graph into a given number of clusters, so as to minimize the sum of squares of distances between vertices and centroids of clusters. The k-means algorithm can be applied when vertex properties are numeric or when it is possible to convert them to numeric.

1.2. The Shortest Path Clustering [12] builds clusters based on the distances between vertices of the graph. This method also belongs to the class of partitioning methods.

2. Hierarchical clustering methods [13] are a class of methods that enable creation of a hierarchical structure of clusters. These methods can be used to cluster graph vertices based on their properties by hierarchically grouping nearby vertices. The general idea behind the methods of this group is a sequential hierarchical decomposition of a set of objects. Depending on the direction of hierarchy construction the methods fall into two categories: divisive and agglomerative clustering. In the case of the agglomerative approach, the decomposition process starts with each object being an independent cluster. Then, at each iteration, pairs of nearby clusters are successively grouped into a common cluster. Iterations continue until all objects are grouped into a single cluster or until some stopping condition is met. Also, we have the divisive approach. The approach is contained all objects into a single cluster in the initial stage. Further

from single cluster the method form to smaller clusters while each object is in a separate cluster or the stopping condition is met. In what follows, we will review some of them.

2.1. The Louvain method [14] has graph vertices gradually clustered together to maximize connectivity inside clusters and minimize connectivity between clusters.

2.2. Spectral Clustering [15] is based on analysing the eigenvalues of the graph's adjacency matrix and allows partitioning the graph into clusters based on its structural information.

2.3. The Edge Betweenness Clustering [16] uses a measure of connectivity between clusters (between center vertices) to partition the graph into clusters.

3. Density-based clustering methods are based on analysing the density distribution of graph nodes in the feature space. They can be particularly useful if vertices of the graph are not numeric.

4. Network analysis methods [17] can also be used for clustering of graphs. Network methods are an approach to data analysis that are based on representing data as networks or graphs, which has an identical structure. The main direction the methods consist that the object space is separate into a cells forming finite number a network structure. All clustering operations are performed in the structure. The main advantage of methods of this group is the short execution time, which usually does not depend on the number of data objects.

4.1. The CLIQUE algorithm [18], adapted to clustering of high dimensional data, is one of the classical network algorithms. The method is based on the assumption that if in a multidimensional data space, the distribution of objects is not uniform (there are dense and sparse regions), then the projection of the dense region into a subspace of lower dimensionality will be a part of the dense region in this subspace.

4.2. Infomap method [19]. The method relies on information theory to determine the most likely communities in a graph.

We will focus specifically on the modularity-based algorithm [20], which maximizes the difference between the actual number of edges in a set and the expected number of such edges; various modifications and improvements of modularity based on heuristic methods include the annealing method [21], spectral algorithms [15], etc. [22]. One of the most commonly used algorithms for community detection in graphs is the Louvain algorithm [14]. As a result of our analysis of clustering methods, it was found that for the problem to be solved, which is clustering of power system into adequacy assessment areas, the most suitable is the Leiden algorithm [23].

IV. Clustering of power system into assessment areas based on the Leiden algorithm

The Leiden algorithm is an algorithm for community detection in large networks. The Leiden method is an iterative algorithm that works as follows (Fig. 1 visualizes the operation of the Leiden algorithm):

1. Initially, each node is considered as a separate cluster.
2. The algorithm sequentially aggregates nearby clusters by iteratively optimizing the clustering quality score.
3. The clustering quality score in the Leiden method is based on the distribution of objects between clusters and shows how similar objects within a cluster are to each other and how different are objects from different clusters.
4. To aggregate clusters, the Leiden method uses the so-called "modularity gain", which shows how much the clustering quality will improve when two clusters are aggregated.
5. Clusters are aggregated as long as further aggregation does not lead to degradation of the clustering quality score.
6. After all clusters are aggregated, the result of the algorithm is the final partitioning of objects into clusters.

The Leiden method is greedy in that it selects a locally optimal solution at each step. It is also relatively fast and performs well when clustering large graphs and networks.

