

УДК 004.048+004.94

С.А. БАБІЧЕВ

Університет Яна Евангелиста Пуркіне в Усті на Лабі, Чехія  
Херсонський державний університет, Херсон, Україна

Л.М. ЯСІНСЬКА-ДАМРІ

Українська академія друкарства, Львів, Україна

## МОДЕЛЮВАННЯ ПРОЦЕСУ РЕКОНСТРУКЦІЇ ГЕННОЇ РЕГУЛЯТОРНОЇ МЕРЕЖІ НА ОСНОВІ КОМПЛЕКСНОГО ЗАСТОСУВАННЯ ТОПОЛОГІЧНИХ ПАРАМЕТРІВ

*У статті представлено результати моделювання щодо визначення оптимальної топології генної регуляторної мережі в процесі її реконструкції із застосуванням кореляційного алгоритму реконструкції мережі. Генна регуляторна мережа представлялася у вигляді неорієнтованого графу, у якому вузли є гени або метаболіти, а дуги визначають зв'язок між відповідними елементами мережі. Оцінка топології мережі здійснювалася шляхом розрахунку значень одиночних топологічних параметрів, які враховують як структуру мережі, так і кількість і характер зв'язків між відповідними елементами. Були досліджені наступні одиночні топологічні параметри: кількість вузлів мережі, ступінь вузлів або їх зв'язність, щільність мережі, коефіцієнти кластеризації та централізації і гетерогенність. Остаточне рішення щодо топології мережі приймалося на основі аналізу узагальненого топологічного параметру, що розраховувався із застосуванням функції бажаності Харрінгтона. Моделювання процесу реконструкції генної мережі на основі профілів експресії генів було проведено у програмному середовищі Cytoscape з використанням профілів експресії генів даних toe430 бази даних ArrayExpress, що містять інформацію про експресію генів мезенхімальних клітин двох типів: нервового гребня та мезодерми. Процес реконструкції генної регуляторної мережі здійснювався із застосуванням кореляційного алгоритму, практична реалізація якого передбачає розрахунок коефіцієнтів парної кореляції між профілями експресії генів, що досліджуються. Топологія мережі у даному випадку формується на основі значення трешолдінгового коефіцієнта  $\tau$ , що визначає порогове значення наявності зв'язку між парою відповідних генів мережі. У результаті моделювання були створені діаграми розподілу одиночних топологічних параметрів і узагальненого топологічного індексу від значення коефіцієнту трешолдінгу. Аналіз отриманих діаграм дозволив визначити оптимальну топологію генної мережі.*

*Ключові слова: генна регуляторна мережа, топологія мережі, топологічні параметри, функція бажаності Харрінгтона, трешолдінг, алгоритм кореляційного виводу.*

S.A. BABICHEV

Jan Evangelista Purkyně University in Ústí nad Labem, Czech Republic  
Kherson State University, Kherson, Ukraine

L.M. YASINSKA-DAMRI

Ukrainian Academy of Printing, Lviv, Ukraine

## MODELLING OF GENE REGULATORY NETWORK RECONSTRUCTION PROCEDURE BASED ON THE COMPLEX USE OF TOPOLOGICAL PARAMETERS

*The paper presents the simulation results concerning the determination of the gene regulatory network optimal topology during its reconstruction using the correlation inference algorithm. The gene regulatory network was presented as an undirected graph, in which the nodes are genes or metabolites, and the arcs define the connection between the corresponding network elements. The assessment of the network topology was carried out by calculating the values of single topological parameters, taking into account both the structure of the network and the nature of the connections between the corresponding elements. The following single topological parameters were investigated: the number of nodes in the network, the degree of nodes or their connectivity, the density of the network, the coefficients of clustering and centralization, and network heterogeneity. The final decision regarding the network topology was done based on the analysis of the generalized topological index, which was calculated using the Harrington desirability function. Modelling of the gene network reconstruction process based on gene expression profiles was carried out in the Cytoscape software environment using the toe430 gene expression profiles of the ArrayExpress database, which contains information concerning the genes expression of two types of mesenchymal cells: neural crest and mesoderm. The*

*process of gene regulatory network reconstruction was carried out using a correlation inference algorithm, the practical implementation of which involves calculating the pair correlation coefficients between the studied gene expression profiles. The network topology, in this case, was formed on the basis of the thresholding coefficient  $\tau$ , which determines the threshold value for the presence of a connection between a pair of corresponding network genes. As a result of modelling, the diagrams of the distribution of single topological parameters and the generalized topological index versus the value of the thresholding coefficient were created. The analysis of these diagrams can allow us to determine the gene regulatory network optimal topology.*

