DOI 10.29042/2019-5607-5614

Helix Vol. 9 (5): 5607- 5614

Application of Spectral Clustering Methods in Pipeline Systems Graph Models

[1]Gul'naz I. Galimova, [2]Il'nur D. Galimyanov, [3]Dinar T. Yakupov, [4]Vladimir V. Mokshin
[1,2] Kazan Federal University, Naberezhnye Chelny Institute
[3,4] Kazan National Research Technical University named after A.N.Tupolev – KAI
Email: plosatik1208@mail.ru

Received: 20th August 2019, Accepted: 30th September 2019, Published: 31st October 2019

Abstract
Among the many tasks on graphs, studies related to partitioning an initial graph into a predetermined number of connected disjoint components have found wide practical application - graph clustering, for example, is used in computer networks, transport, pattern recognition, and in many other areas. The decomposition methods of graph structures make a significant contribution to the performance of search algorithms, which is especially important in the context of restrictions on computing and time resources. And here we should pay special attention to the class of spectral clustering methods that combine elements of graph theory and linear algebra. In this paper, we consider the basic principles of the theory of spectral clustering, describe the main approach of normalized spectral clustering of graphs. Decomposition of any graph as a structure with inherent topology meets the criteria for optimality in connectedness and balanced subgraphs with a small number of clusters. With an increase in the number of sub-areas above a certain value, the probability of appearance of the disconnected subgraphs in the decomposition structure increases. To solve this problem, the authors propose an algorithm for the priority distribution of nodes based on iterative transfer of nodes of isolated areas to the most priority neighboring subgraphs. It is considered the question of optimal placement of objects in graph models of hydraulic networks by methods based on trial and error algorithms, greedy and spectral clustering.

Keywords
Graph, Spectral Clustering, Eigenvector, Pipeline Systems, Modeling.

Introduction

The piping system is a complex structure distributed in space. Its functioning is carried out in the conditions of information incompleteness and on the basis of a probabilistic assessment of its real state [1]. Thus, the management effectiveness of the pipeline system depends on the estimates of certain parameters that affect its functioning. The reliability of these estimates, in turn, depends on the number and location of objects. This paper discusses alternative solutions.
The use of greedy algorithm [2] is based on iterative expansion of the initial sub-areas by including vertices adjacent to the node selected as the starting one [3].
Kernigan and Lin [4] proposed a method of phased transformations of sub-areas of the initial partition: a small number of vertices are exchanged at each iteration between subgraphs. The particle pairs move from one sub-area to another, provided that the maximum improvement in the partition quality is achieved [5].
The class of spectral decomposition methods [6–9] assumes a fundamentally different formulation of the optimal partitioning problem and combines elements of graph theory and linear algebra.
Fedler [10] demonstrated that the eigenvector provides a solution to the problem of bipartite partitioning of the graph. In [11], the authors use the Fedler's conclusions in solving the dicot partitioning problem and introduce the normalized cut criterion (*NCut*). Ng, Jordan, Weis [12] described in the article the spectral clustering algorithm based on the symmetrically normalized discrete Laplace operator. The development and application of the theory of spectral clustering are also considered in [13-17].

Methods

1. Research Problem Statement

There is a graph G, whose nodes are characterized by estimates E_i of the determinism of a certain parameter, and the edges - by a combination of significant features P_j. After installing the next control element, the estimates are recalculated according to the formulas:

