

Article

An Evaluation of the Information Technology of Gene Expression Profiles Processing Stability for Different Levels of Noise Components [†]

Sergii Babichev ^{1,2} 

¹ Department of Informatics, Jan Evangelista Purkyně University in Ústí nad Labem, 400 96 Ústí nad Labem-město, Czech Republic; sergii.babichev@ujep.cz; Tel.: +420-777-843-785

² Department of Information Technologies, IT Step University, 79019 Lviv, Ukraine

[†] This paper is an extended version of conference paper: Babichev, S.; Lytvynenko, V.; Škvor, J.; Korobchynskiy, M.; Voronenko, M. Information Technology of Gene Expression Profiles Processing for Purpose of Gene Regulatory Networks Reconstruction. In Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining and Processing (DSMP), Lviv, Ukraine, 21–25 August 2018; pp. 336–341.

Received: 10 September 2018; Accepted: 1 November 2018; Published: 5 November 2018



Abstract: This paper presents the results of research concerning the evaluation of stability of information technology of gene expression profiles processing with the use of gene expression profiles, which contain different levels of noise components. The information technology is presented as a structural block-chart, which contains all stages of the studied data processing. The hybrid model of objective clustering based on the SOTA algorithm and the technology of gene regulatory networks reconstruction have been investigated to evaluate the stability to the level of the noise components. The results of the simulation have shown that the hybrid model of the objective clustering has high level of stability to noise components and vice versa, the technology of gene regulatory networks reconstruction is rather sensitive to the level of noise component. The obtained results indicate the importance of gene expression profiles preprocessing at the early stage of the gene regulatory network reconstruction in order to remove background noise and non-informative genes in terms of the used criteria.

Keywords: objective clustering; biclustering; gene regulatory networks; reconstruction; validation; gene expression profiles; noise component; systems stability

1. Introduction

Relevance of the problem is determined by the current works in the field of gene expression profiles processing for the purpose of gene regulatory networks reconstruction. Gene regulatory network is a set of genes, which interact with each other and with other elements in the cells to control the specific cells functions [1–3]. Qualitatively reconstructed gene regulatory network promotes to better understanding of the genes interaction mechanism in order to develop new methods to early diagnostics and treatment of complex diseases and for making new effective medicines. In [4–6] authors provide a comparative study of the main available association measures for characterizing gene regulatory strengths. They compare different measures in their consistency and specificity of detecting gene regulatory relationships. In these works the authors summarize and categorize the main frameworks and methods currently available for inferring transcriptional regulatory networks from microarray gene expression profiling data. The gene expression profiles, which are obtained by DNA microarray experiments [7,8] or by RNA-molecules sequencing technology [9,10] are the basis for gene regulatory networks reconstruction. High dimension of feature space in all cases and existence of complex noise components in the case of DNA-microarray technology use are the

distinctive peculiarities of the studied data. The reconstruction of gene networks based on the whole dataset of gene expression profiles is very complicated task due to the following aspects: it requests large computer resources and the complexity of the obtained networks complicates the obtained result interpretation. Therefore, it is necessary at the early stage of gene regulatory network reconstruction to process gene expression profiles with the use of current computational and information technologies of complex data processing. This process includes: data filtering in the case of DNA-microchip experiment use, non-informative genes reducing in terms of statistical criteria and Shannon entropy and data clustering and biclustering in order to allocate mutually correlated genes and samples.

In [11,12] authors present the results of research concerning the use of a wavelet filter for decreasing of the noise power in the studied data. The authors have shown the advantages of this technology in comparison with the fast Fourier transform. The bicluster analysis is relevant to allocate the mutually correlated genes and samples nowadays [13–16]. However, it should be noted that the main disadvantages of this technology implementation are the large quantity of small biclusters, and large amount of loss of useful information at the stage of biclusters formation. In [17] authors proposed the information technology of the step by step gene expression profiles processing in order to reconstruct gene regulatory networks. Practical implementation of this technology involves gene expression profiles clustering with the use of the DBSCAN clustering algorithm [18] at the first step and the SOTA clustering algorithm [19,20] at the second step. Further, bicluster analysis should be implemented on the obtained clusters. According to the authors' research, the implementation of this technology allows us to save more useful information due to the paralleling of the investigated data processing. The objective clustering inductive technology was proposed in [21–23] to determine the optimal parameters of the appropriate clustering algorithm. The results of the research concerning evaluation of the stability of the objective clustering inductive technology based on k-means clustering algorithm with the use of the data containing different levels of noise components were presented in [24]. However, k-means clustering algorithm is not effective for gene expression profiles clustering. The issues concerning creation of the method of fuzzy clustering task for multi-variate short time series with the unevenly distributed observations were investigated in [25]. The proposed method allows the authors to process the time series in both the batch mode and sequential on-line mode. However, it should be noted that the authors' researches are primarily focused on low-dimensional data processing. High-dimensional data processing is not considered in these works. Moreover, the authors did not investigate the effectiveness of the proposed methods to process the noise data with different levels of noise amplitude. Thus, it should be noted that in spite of the achieved successful results in this subject area the evaluation of the stability of information technology of the gene expression profiles processing with the use of the gene expression profiles which contain different levels of noise components has not been sufficiently investigated until present.

The aim of the paper is the evaluation of the stability of both the hybrid model of objective clustering based on the self-organizing SOTA clustering algorithm and the technology of the gene regulatory networks reconstruction based on the obtained biclusters to the level of noise components in the case of gene expression profiles use.

2. Materials and Methods

The structural block-chart of the information technology of the gene expression profiles processing for the purpose of the gene regulatory networks reconstruction and validation of the obtained models is presented in Figure 1 [17]. The implementation of this technology involves the following stages:

Stage I. Formation of the gene expression profiles array in the case of DNA microchip experiments use

Two technologies are relevant for the formation of the gene expression array nowadays: DNA microchip technology and mRNA molecules sequencing method. In the case of mRNA molecules sequencing method use we have the array of gene expression profiles.