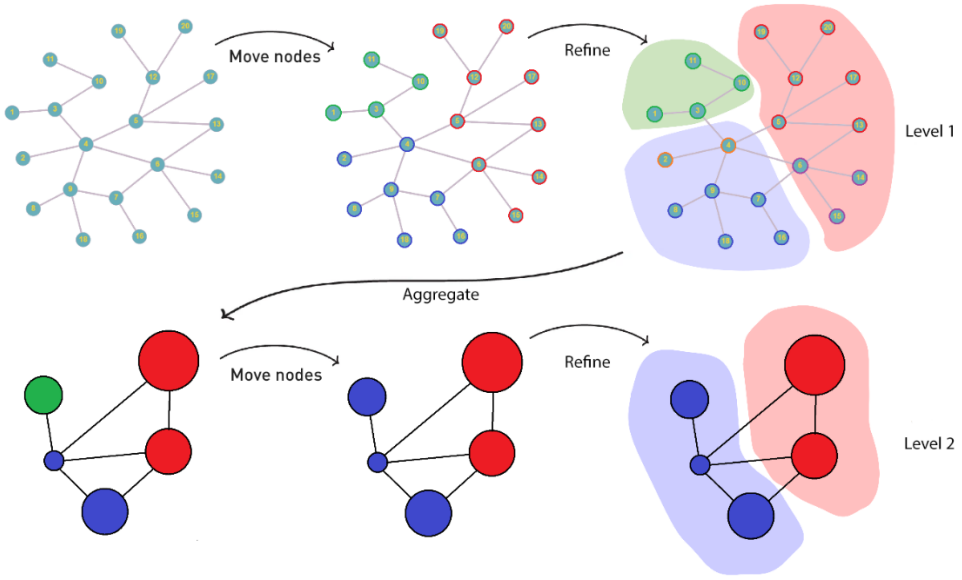

Figure 1: *A visualization of the operation of the Leiden algorithm.*

More formally speaking, the algorithm divides nodes into non-overlapping clusters to maximize the modularity score for each cluster. Modularity quantifies the quality of node assignment to clusters, i.e., how tightly connected the nodes in a cluster are compared to how connected they would be in a random network. The Leiden algorithm performs hierarchical clustering by greedily optimizing modularity, and the process is repeated in the compressed graph. The Leiden algorithm consists of three main steps:

1) Local moving of nodes into clusters based on the modularity score (metric), which is defined as:

$$Q = \frac{1}{2m}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})\delta(c_i, c_j)\,; \tag{2}$$

where: A_{ij} - edge weight formed according to the features used; k_i and k_j - sum of weights of edges attached to nodes i and j; m - sum of all edge weights in the graph; c_i and c_j - clusters; δ - the Kronecker delta.

$$\delta(c_i, c_j) = \begin{cases} 1, c_i = c_j \\ 0, c_i \neq c_j \end{cases}. \tag{3}$$

The following expression is used to determine the modularity of the cluster c:

$$Q_c = \frac{\sum in}{2m} - (\frac{\sum tot}{2m})^2, \tag{4}$$

where: $\sum in$ - the sum of edge weights between nodes within the cluster c (each edge is counted twice); $\sum tot$ - the sum of all edge weights for nodes within the cluster.

In the iterative process, nodes are moved from one cluster to another. For each option of moving, the difference of modularities ΔQ is calculated:

$$\delta(c_i, c_j) = \begin{cases} 1, c_i = c_j \\ 0, c_i \neq c_j \end{cases}. \quad (5)$$

Based on the best result, the corresponding node is assigned to the corresponding cluster.

2) Cluster refinement, namely the identification of the clusters proposed at the first step. The clusters proposed at the first step can be divided into several clusters. The refinement phase does not follow a greedy approach and can aggregate a node with a randomly selected cluster, which increases the quality (modularity) score. This randomness allows the cluster space to be opened up more widely.

3) Aggregation and repetition of steps 1 and 2 until the quality of moving and aggregation of nodes cannot be improved.

V. Experimental studies

To evaluate the performance of the Leiden algorithm, experimental studies were performed on a test circuit of the Unified Power System (UPS) of Siberia, which includes the Altai, Buryat, Irkutsk, Krasnoyarsk, Kuzbass, Novosibirsk, Omsk, Tomsk, Khakassia and Transbaikal energy systems. For subsequent clustering, the UPS of Siberia was presented as an undirected graph, in which each of the vertices is characterized by maximum and minimum generation and load, the lines are characterized by length and voltage. The initial data for clustering the UPS of Siberia were adopted in accordance with [24]. Initial circuit of the UPS of Siberia during clustering consisted of 540 generating and loading nodes and 1025 power transmission lines.

Figure 2 shows the result of the Leiden clustering method (algorithm) in the form of a graphical representation of the division of the nodes of the Siberian UPS into reliability zones.