$$E_S = 1, \quad (1)$$

$$E_i = \left(E_{i-1} \cdot f\left(P_{i,i-1}\right)\right), \quad (2)$$

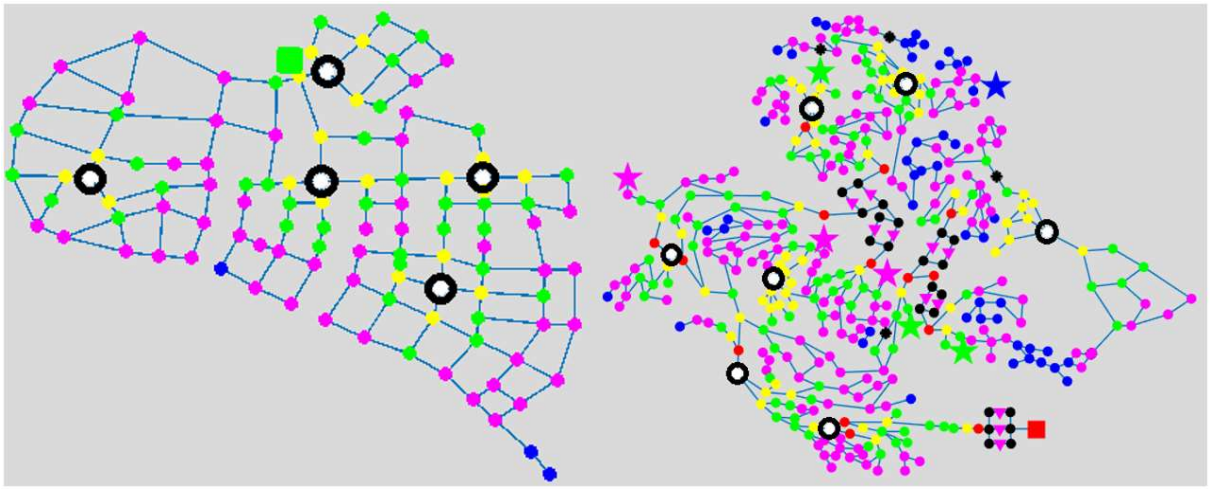
where E_S - assessment of the determinism of the value of considered parameter in the installation node of the control element, E_{i-1} - assessment of the determinism of the value of the considered parameter of the neighbor node, $f\left(P_{i,i-1}\right)$ – a function that characterizes the change in the estimate of the considered parameter depending on the features of pipeline section to the neighboring node.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

There is a certain number of k control elements; it is necessary to find such an arrangement of these elements in nodes that provides a minimum of the average value of the estimation of the target parameter uncertainty in the network:

$$F = 1 - mean(E) \rightarrow min. \quad (3)$$

Solving the problem of finding the best option for placing objects in k nodes for a network with the number of vertices N by exhaustive search method requires $\frac{N!}{k!(N-k)!}$ iterations. Figure 1 shows the search results for such vertices by the exponential trial and error method: 5 for ZJ network (114 nodes and 164 arcs) and 7 - for D-Town (407 nodes and 459 arcs).



a. **b.**

Fig. 1: Networks of Hydraulic Models with Highlighting the Objects Placement Using Trial and Error Method a. ZJ-Network; b. D-Town-Network

2. Theory of Spectral Clustering

Spectral methods are based on the use of eigenvalue properties and discrete Laplace operator vectors (Kirchhoff matrix) [10, 18, 19].

Eigenvectors contain information about the graph structure.

1. The main eigenvector corresponds to the largest eigenvalue of the adjacency matrix of connected graph. [20] The main eigenvector is used, for example, in search engines [21-22].
2. Fedler eigenvector - corresponds to the second smallest eigenvalue of the Kirchhoff matrix of connected graph.
3. The first k of eigenvectors - corresponds to the first k of the smallest eigenvalues of the Kirchhoff matrix of connected graph.

The Shi-Malik method (Shi & Malik, SM) [11] is one of the basic and widely used approaches of spectral clustering and includes the following steps:

1. Formation of the adjacency matrix W. This matrix is a way of representing graph G with many nodes V and many edges E in the form of square symmetric matrix.
2. Formation of the matrix D - matrix.
3. Calculation of the non-normalized Kirchhoff (Laplace) matrix:

$$L = D - W. \quad (4)$$

4. Finding the Kirchhoff (Laplace) matrix L_{RW}, normalized by the random walk method:

$$L_{RW} = D^{-1}L = I - D^{-1}W, \quad (5)$$

where I – single matrix.

5. Formation of the matrix U using the first k of eigenvectors corresponding to the first k of the smallest eigenvalues of the normalized Kirchhoff matrix.
6. Division of the set of nodes into k clusters using classical clustering methods.