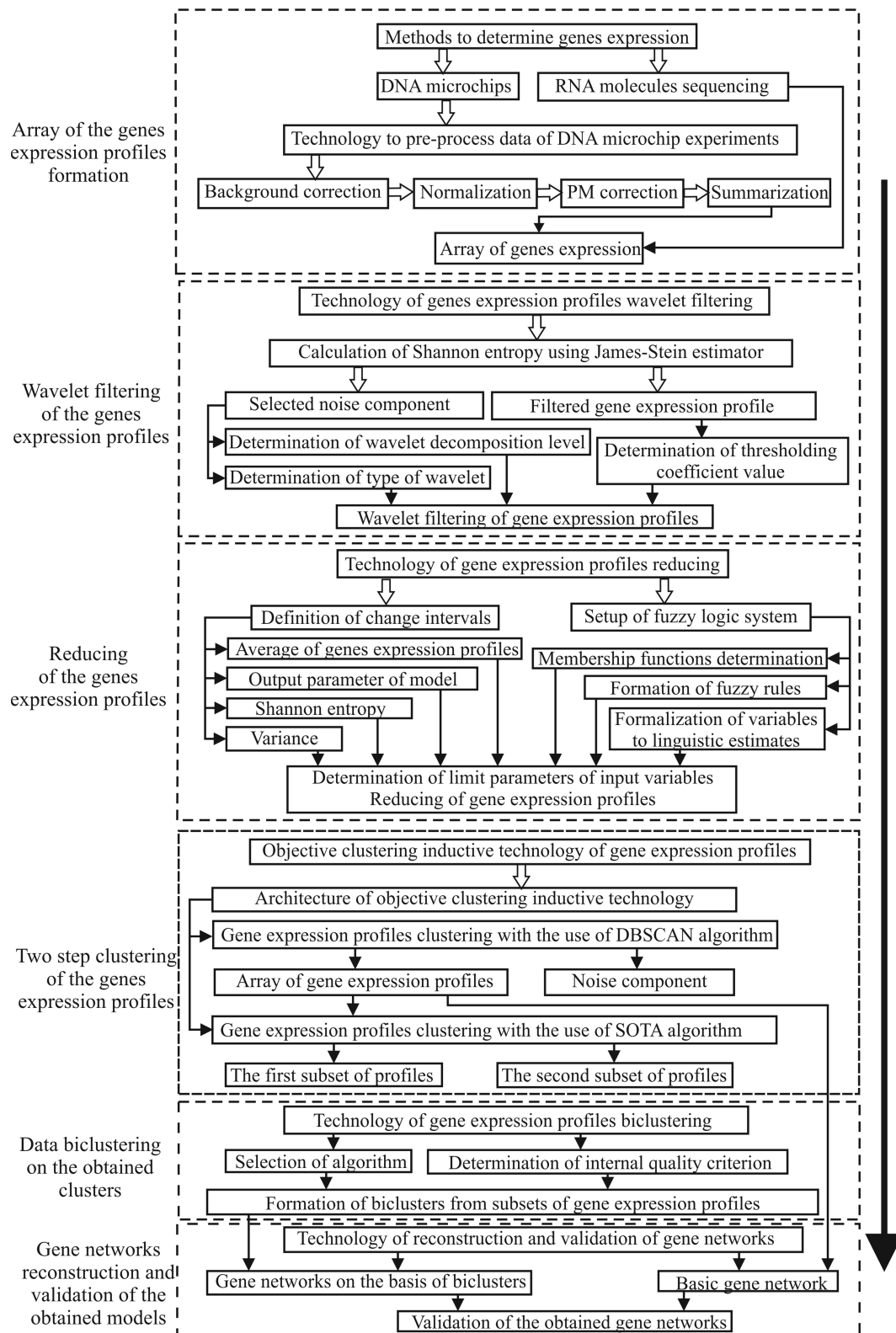


Figure 1. A structural flow chart of the information technology of gene expression profiles processing for the purpose of gene regulatory network reconstruction.

Moreover, implementation of this technology allows us to determine gene expression more exactly in comparison with the use of DNA microchip technology use. However, the use of mRNA molecules sequencing method is very expensive. Implementation of DNA microchip technology involves four steps: background correction, normalization, PM-correction and summarization. Each of the steps involves the use of different methods. The determination of the optimal combination of the methods in terms of minimum value of Shannon entropy calculated based on James-Stein shrinkage estimator [26] was performed at this stage.

Stage II. Wavelet filtering of the gene expression profiles

The necessity of this stage is determined by the existence of the background noise, which can appear during the scanning of the information from the DNA microchips. The wavelet filtering process was used at this stage. The implementation of this process involves calculation of the approximation and detail coefficients at the levels of wavelet decomposition from 1 to N. The model, which is implemented within the framework of the proposed information technology, involves determination of the wavelet filter optimal parameters on the basis of concurrent evaluation of Shannon entropy for both the filtered data and allocated noise component. The type of the wavelet and the level of wavelet decomposition are determined based on the maximum value of Shannon entropy for the allocated noise component. The optimal value of thresholding coefficient for the detail coefficients processing is determined on the basis of the minimum value of Shannon entropy for the filtered data. The algorithm works in such a way that if the value of Shannon entropy for the filtered data increases at the first step of thresholding coefficient change, the filtering process is stopped. In this case the studied data do not need any filtering.

Stage III. Gene expression profiles reduction

The aim of this stage is the division of the studied gene expression profiles into informative and non-informative in terms of complex use of statistical criteria and Shannon entropy. It is assumed that if the variance and average of absolute value of gene expression profiles are less and the value of Shannon entropy is greater than the corresponding boundary values, then these profiles are removed from the studied data as non-informative without significant loss of useful information. The fuzzy logic system was used to determine the boundary values of the appropriate parameters within the framework of the proposed technology [17]. After the determination of the boundary values a stepwise comparison of the variance and the average of absolute value and Shannon entropy of the gene expression profiles with the appropriate boundary values are performed. If the following conditions are true:

$$var \leq var_{lim}; abs \leq abs_{lim}; entr \geq entr_{lim}, \quad (1)$$

then this gene is allocated from the data as non-informative. Otherwise, the gene profile is recognized as informative for the further analysis.

Stage IV. Step-by-step gene expression profiles clustering within the framework of the objective clustering inductive technology

The implementation of the objective clustering inductive technology involves the division of the initial dataset into two the equal power subsets (containing the same quantity of pairwise similar objects). Then, the clustering process is carried out on both subsets concurrently and the calculation of both the internal and external clustering quality criteria at each step of the appropriate algorithm operation is performed. At the final step the complex balance criterion is calculated based on the internal and external criteria. The maximum value of the balance criterion corresponds to the optimal parameters of the appropriate clustering algorithm operation. The use of DBSCAN clustering

algorithm allows us to allocate the genes, which are identified as noise. The densities of these genes distribution in the feature space are significantly less in comparison with the density of other genes distribution. These genes are removed from the studied data. At the second step of the clustering process the gene expression profiles are divided into two clusters with the use of SOTA clustering algorithm. These subsets are used for the following bicluster analysis.

Stage V. Bicluster analysis of the obtained subsets of gene expression profiles

The allocation of small groups of mutually correlated genes and samples from the studied gene expression array is carried out during the biclustering process. The comparative analysis of different bicluster algorithms effectiveness with the use of both the internal and external biclustering quality criteria is presented in [27]. The tested biclusters and gene expression profiles were used during the simulation process. The model of the gene expression profiles biclustering based on “ensemble” biclustering method [28] has been proposed as the result of the research. The implementation of this model allows us to determine the optimal parameters of “ensemble” biclustering method in terms of the minimal value of the internal biclustering quality criterion. Then, the biclustering process is performed with the implementation of the “ensemble” biclustering method using the optimal parameters of this algorithm operation.

Stage VI. Gene regulatory networks reconstruction and validation of the obtained models

The reconstruction of the gene regulatory networks was performed based on the correlation inference algorithm with the use of software Cytoscape [29]. The optimal topology of the obtained gene networks was determined on the basis of the maximum value of the general Harrington desirability index [30], which contains the topological parameters of networks as the components. The validation of the obtained models was performed based on the comparison analysis of the existence of the direct links between the appropriate genes in the basic network (reconstructed on the basis of the complete set of the studied genes) and in the networks reconstructed based on the obtained biclusters. The basis of ROC-analysis theory was used to calculate the complex relative validation criterion, which indicates the quality of the obtained gene networks. The larger value of this criterion corresponds to the larger level of adequacy of gene networks, reconstructed on the basis of the biclusters to the basic network in terms of existence of direct links between the appropriate genes in different networks.