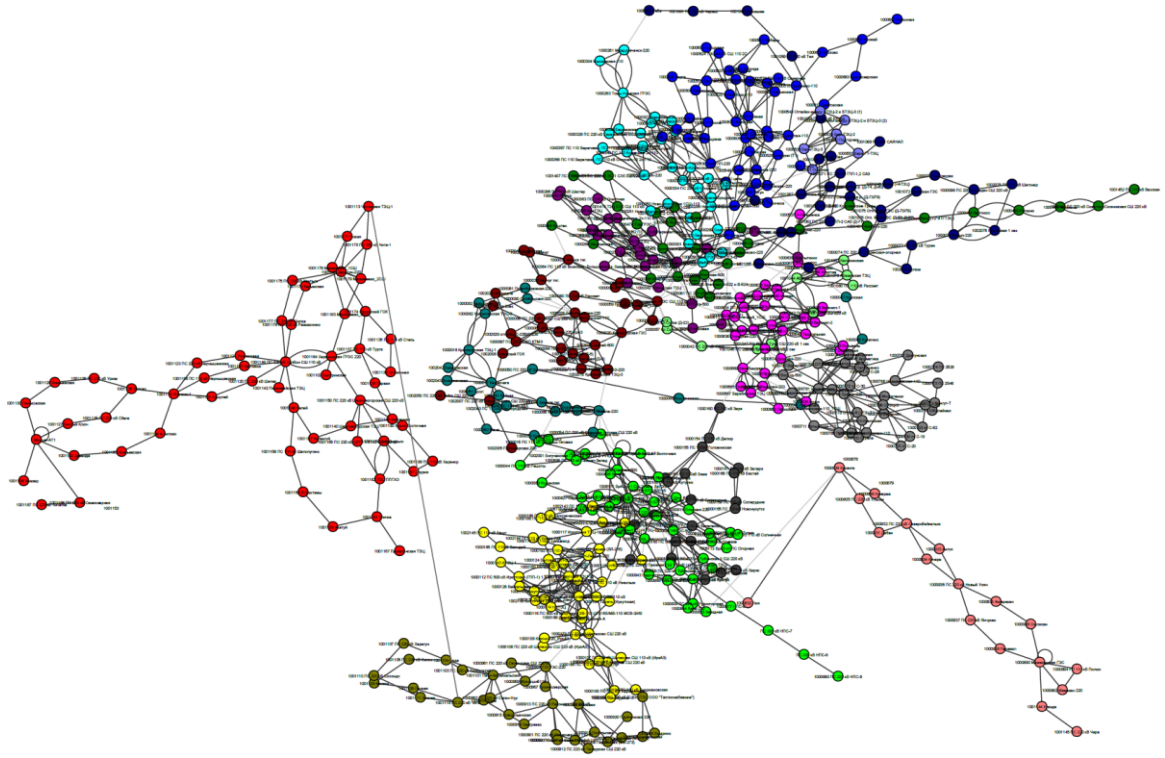

Figure 2: *Graphical representation of assignment of power system nodes to assessment areas based on the application of the Leiden algorithm.*

Figure 3 presents the same clustering but mapped to the power system circuit.

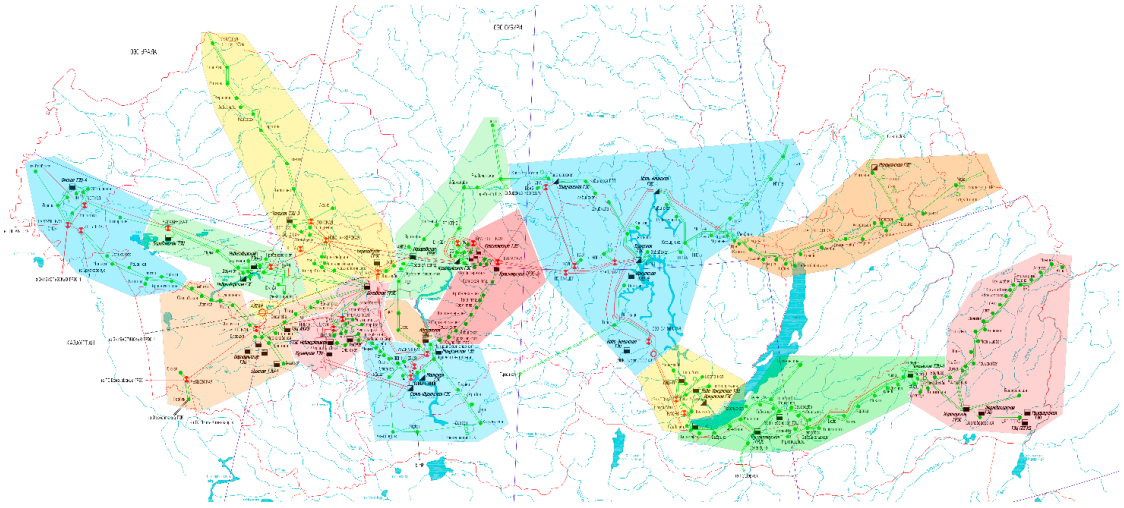

Figure 3: *The result of clustering the Siberian UPS into reliability zones based on the application of the Leiden algorithm (geographical referencing).*

As a result of clustering using the Leiden method, 14 reliability zones were obtained. The time spent on clustering was 1.16 s, which is sufficient for clustering in the process of assessing adequacy with the preparatory work with the initial data.