3. Subgraph Connectivity Problem

Figure 2 shows the first 10 eigenvalues of the normalized Laplace matrix for the graph of the D-Town network of the EPANET hydraulic simulation system.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

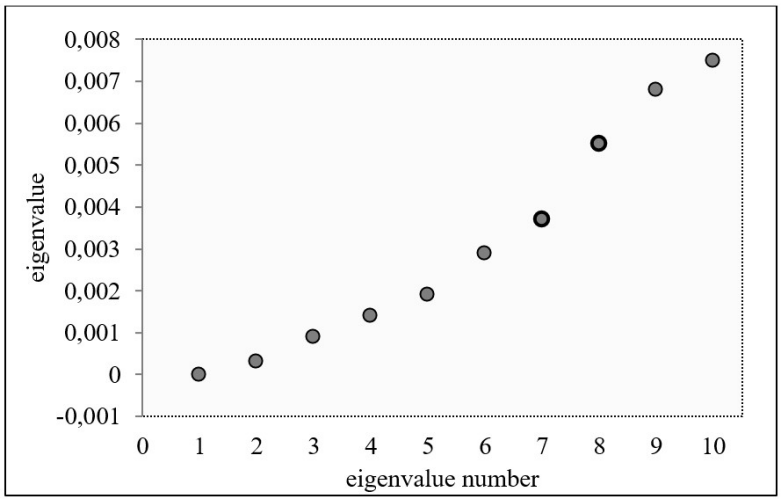


Fig. 2: Eigenvalues of the Normalized Laplace Matrix

Let us consider the diagram (Figure 3) of the values of the elements of the Fedler eigenvector for the graph in question.

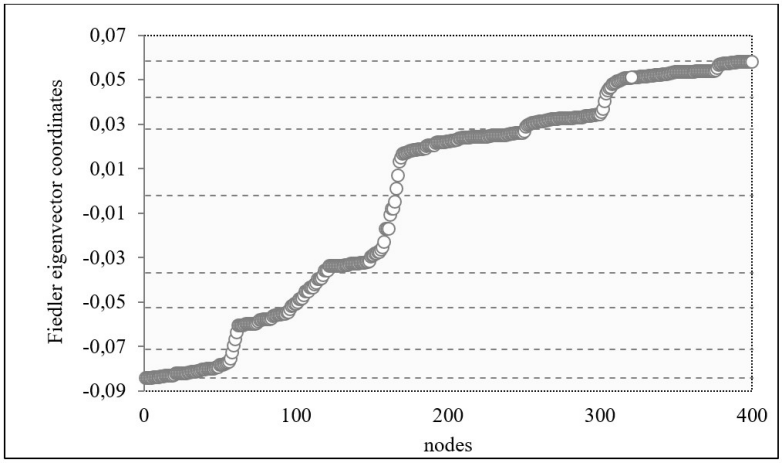


Fig. 3: Fiedler Eigenvector Coordinates (7 Subgraphs)

Figure 4 shows the values of the elements of the Fedler eigenvector.

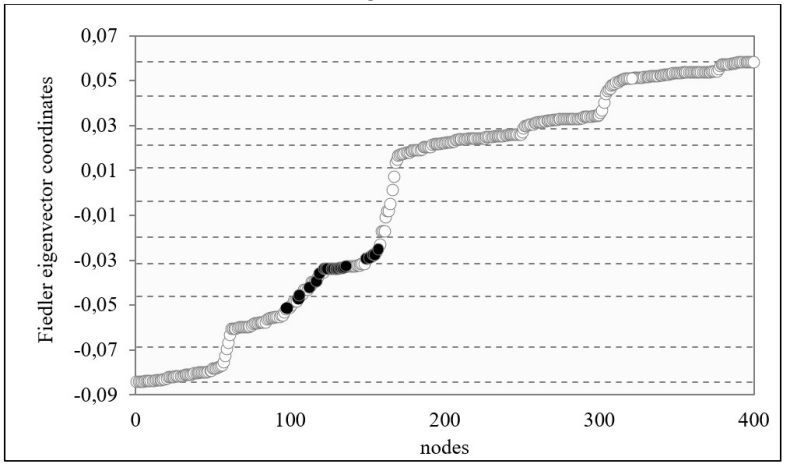


Fig. 4: Fiedler Eigenvector Coordinates (10 Subgraphs)