2.1. The Evaluation of the Stability of the Objective Clustering Inductive Model Based on the SOTA Algorithm to the Level of the Noise Component

The objective clustering inductive model based on the self-organizing SOTA clustering algorithm was investigated to evaluate the stability of the model to the level of the noise component. Gene expression profiles of 2000 patients who were examined on lung cancer [31] were used in this case. The length of the studied vectors was equal to the number of the studied samples (96). The simulation process involved the following steps:

1. Generation of random values vector. The length of this vector is equal to the length of the studied gene expression profiles and its amplitude corresponds to the minimum value of the studied data gene expression (“white noise”).
2. Setup of the vector of coefficients to change the amplitude of the noise component. In the case of the studied gene expression profiles the values of coefficients were changed within the range from 0.2 to 4 with step 0.2. These parameters were determined empirically during the simulation process.
3. Formation of gene expression profiles with the noise by adding of the appropriate noise components to the studied gene expression profiles.

4. Division of the obtained data into two equal power subsets by the use of the algorithm presented in [21].
5. Gene expression profiles clustering with the use of the method described in detail in [21] using SOTA clustering algorithm. The value of the sister cell weigh coefficient (*scell*) was changed within the small range from 8×10^{-4} to 11×10^{-4} with the step 2×10^{-5} . This range was determined empirically during the previous simulation process. The value of the variation coefficient was taken as zero.
6. Calculation of the complex balance criterion (general Harrington desirability index) for each value of the sister cell weigh coefficient. Creation of the plots of complex balance criterion versus the weigh coefficient value for both the data without noise and the data with different levels of noise component. Determination of the SOTA clustering algorithm optimal parameters, which correspond to the maximum value of the complex balance criterion. Data clustering with the use of SOTA algorithm with its optimal parameters.
7. Calculation of the external clustering quality criteria, which allows us to compare the clustering results for both the data without noise and the data with noise component. The following criteria were used as the external clustering quality criteria in this case:

- Jaccard index:

$$J = \frac{a}{a + b + c}. \quad (2)$$

- Kulczynski index:

$$K = \frac{a}{2 \times (a + b)} + \frac{a}{2 \times (a + c)}, \quad (3)$$

where a is the number of objects distributed in the same clusters in different clustering; b is the number of objects in the clusters of the first clustering, which did not coincide with the appropriate objects in the clusters of the second clustering; c is the number of objects in the clusters of the second clustering, which did not coincide with the appropriate objects in the clusters of the first clustering.

8. Analysis of the obtained results.

2.2. Evaluation of the Stability of the Model of Gene Regulatory Networks Reconstruction to the Level of the Noise Component

In this case the gene expression profiles of data moe430a [32] from database ArrayExpress were used during the simulation process. This data contains the gene expression profiles of the mesenchymal cells from the two distinct lineages, neural crest and mesoderm derived. 1000 of the gene expression profiles from 20 samples were used during the simulation process. The random “white noise” was added to each of the studied gene expression profiles. An amplitude of the noise components was determined by the following:

$$A = k \times (\max(v) - \min(v)), \quad (4)$$

where v is the vector of gene expression, the length of which is equal to the number of the studied samples; k is the coefficient, which limits the amplitude of the noise vector. The value of k coefficient was changed within the range from 0.025 to 0.1 with the step 0.025 during the simulation process. So, the four database of gene expression profiles with different level of noise were generated as the result of this process implementation.

The data biclustering was performed with the use of “ensemble” biclustering method according to the method described in detail in [27]. Finally, the gene regulatory network reconstruction and validation of the obtained model were carried out. The relative criterion of validation was calculated for the reconstructed networks based on both the data without noise and data with different levels of noise component.

3. Results and Discussion

3.1. Results of the Simulation Concerning the Use of the Objective Clustering Inductive Technology Based on the SOTA Clustering Algorithm

Figure 2 presents the charts of the complex balance criterion versus the sister cell weigh coefficient (*scell*) of SOTA clustering algorithm, which was implemented within the framework of the objective clustering inductive technology [21–23]. The noised gene expression profiles of the patients who were examined on lung cancer disease were used in this case. The optimal value of the *scell*, which corresponds to the maximum value of general Harrington desirability index was determined during the simulation process. The results of the simulation have shown that the increase of the amplitude coefficient of the noise components from 0.2 to 3.2 does not significantly influence the character of the balance criterion change. Figure 3 shows the charts of the number of objects in the clusters, the values of Jaccard and Kulczynski indexes and the relative changes of these indexes in percent versus the amplitude coefficient of the noise component.

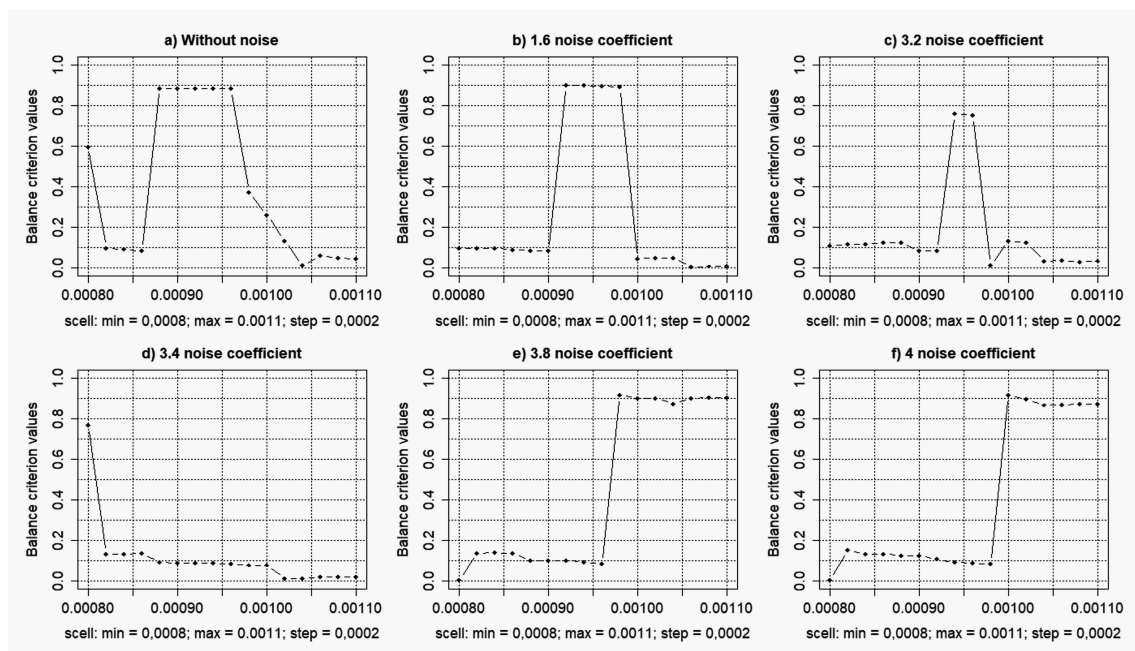


Figure 2. Charts of the complex balance criterion versus the sister cell weigh coefficient (*scell*) for the gene expression profiles with the different levels of noise component.