Analysis of the obtained clusters shows that the algorithm performed clustering by the weakest points of the EPS in terms of the density of interrelations and their quality. Therefore, the obtained reliability zones turned out to be maximally independent with the smallest number of connections between each other. Such a division in the subsequent assessment of adequacy will well determine the workload of connections, which were already identified at the clustering stage as bottlenecks.

VI. Conclusion

The presented study considers the problem of forming energy calculation models that are used in electric power systems adequacy assessment. An analysis of clustering methods on graphs was carried out and it was determined that one of the most suitable methods for clustering EPS into reliability zones is the Leiden algorithm.

Based on the results of the experimental calculations, it can be noted that in comparison with known engineering approaches to the formation of energy calculation models, the proposed approach gives a close result. Figure 3 shows that, based on the results of forming reliability zones, all the main restrictions on power transmission in the UPS of Siberia were taken into account. The boundaries of the formed reliability zones correlate with the corresponding controlled sections in the electric network, despite the fact that the controlled sections were not taken into account in any way in the proposed approach.

Thus, it can be concluded that with the help of the Leiden algorithm, it becomes possible to formalize the definition of reliability zones directly in the cycles of performing the corresponding calculations, which allows optimizing the solution of problems of analysis and synthesis of adequacy.

Funding

The research was supported by RSF (project No. 23-71-01090).

References

[1] Rudenko Y.N., Cheltsov M.B. Reliability and redundancy in electric power systems. "Nauka", Siberian branch, Novosibirsk. 1974.

[2] Kovalev G.F., Krupenev D.S., Lebedeva L.M. System reliability of the UES of Russia at the level of 2030. *Power stations*. 2011. No. 2. P. 44-47.

[3] Probabilistic Adequacy and Measures. Technical Reference Report Final, NERC, July, 2018.

[4] 2021 Long-Term Reliability Assessment. NERC. 2021.

[5] Kovalev G.F., Lebedeva L.M. Reliability of Power Systems. Springer-Verlag GmbH, 2019. DOI: 10.1007/978-3-030-18736-1.

[6] Krupenev D, Boyarkin D, Iakubovskii D. Improvement in the computational efficiency of a technique for assessing the reliability of electric power systems based on the Monte Carlo method. *Reliability Engineering & System Safety*. № 204. 2020. doi:10.1016/j.ress.2020.107171

[7] M. Rouhani, M. Mohammadi, M. Aiello. Soft clustering based probabilistic power flow with correlated inter temporal events. *Electric Power Systems Research*. № 204. 2022. 107677. https://doi.org/10.1016/j.epsr.2021.107677.

[8] Gheorghe G., Scarlatache F., Neagu B. Clustering in Power Systems. Applications. 2016.

[9] D.L. Faibisovich. Reference book on the design of electrical networks. Ed. 4th, revised. and additional M.: Publishing house of NTs ENAS. 2012.

[10] Mandel I. D. Cluster analysis. M.: Finance and statistics. 1988.

[11] MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proc. 5th Berkeley Symp. On Math. Statistics and Probability, 1967.

[12] A. Sarkar, S. Ramalingam. Graph Clustering: A Review. 2020.

[13] Kaufman, L.; Rousseeuw, P.J. Finding Groups in Data: An Introduction to Cluster Analysis (1 ed.). New York: John Wiley, 1990.

[14] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 10008. 6. 2008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

[15] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. Phys. Rev. E 69. 026113. 2004.

[16] Ester, M., H. P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, OR, AAAI Press. P. 226-231. 1996.

[17] F. Beligianni, G. Tsatsaronis, T. Palpanas. A Survey on Clustering Techniques for Graph Structured Data. 2019.

[18] Palla, G., Derényi, I., Farkas, I. et al. Uncovering the overlapping community structure of complex networks in nature and society. *Nature,* 435. P. 814–818. 2005. https://doi.org/10.1038/nature03607

[19] M. Rosvall, C. T. Bergstrom. Maps of information flow reveal community structure in complex networks. *PNAS*. 105. 1118. 2008.

[20] Orlov A.O., Chepovsky A.A. About modularity properties and actual adjustments of the blondel algorithm. *Vestnik NSU. Series: Information technologies*. №3. 2017.

[21] Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., and Teller E. Equation of State Calculations by Fast Computer Machines. *Chemical Physics,* 21. 6. June. 1953. P. 1087-1092.

[22] Lancichinetti, A., Fortunato, S. Community detection algorithms: A comparative analysis. *Phys. Rev.* E 80, 056117, 2009. https://doi.org/10.1103/PhysRevE.80.056117.

[23] Traag, V.A., Waltman, L., van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 9. 5233. 2019. https://doi.org/10.1038/s41598-019-41695-z.

[24] Order of the Ministry of Energy of Russia dated February 28, 2022 No. 146 "On approval of the scheme and program for the development of the Unified Energy System of Russia for 2022-2028".