Figure 5 shows the pipeline network represented by the original graph model, divided into 10 subgraphs. Multi-colored sections correspond to different subgraphs, white nodes are the central points of the subgraphs.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

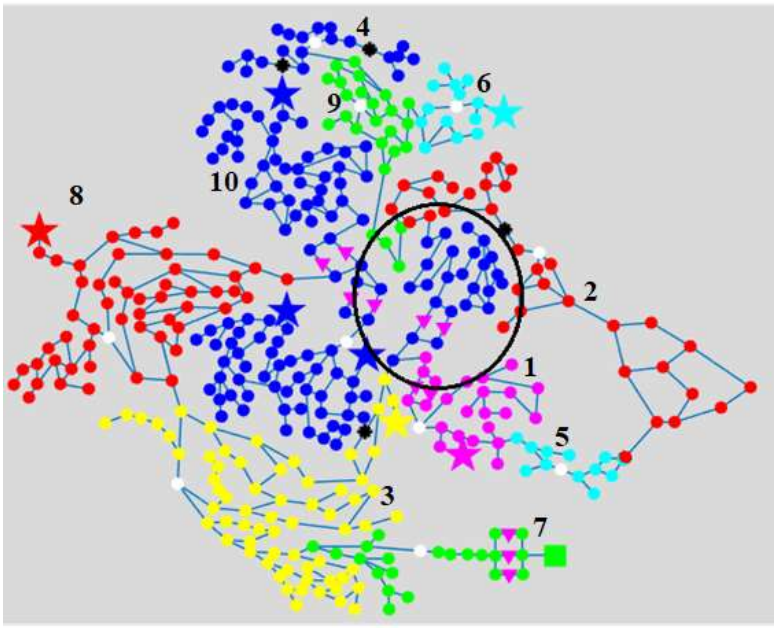


Fig. 5: The Graph Divided into 10 Sub-Areas. The "Cut Off" Area is Highlighted

To solve the connectivity problem of subgraphs, it is proposed an algorithm for the priority distribution of nodes
Input: *graph G (V, E),* the number of sub-areas *k*
Exit: *k* of connected subgraphs.
Steps:
1. A standard graph spectral clustering procedure is carried out.
2. Subgraphs are formed based on the results obtained in the previous step.
3. Checking the subgraph for connectivity.
4. Definition of the boundary node.
5. Determination of neighboring nodes that do not belong to the current subgraph.
6. Determination of the subgraphs of neighboring vertices.
7. Determination of the subgraph with the closest centroid to the boundary node under consideration.
8. Transfer of this boundary node to a subgraph.
9. Passing to Step 2.
10. If the subgraph is last - exit, otherwise - passing to the next subgraph (Step 3).

A graph with transformations according to the described algorithm is shown in Figure 6.

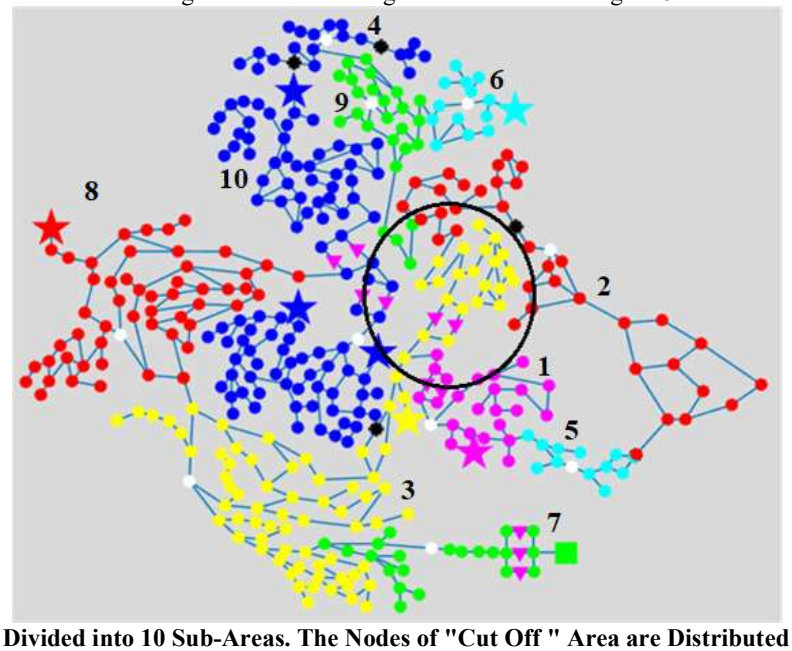


Fig. 6: The Graph Divided into 10 Sub-Areas. The Nodes of "Cut Off " Area are Distributed between Neighbor Subgraphs