The analysis of the obtained charts allows us to conclude that the character of the objects distribution within the clusters is changed slightly during the increase of the noise amplitude coefficient. It is natural since the existence and the increase of the amplitude of the noise components in the studied data changes the gene expression profiles. In this case the movement of the object between clusters is possible. The values of Jaccard and Kulczynski indexes decrease monotonically in this case but the speed of these indexes changes chaotically in the defined range. This character of these parameters change is observed to value of the amplitude coefficient of noise 3.2. The charts of the appropriate parameters are changed significantly in the case of larger value of the noise amplitude. The *scell* optimal value of SOTA clustering algorithm, which corresponds to the maximum value of the complex balance criterion is changed chaotically too. This fact indicates the non-stability of the system. The number of the objects in the clusters and the values of Jaccard and Kulczynski indexes in the case of large values of the noise amplitude coefficient are changed very slowly.

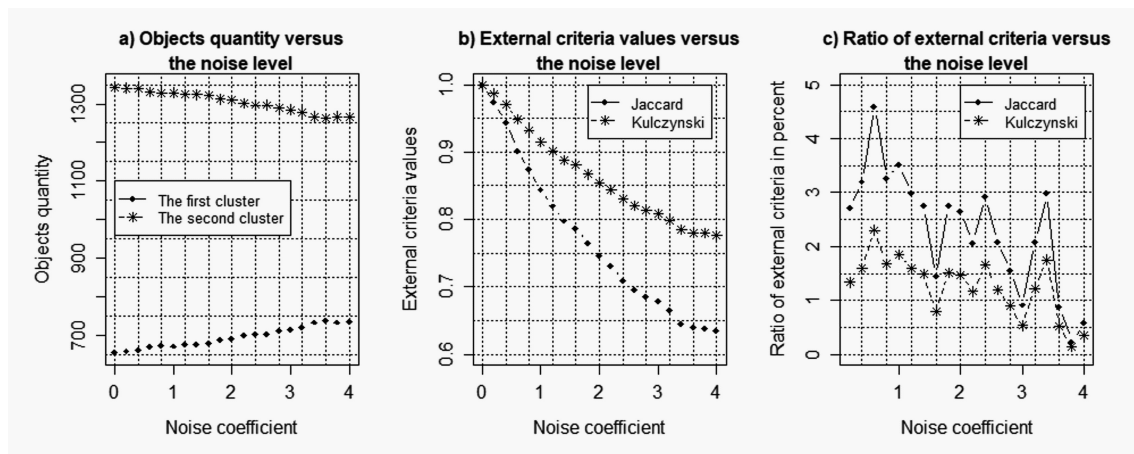


Figure 3. Charts of: (a) the quantity of gene expression profiles in different clusters; (b) Jaccard and Kulczynski indexes values; (c) the relative changes of Jaccard and Kulczynski indexes versus the amplitude coefficient of noise component.

As it can be seen from chart Figure 3c, the speed of these parameters changes in this case tends to zero. This fact can be explained in the following way. In the case of high level of the noise components local particularities of the gene expression profiles become smoother and clustering in this case is carried out by the estimation of the coarse component of the appropriate vector. Therefore, the *scell* value in this case is not determinative. The results of the simulation have shown that the clustering results in the case of the high level of noise components are almost the same and they do not depend on the *scell* value. The conducted research has shown also that the objective clustering inductive technology is effective and efficient in the case of the analysis of the complex data with the local particularities. The use of this technology to group the gene expression profiles is reasonable in the case of low level of the noise component.

3.2. Results of the Simulations Concerning the Influence of the Level of noise components to the Quality of the Reconstructed Gene Networks

Figures 4–7 show the charts of both number of the obtained biclusters and values of the biclustering quality criteria versus the parameters of the “ensemble” biclustering method (thresholding coefficient value (*thr*) and ratio of the number of the rows and columns in biclusters (*simthr*)) in the case of bicluster analysis of the gene expression profiles of the data *moe430a* with different levels of noise component. The following parameters were determined as the result of the obtained charts analysis:

- $k = 0.025$: $thr = 0.33$; $simthr = 0.29$;
- $k = 0.05$: $thr = 0.35$; $simthr = 0.11$;
- $k = 0.075$: $thr = 0.26$; $simthr = 0.07$;
- $k = 0.1$: $thr = 0.48$; $simthr = 0.33$;

Ten largest biclusters from each of the studied data were selected for the further analysis. The reconstruction of the gene regulatory networks and validation of the obtained models were performed based on Cytoscape software with the use of correlation inference algorithm [29]. The detailed description of the used information technology for the reconstruction and the validation of gene networks is presented in [33].

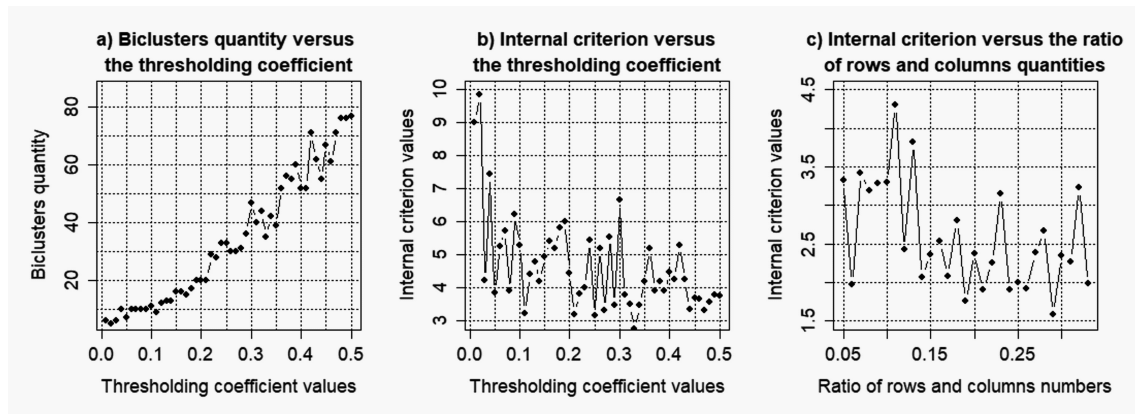


Figure 4. Results of the simulation to determine the optimal parameters of “ensemble” biclustering method for the noise coefficient $k = 0.025$.

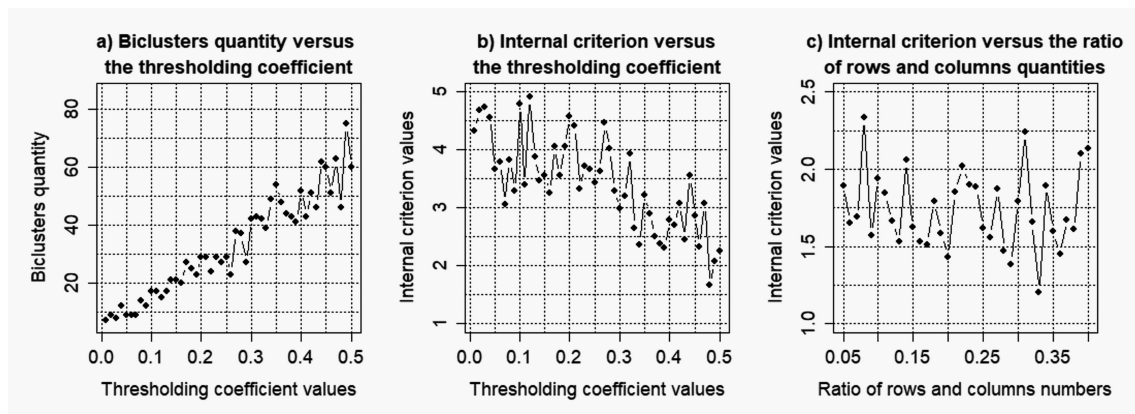


Figure 5. Results of the simulation to determine the optimal parameters of “ensemble” biclustering method for the noise coefficient $k = 0.05$.