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

The vertices of the area selected in Figure 5 are distributed between adjacent subgraphs in accordance with the minimum distance from the eigenvector components to the centroids of the clusters for each vertex.

4. Comparison algorithms and criteria

The work considers alternative solutions: trial and error method, greedy algorithm, a method based on spectral clustering.

Trial and error (TE) algorithm:

1. It is selected the node in the network, and it is installed the control element. The uncertainty estimates of the target parameter value and the objective function are recalculated.
2. Repetition of Step 1 for each network node.
3. Selecting a node, installing a control element, in which a minimum of objective function is provided.
4. Repetition of Steps 1-3 *k* times.

Greedy algorithm (Gr):

1. It is selected the node, having a particular property, and it is installed the control element. The uncertainty estimates and the objective function are recalculated.
2. Repetition of Step 1 *k* times.

Spectral clustering (SC) algorithm:

1. The network is divided into *k* parts.
2. A node is selected in each cluster. The uncertainty estimates and the objective function are recalculated.

To assess the effectiveness of the algorithms, the following criteria are applied: a) the minimum number k_{min} of control elements to achieve a given value of the objective function, b) the number of iterations $Iter$ of the calculation of objective function when solving the problem.

Results and Discussion

The problem considered within the framework of this work was solved with respect to the placement of pressure sensors in the water supply network of a settlement, the nodes (consumers) of which are characterized by the estimates of the pressure determinism, and the ribs (pipelines) - by lengths L_j. After installing another sensor in the network, the determinism estimates are recalculated according to the formula:

$$E_S = 1 \quad (6)$$

$$E_i = \max(E_{i-1} \cdot \alpha_{resist.} \cdot \alpha^2_{consum.} \cdot f(L_{i,i-1})), \quad (7)$$

where E_S - assessment of the determinism of the pressure value in the sensor installation unit, E_{i-1} - assessment of the determinism of the pressure value of a neighbor node, $\alpha_{resist.}$ – estimation of the error in determining the pipeline resistivity, $\alpha_{consum.}$ - estimation of the error in determining the water consumption value, $f(L_{i,i-1})$ – function from the pipeline length to the neighboring node.

If these parameters change, the final solution may change.

As a property of nodes, according to which it is decided to install sensors in a greedy algorithm and an algorithm based on spectral clustering, the following are used in the work:

a) assessment of pressure uncertainty in the node:

$$nGW = max[f(E_i)], \quad (8)$$

b) topological feature of the node:

$$GnW = max[f(I_i)], \quad (9)$$

- the node with the highest function significance value is selected as the sensor installation location.

c) combined feature:

$$GW = max[f(E_i) * f(I_i)], \quad (10)$$

The algorithms were tested on two water supply networks - ZJ and D-Town of the EPANET hydraulic simulation system. ZJ - a network without pumps, with 114 nodes and 164 pipes; there are no installed sensors at the initial moment. D-Town - a network with 11 pumps, 407 nodes and 459 pipes, the initial estimates of pressure determinism are calculated based on the fact that the sensors are already installed in the pumping stations. For the ZJ network, it is considered the sensor installation options in the amount from 0 to 10, for the D-Town network - from 0 to 20, the values $\alpha_{resist.} = 0{,}95$ и $\alpha_{consum.} = 0{,}95$.

Figures 7 and 8 show value graphs of the function *I* depending on the number of installed sensors I.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

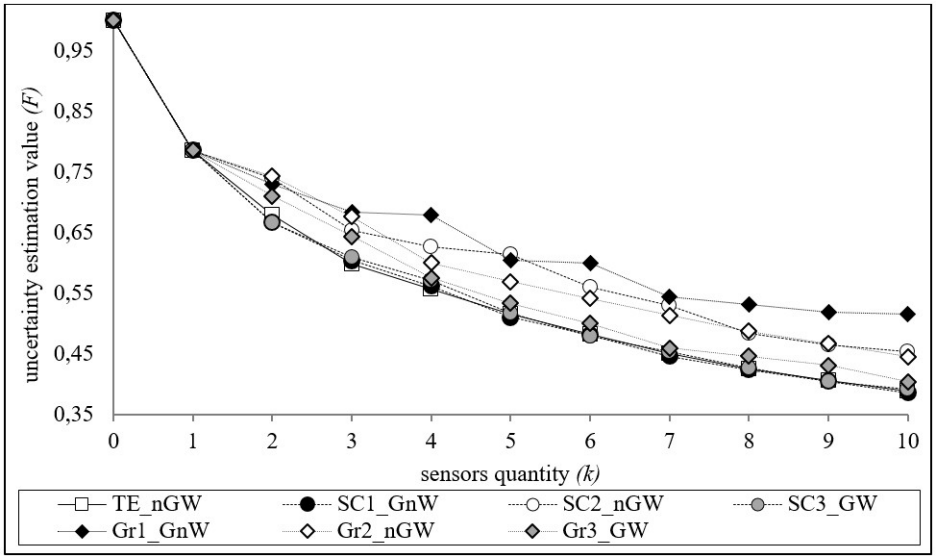


Fig. 7: Uncertainty Estimation Value *F* according to Sensors Quantity (ZJ)

The best result for the average value of the uncertainty function was shown by the SC1_GnW (0.570) algorithm, slightly ahead - TE_nGW (0.571) and SC3_GW (0.575), then - Gr3_GW (0.590), Gr2_nGW (0.621), SC2_nGW (0.628) and Gr1_GnW (0.654).

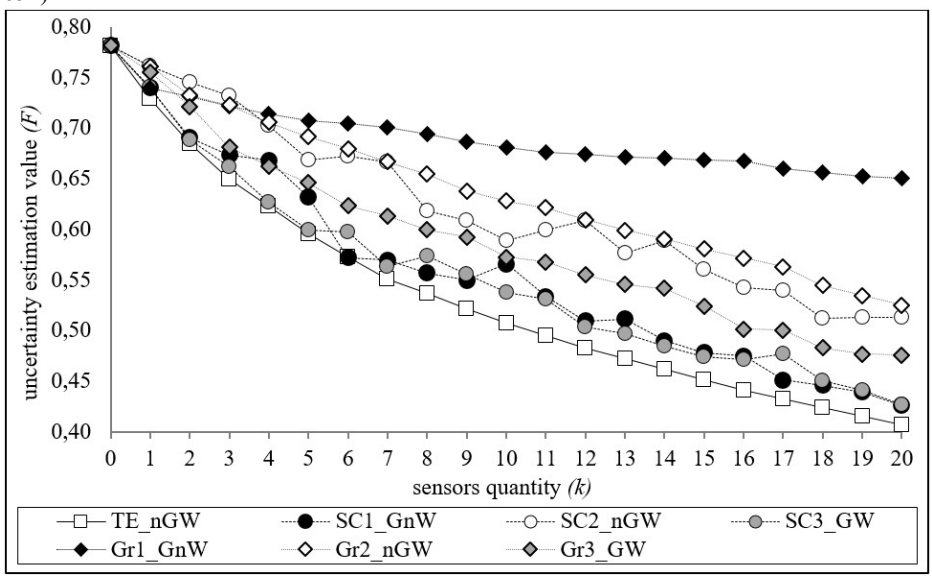


Fig. 8: Uncertainty Estimation Value *F* according to Sensors Quantity (D-Town)

The best result for the average value of the uncertainty function was shown by the TE_nGW (0.535) algorithm, then - SC1_GnW (0.560), and SC3_GW (0.561), and then - Gr3_GW (0.591), SC2_nGW (0.624), Gr2_nGW (0.638) and Gr1_GnW (0.691).