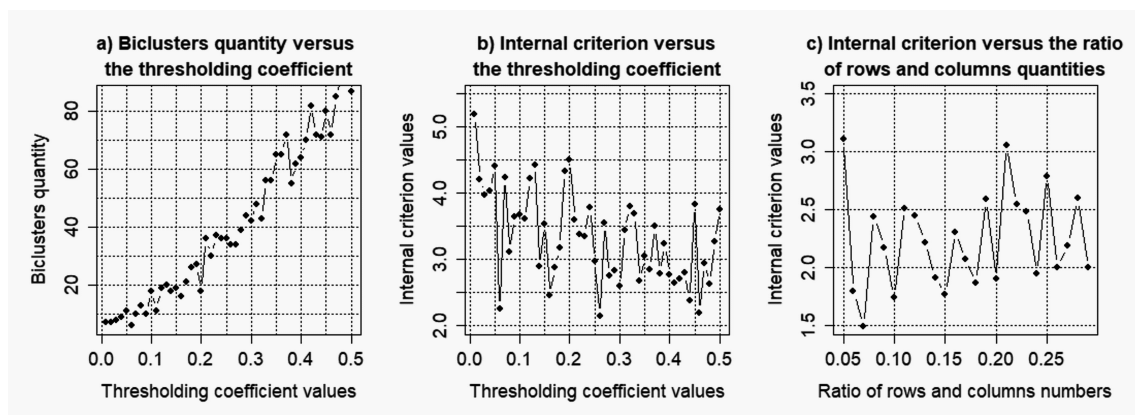


Figure 6. Results of the simulation to determine the optimal parameters of “ensemble” biclustering method for the noise coefficient $k = 0.075$.

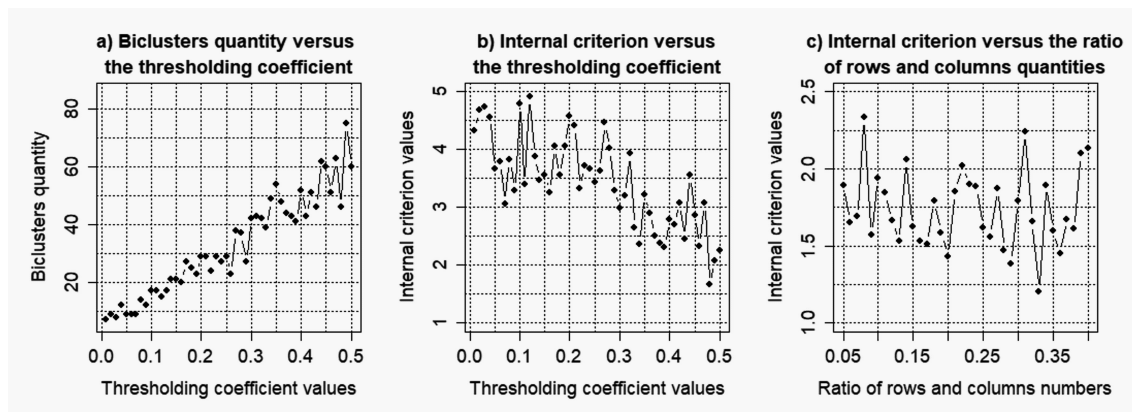


Figure 7. Results of the simulation to determine the optimal parameters of “ensemble” biclustering method for the noise coefficient $k = 0.1$.

Figures 8–11 presents the charts of general Harrington desirability index versus the value of thresholding coefficient for both the data without noise and the data with the different levels of noise component. The values of thresholding coefficient, which correspond to the maximum of Harrington desirability index for both the complete set of the studied gene expression profiles and the data in the obtained biclusters are presented in Table 1.

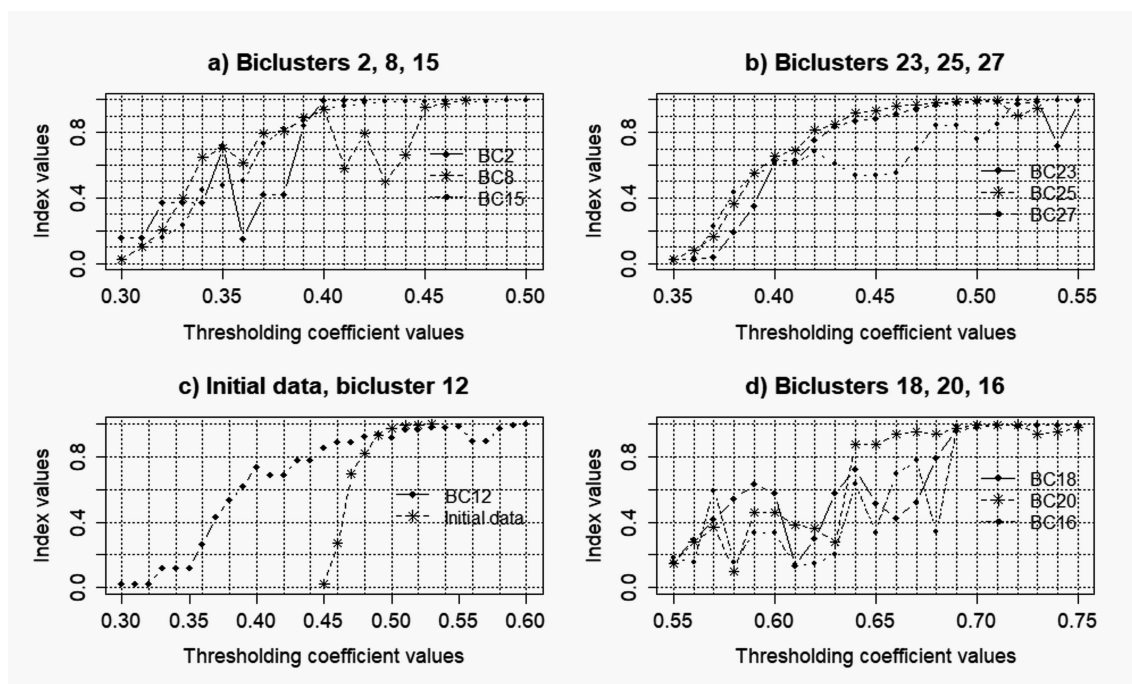


Figure 8. Charts of the general Harrington desirability index versus the value of thresholding coefficient for the gene networks reconstructed on the basis of the noise data with the noise coefficient $k = 0.025$.

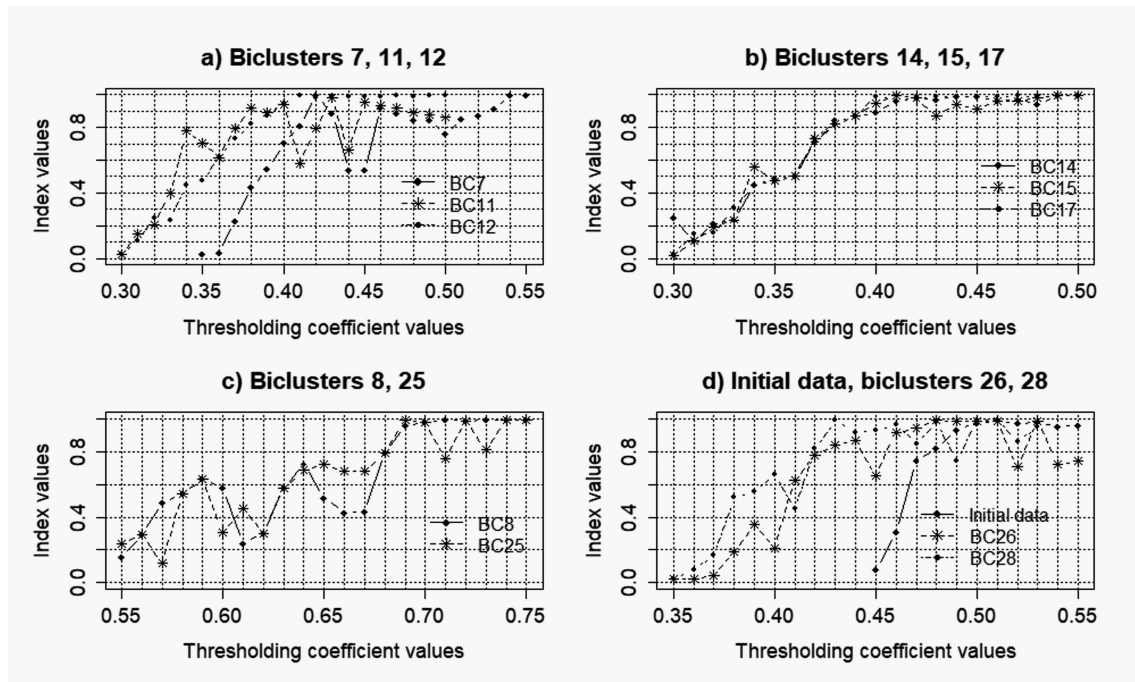


Figure 9. Charts of the general Harrington desirability index versus the value of thresholding coefficient for the gene networks reconstructed on the basis of the noise data with the noise coefficient $k = 0.05$.

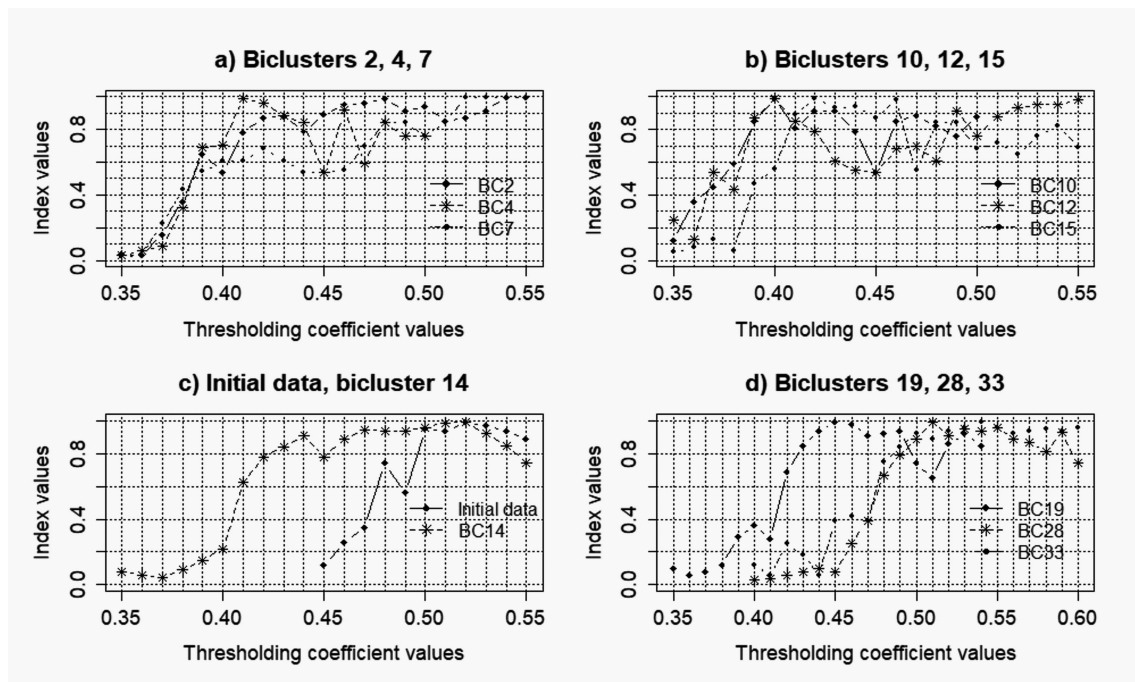


Figure 10. Charts of the general Harrington desirability index versus the value of thresholding coefficient for the gene networks reconstructed on the basis of the noise data with the noise coefficient $k = 0.075$.

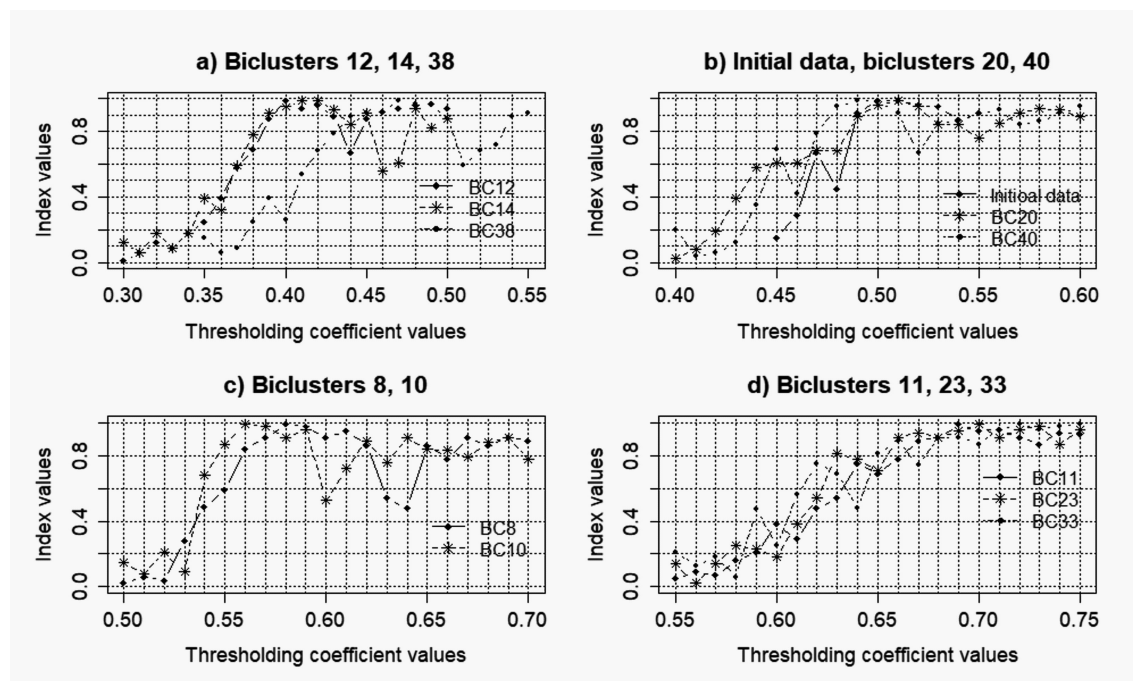


Figure 11. Charts of the general Harrington desirability index versus the value of thresholding coefficient for the gene networks reconstructed on the basis of the noise data with the noise coefficient $k = 0.1$.