Tables 1 and 2 show the results showing how many sensors need to be installed in the network to achieve the specified pressure uncertainty estimation values *F*.

***F* function value**	**Algorithms**						
	TE_nGW	**SC1_GnW**	**SC2_nGW**	**SC3_GW**	**Gr1_GnW**	**Gr2_nGW**	**Gr3_GW**
0.8	1 (114)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
0.7	2 (228)	2 (2)	3 (3)	2 (2)	3 (3)	3 (3)	3 (3)
0.6	3 (342)	4 (4)	6 (6)	4 (4)	6 (6)	4 (4)	4 (4)
0.5	6 (684)	6 (6)	8 (8)	6 (6)	11 (11)	8 (8)	6 (6)

Table 1: Number of Sensors and Iterations to reach F function Value (ZJ)

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

F function value	Algorithms						
	TE_nGW	SC1_GnW	SC2_nGW	SC3_GW	Gr1_GnW	Gr2_nGW	Gr3_GW
0.8	1 (407)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
0.7	2 (814)	2 (2)	5 (5)	2 (2)	7 (7)	5 (5)	3 (3)
0.6	5 (2035)	6 (6)	10 (10)	5 (5)	25 (25)	13 (13)	8 (8)
0.5	11 (4477)	15 (15)	22 (22)	13 (13)	34 (34)	23 (23)	18 (18)

Table 2: Number of Sensors and Iterations to achieve F Value (D-Town)

It is considered the options of the function value $F = 0.8; 0.7; 0.6; 0.5$. For each of the algorithms, the number of sensors required for placement is given, and in parentheses - the number of iterations. The best solutions for the number of sensors in the tables are grayed out.
By the number of iterations required to solve the problem, the trial and error method is much inferior to the greedy algorithm and spectral clustering, for example, for ZJ with $k = 6$: 684 iterations opposite 6.

Summary
Despite the fact that the trial and error algorithm TE_nGW provides high accuracy indicators, the calculations require significant resources, which increase with an increase in the number of nodes in the network. Algorithms SC1_GnW - spectral clustering with a topological feature, and SC3_GW - spectral clustering with a combined feature show results slightly lagging behind the trial and error algorithm, while being far ahead of the solution time. Gr3_GW - greedy algorithm with a combined feature showed an average result. SC2_nGW - spectral clustering with an uncertainty estimate, Gr2_nGW - greedy algorithm with an uncertainty estimate, Gr1_GnW - greedy algorithm with a topological feature showed low results on the solution accuracy.
Thus, the use of algorithms based on spectral clustering with topological and combined features to solve the placement problem makes it possible to obtain a quasi-optimal solution in an acceptable time. This conclusion is confirmed by the results of experiments on two networks.

Conclusions
In this paper, we consider the basic principles of the theory of spectral clustering, describe the main approach of normalized spectral clustering of graphs. To solve the problem of forming the disconnected subgraphs, we proposed an algorithm for the priority distribution of nodes based on iterative transmission of the vertices of isolated areas to the most priority neighboring subgraphs. It is considered the question of optimal placement of objects in graph models of hydraulic networks by methods based on trial and error algorithms, greedy and spectral clustering. It is shown that the application of spectral clustering algorithms to solve this problem makes it possible to obtain a quasi-optimal solution in a short time.