Table 1. The values of thresholding coefficient for the gene networks reconstruction based on both the complete data and the obtained biclusters for the different levels of the noise component.

Noise Coef.	Full Data	BC2	BC8	BC12	BC15	BC16	BC18	BC20	BC23	BC25	BC27
0.025	0.52	0.43	0.4	0.52	0.4	0.7	0.7	0.7	0.5	0.5	0.52
Noise Coef.	Full Data	BC7	BC8	BC11	BC12	BC14	BC15	BC17	BC25	BC26	BC28
0.05	0.51	0.42	0.71	0.43	0.41	0.44	0.41	0.4	0.69	0.48	0.43
Noise Coef.	Full Data	BC2	BC4	BC7	BC10	BC12	BC14	BC15	BC19	BC28	BC33
0.075	0.52	0.48	0.41	0.4	0.43	0.4	0.52	0.42	0.45	0.51	0.54
Noise Coef.	Full Data	BC8	BC10	BC11	BC12	BC14	BC20	BC23	BC33	BC38	BC40
0.1	0.5	0.58	0.56	0.69	0.44	0.41	0.51	0.7	0.72	0.47	0.49

The results of the validation of the reconstructed gene regulatory networks are presented in Figure 12. The comparative analysis of the character of the appropriate genes interconnection in the gene networks reconstructed based on both the complete data and the obtained biclusters with the calculation of the errors of both the first and second types was performed at this stage. Then, the relative validation criterion was calculated according to the method described in details in [33]. A higher value of this criterion corresponds to a higher level of adequacy of the networks reconstructed based on the obtained biclusters to the network reconstructed based on the complete data in terms of the direct links coincidence between the appropriate genes in different networks.

The analysis of the obtained charts allows us to conclude that the existence of the noise components decreases the level of adequacy of the gene networks, reconstructed based on the biclusters to the network reconstructed on the basis of the complete data. The average of the relative validation criterion for the obtained models of gene networks are significantly less than the appropriate value of this criterion in the case of gene networks reconstruction based on the gene expression profiles without noise [33]. Moreover, the analysis of the charts in Figure 12 has shown that the increase of

the noise level in the data decreases the average of the relative validation criterion. This fact indicates the necessity of qualitative preprocessing of the gene expression profiles at the early stage of gene regulatory networks reconstruction.

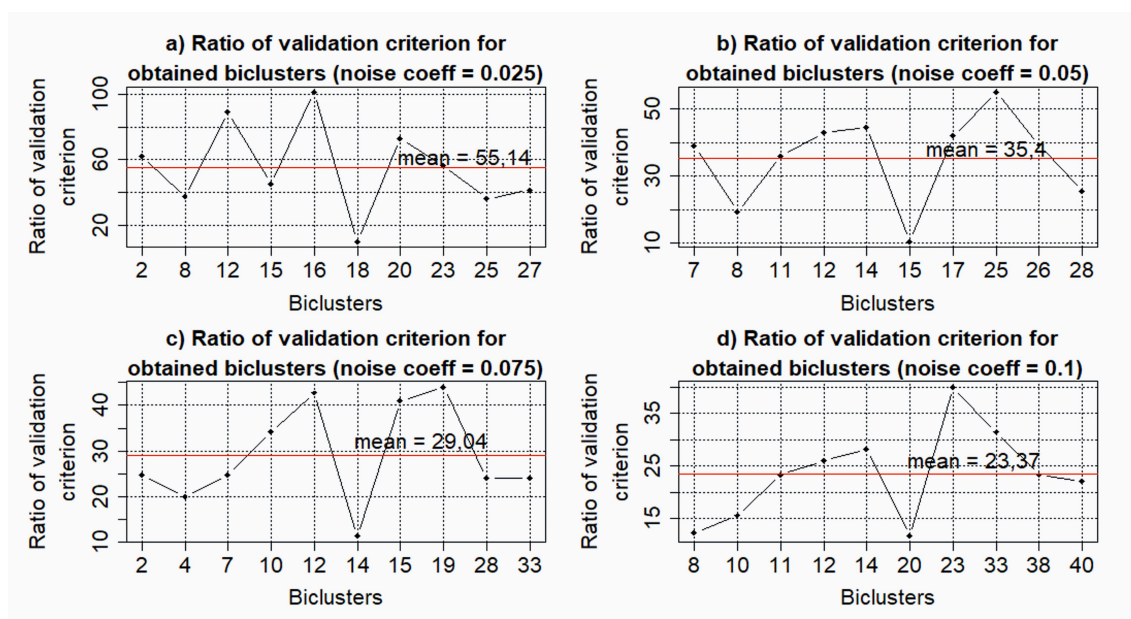


Figure 12. Results of the validation of the gene networks, reconstructed on the basis of the gene expression profiles with the different level of the noise component.

4. Conclusions

The conducted research has shown what gene expression profiles preprocessing in order to decrease the noise components is important and significant. Two stages of the proposed information technology of the gene expression profiles processing for the purpose of gene regulatory networks reconstruction and validation of the obtained models have been investigated to evaluate their sensitivity to the level of the noise component. The first stage is the objective clustering inductive model on the basis of the SOTA clustering algorithm. The second stage is the information technology of the gene regulatory networks reconstruction based on the selected biclusters and validation of the obtained models. The results of the research concerning the evaluation of the stability of the objective clustering inductive technology based on the SOTA clustering algorithm have shown that this technology is not sensitive to noise in the case of low level of amplitude of the noise component. However, in the case of high level of the noise amplitude the variation of the algorithm parameters does not change the clustering results. This fact indicates the effectiveness of the proposed technology to clustering data, which contain some quantity of noise. The results of the simulation concerning the evaluation of the stability of the information technology of the gene regulatory network reconstruction to the level of the noise components has shown, what this technology is very sensitive to the noise component. A slight increase of the noise amplitude promotes to the decrease of the level of adequacy of the networks reconstructed based on the gene expression profiles of the obtained biclusters in relation to the network reconstructed on the basis of the complete set of the studied data in terms of the direct links coincidence between the appropriate genes in different networks.

The following perspectives of the author's research is the investigation of different algorithms of the gene regulatory networks reconstruction, estimation of their effectiveness, development of the validation methods for the reconstructed networks and the simulation of the reconstructed network with the use of Bayes and Petri networks.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SOTA	Self-Organizing Tree Algorithm
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
PM	Perfect Match

References

1. Zak, D.E.; Vadigepalli, R.; Gonye, E.G.; Doyle, F.J.; Schwaber, J.S.; Ogunnaike, B.A. Unconventional systems analysis problem in molecular biology: A case study in gene regulatory network Modeling. *Comput. Chem. Eng.* **2005**, *2*, 547–563. [\[CrossRef\]](#)
2. Davidson, E.; Levin, M. Gene regulatory networks for development. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 4936–4942. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Liang, S.; Fuhrman, S.; Somogyi, R. *REVEAL, A General Reverse Engineering Algorithm for Inference of Genetic Network Architectures*; Pacific Symposium on Biocomputing, World Scientific Publishing Co.: Singapore, 1998; Volume 3, pp. 18–29.
4. Liu, Z.-P. Quantifying Gene Regulatory Relationships with Association Measures: A Comparative Study. *Front. Genet.* **2017**, *8*, 1–12. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Liu, Z.-P. Reverse Engineering of Genome-wide Gene Regulatory Networks from Gene Expression Data. *Curr. Genom.* **2015**, *16*, 3–22. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Liu, Z.-P.; Wu, C.; Miao, H.; Wu, H. RegNetwork: An integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database* **2015**, *2015*, bav095. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Zhang, X.; Zhou, Z.; Jiao, Y.; Niu, Y.; Wang, Y. A visual cryptography scheme-based DNA microarrays. *Int. J. Perform. Eng.* **2018**, *14*, 334–340. [\[CrossRef\]](#)
8. Shukla, S.; Agarwal, A.K.; Lakhmani, A. MICROCHIPS: A leading innovation in medicine. In Proceedings of the 3rd International Conference on Computing for Sustainable Global Development, New Delhi, India, 16–18 March 2016; pp. 205–210. [\[CrossRef\]](#)
9. Wu, X.; Yang, B.; Udo-Inyang, I.; Ji, S.; Ozog, D.; Zhou, L.; Mi, Q.-S. Research Techniques Made Simple: Single-Cell RNA Sequencing and its Applications in Dermatology. *J. Investig. Dermatol.* **2018**, *138*, 1004–1009. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Wang, L.Y.; Guo, J.; Cao, W.; Zhang, M.; He, J.; Li, Z. Integrated sequencing of exome and mRNA of large-sized single cells. *Sci. Rep.* **2018**. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Puchala, D.; Szczepaniak, B.; Yatsymirskyy, M. Effective Realizations of Biorthogonal Wavelet Transforms of Lengths $2K + 1/2K - 1$ with Lattice Structures on GPU and CPU. In Proceedings of the 16th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL), Wroclaw, Poland, 14–16 October 2015; pp. 130–137. [\[CrossRef\]](#)
12. Lipinski, P.; Yatsymirskyy, M. Efficient 1D and 2D daubechies wavelet transforms with application to signal processing. In Proceedings of the 8th International Conference on Adaptive and Natural Computing Algorithms (ICANNGA), Warsaw, Poland, 11–14 April 2007. [\[CrossRef\]](#)
13. Pontes, B.; Giráldez, R.; Aguilar-Ruiz, J.S. Biclustering on expression data: A review. *J. Biomed. Inform.* **2015**, *57*, 163–178. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Chi, E.C.; Allen, G.I.; Baraniuk, R.G. Convex Biclustering. *Biometrics* **2017**, *73*, 10–20. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Rocha, O.; Mendes, R. JBiclustGE: Java API with unified biclustering algorithms for gene expression data analysis. *Knowl.-Based Syst.* **2018**, *155*, 83–87. [\[CrossRef\]](#)
16. Puleo, G.J.; Milenkovic, O. Correlation Clustering and Biclustering with Locally Bounded Errors. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2018; pp. 4105–4119.

17. Babichev, S.; Lytvynenko, V.; Korobchynskiy, M.; Skvor, J.; Voronenko, M. Information Technology of Gene Expression Profiles Processing for Purpose of Gene Regulatory Networks Reconstruction. In Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining and Processing, Lviv, Ukraine, 21–25 August 2018; pp. 336–342.
18. Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial datasets with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; pp. 226–231.
19. Dorazo, J.; Corazo, J. Phylogenetic reconstruction using an unsupervised growing neural network that adopts the topology of a phylogenetic tree. *J. Mol. Evol.* **1997**, *44*, 226–259.
20. Fritzke, B. Growing cell structures a self-organizing network for unsupervised and supervised learning. *Neural Netw.* **1994**, *7*, 1441–1460. [[CrossRef](#)]
21. Babichev, S.; Lytvynenko, V.; Korobchynskiy, M.; Taif, M.A. Objective Clustering Inductive Technology of Gene Expression Sequences Features. *Commun. Comput. Inf. Sci.* **2017**, *716*, 359–372. [[CrossRef](#)]
22. Babichev, S.; Lytvynenko, V.; Osypenko, V. Implementation of the Objective Clustering Inductive Technology Based on DBSCAN Clustering Algorithm. In Proceedings of the 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 5–8 September 2017; pp. 479–484.
23. Babichev, S.; Taif, M.A.; Lytvynenko, V.; Osypenko, V. Criterial Analysis of Gene Expression Sequences to Create the Objective Clustering Inductive Technology. In Proceedings of the 2017 37th IEEE International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 18–20 April 2017; pp. 244–248.
24. Babichev, S.; Taif, M.A.; Lytvynenko, V. Estimation of the inductive model of objects clustering stability based on the k-means algorithm for different levels of data noise. *Radio Electron. Comput. Sci. Control* **2016**, *4*, 54–60. [[CrossRef](#)]
25. Setlak, G.; Bodyanskiy, Y.; Pliss, I.; Vynokurova, O.; Peleshko, D.; Kobylin, I. Adaptive Fuzzy Clustering of Multivariate Short Time Series with Unevenly Distributed Observations Based on Matrix Neuro-Fuzzy Self-organizing Network. *Adv. Intell. Syst. Comput.* **2018**, *643*, 308–315. [[CrossRef](#)]
26. Hausser, J.; Strimmer, K. Entropy inference and the james-stein estimator with application to nonlinear gene association networks. *J. Mach. Learn. Res.* **2009**, *10*, 1469–1484.
27. Babichev, S.; Lytvynenko, V.; Osypenko, V.; Korobchynskiy, M.; Voronenko, M. Comparison Analysis of Biclustering Algorithms With the Use of Artificial Data and Gene Expression Profiles. In Proceedings of the 2018 38th IEEE International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 24–26 April 2018; pp. 292–297.
28. Kaiser, S. Biclustering: Methods, Software and Application. Ph.D. Thesis, Faculty of Mathematics, Computer Science and Statistics, Loyola Marymount University, Los Angeles, CA, USA, 2011.
29. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]
30. Harrington, J. The desirability function. *Ind. Qual. Control* **1965**, *21*, 494–498.
31. Beer, D.G.; Kardia, S.L.; Huang, C.C.; Giordano, T.J.; Levin, A.M.; Misek, D.E.; Lin, L.; Chen, G.; Gharib, T.G.; Thomas, D.G.; et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* **2002**, *8*, 816–824. [[CrossRef](#)] [[PubMed](#)]
32. Bhattacharjee, V.; Mukhopadhyay, P.; Singh, S.; Johnson, C.; Philipose, J.T.; Warner, C.P.; Greene, R.M.; Pisano, M.M. Neural crest and mesoderm lineagedependent gene expression in orofacial development. *Differentiation* **2007**, *75*, 463–477. [[CrossRef](#)] [[PubMed](#)]
33. Babichev, S.; Korobchynskiy, M.; Lahodynskiy, O.; Korchomnyi, O.; Basanets, V.; Borynskiy, V. Development of technology of gene network reconstruction and validation based on gene expression profiles. *East.-Eur. J. Enterp. Technol.* **2018**, *1*, 19–32. [[CrossRef](#)]