Acknowledgements
The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References
[1] G.I.Galimova, D.T.Yakupov. Statistical analysis of the urban water consumption for administrative-business sector//Amazonia Investiga.Vol 7,No 17 (2018): 414- 425.
[2] Yakimov, I., Kirpichnikov, A., Mokshin, V., Yakhina, Z., Gainullin, R. The comparison of structured modeling and simulation modeling of queueing systems. Communications in Computer and Information Science (CCIS) volume 800. Springer. 2017. P. 256-267.
[3] Schloegel K., Karypis G., Kumar V. Graph partitioning for high performance scientific simulations / In J. Dongarra, I. Foster, G. Fox, K. Kennedy, and A. White, editors, CRPC Parallel Computing Handbook, Morgan Kaufmann, 2001.
[4] Kernighan B.W., Lin S. An efficient heuristic procedure for partitioning graphs // The Bell System Technical Journal. Vol. 49, №2, 1970. P. 291–307.
[5] A.L. Zheleznyakova. Effective decomposition methods of unstructured adaptive grids for high-performance calculations in solving computational aerodynamics problems // Physicochemical Kinetics in Gas Dynamics 2017, V. 18 (1).
[6] Mokshin V.V., Yakimov I.M., Yulmetyev R.M., Mokshin A.V. Recursive-regression self-organization of models for analysis and control of complex systems // Nonlinear World. 2009. V. 7. No. 1. P. 66-76.
[7] Sayfudinov I.R., Mokshin V.V., Tutubalin P.I., Kirpichnikov A.P. The graph representation optimization model for highlighting significant structures using the example of visual data preprocessing // Bulletin of the Technological University. 2018. V. 21. No. 5. P. 121-129.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495

[8] Sayfudinov I.R., Mokshin V.V., Kirpichnikov A.P., Tutubalin P.I., Sharnin L.M. A method for highlighting significant areas in graphic information for decision support systems / Bulletin of the Technological University. 2018. V. 21. No. 6. P. 146-152.
[9] Mokshin V.V., Yakimov I.M. The method of forming a model analysis of a complex system / Information Technology. 2011. No. 5. P. 46-51.
[10] Fiedler M. Algebraic connectivity of graphs. Czechoslov. Math. J. 1973, 23, 298–305.
[11] Shi J., Malik J. Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2000, 22, 888–905.
[12] A.Y. Ng, M.I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In Advances in Neural Information Processing Systems (NIPS), volume 14, pages 849–856, 2002.
[13] P.I. Tutubalin, V.V. Mokshin The Evaluation of the cryptographic strength of asymmetric encryption algorithms. 2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC), IEEE (2017) 180-183.
[14] G. Chen, S. Atev, and G. Lerman. Kernel spectral curvature clustering (KSCC). In Dynamical Vision Workshop, IEEE 12th International Conference on Computer Vision, pages 765–772, Kyoto, Japan, 2009.
[15] G. Chen and G. Lerman. Spectral curvature clustering (SCC). Int. J. Comput. Vision, 81(3): 317–330, 2009.
[16] Igor M. Yakimov, Mikhail V. Trusfus, Vladimir V. Mokshin, Alexander P. Kirpichnikov / AnyLogic, ExtendSim and Simulink Overview Comparison of Structural and Simulation Modelling Systems. 2018 3rd Russian-Pacific Conference on Computer Technology and Applications (RPC). IEEE (2018). 18-25 Aug. 2018.
[17] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. Ann. Statist., 36(2): 555–586, 2008.
[18] Mohar B. The Laplacian spectrum of graphs. In Graph Theory, Combinatorics, and Applications; Wiley: Hoboken, NJ, USA, 1991; pp. 871–898.
[19] U. Luxburg. A tutorial on spectral clustering. Statistics and Computing, 17(4): 395–416, 2007.
[20] S.A. Lyasheva, M.V. Medvedev, M.P. Shleymovich and V.V. Mokshin The analysis of image characteristics on the base of energy features of the wavelet transform. IPERS-ITNT. Image Processing and Earth Remote Sensing. Information Technology and Nanotechnology. CEUR Workshop Proceedings. Session Image Processing and Earth Remote Sensing Samara, Russia, April 24-27, 2018
[21] Brin S., Page L. The anatomy of a large-scale hypertextual web search engine. In Proceedings of the Seventh International World-WideWeb Conference (WWW 1998), Brisbane, Australia, 14–18 April 1998.
[22] I. R. Saifudinov, V.V. Mokshin, P.I. Tutubalin, L.M. Sharnin and D.G. Hohlov Visible Structures Highlighting Model Analysis Aimed at Object Image Detection Problem. IPERS-ITNT. Image Processing and Earth Remote Sensing. Information Technology and Nanotechnology. CEUR Workshop Proceedings. Session Image Processing and Earth Remote Sensing Samara, Russia, April 24-27, 2018. P.139-148.

  Helix E-ISSN: 2319-5592; P-ISSN: 2277-3495