*Keywords: gene regulatory network, network topology, topological parameters, Harrington desirability function, thresholding, correlation inference algorithm.*

**С.А. БАБИЧЕВ**

Университет Яна Евангелиста Пуркине в Усти на Лабе, Чехия  
Херсонский государственный университет, Херсон, Украина

**Л.М. ЯСИНСКАЯ-ДАМРИ**

Украинская академия книгопечатания, Львов, Украина

## **МОДЕЛИРОВАНИЕ ПРОЦЕССА РЕКОНСТРУКЦИИ ГЕННОЙ РЕГУЛЯТОРНОЙ СЕТИ НА ОСНОВЕ КОМПЛЕКСНОГО ИСПОЛЬЗОВАНИЯ ТОПОЛОГИЧЕСКИХ ПАРАМЕТРОВ**

*В статье представлены результаты моделирования по определению оптимальной топологии генной регуляторной сети в процессе ее реконструкции с применением корреляционного алгоритма реконструкции сети. Генная регуляторная сеть представлялась в виде неориентированного графа, в котором узлы являются генами или метаболитами, а дуги определяют связь между соответствующими элементами сети. Оценка топологии сети осуществлялась путем расчета значений одиночных топологических параметров, учитывающих как структуру сети, так и характер связей между соответствующими элементами. Были исследованы следующие одиночные топологические параметры: количество узлов сети, степень узлов или их связность, плотность сети, коэффициенты кластеризации и централизации и гетерогенность. Окончательное решение касательно топологии сети принималось на основе анализа обобщенного топологического параметра, который рассчитывался с применением функции желательности Харрингтона. Моделирование процесса реконструкции генной сети на основе профилей экспрессии генов проводилось в программной среде СутоScare с использованием профилей экспрессии генов тое430 базы данных ArrayExpress, содержащих информацию о экспрессии генов мезенхимальных клеток двух типов: нервного гребня и мезодермы. Процесс реконструкции генной регуляторной сети осуществлялся с применением корреляционного алгоритма, практическая реализация которого предполагает расчет коэффициентов парной корреляции между исследуемыми профилями экспрессии генов. Топология сети в данном случае формировалась на основе трешолдингового коэффициента  $\tau$ , который определяет пороговое значение наличия связи между парой соответствующих генов сети. В результате моделирования были построены диаграммы распределения одиночных топологических параметров и обобщенного топологического индекса от значения коэффициента трешолдинга. Анализ полученных диаграмм позволил определить оптимальную топологию генной сети.*

*Ключевые слова: генная регуляторная сеть, топология сети, топологические параметры, функция желательности Харрингтона, трешолдинг, алгоритм корреляционного вывода.*

### **Постановка проблеми**

Сучасні системи обробки інформації у більшості випадків ґрунтуються на використанні аналогій функціонування біологічних механізмів і процесів, що протікають у живих організмах. До таких процесів слід віднести функціонування природної нейронної мережі, імунні процеси, генна мережа тощо. Особливістю таких систем є децентралізована паралельна обробка інформації, великий рівень складності, здатність навчатися, розпізнавати інформацію та формувати рішення. Створення штучних моделей сучасних біологічних систем та дослідження їхньої поведінки можливе на основі системного підходу, який передбачає комплексне використання методів молекулярної біології, математики, інформатики, законів фізики, і створює умови для розуміння, які чинники визначають характер функціонування біологічної

системи з метою корегування даного процесу. Реконструкція та моделювання генної регуляторної мережі (ГРМ) формує основу для дослідження та аналізу характеру взаємодій генів і впливів цих взаємодій на функціональні можливості біологічного організму. Складність процесу реконструкції ГРМ визначається тим, що експериментальні дані, які використовуються для реконструкції мережі, зазвичай не дозволяють однозначно визначити структуру мережі та характер взаємозв'язку між її вузлами. Окрім того, велика кількість генів, які визначають структуру та обсяг мережі, ускладнює процес інтерпретації результатів, що отримуються. Тому виникає необхідність у дослідженнях по кількісній оцінці топології мережі та характеру взаємозв'язку між її елементами зі застосуванням у якості експериментальних даних профілів експресій генів, отриманих шляхом ДНК-мікрочіпових експериментів.

### **Аналіз останніх досліджень і публікацій**

Реконструкція ГРМ та інтерпретування результатів її роботи є досить складними задачами, які у даний час не мають однозначного розв'язку. Перші роботи по реконструкції біологічних генних мереж із експериментальних даних були опубліковані у кінці 90-х років минулого століття [1-4]. У даних роботах було запропоновано декілька підходів по даному типу моделювання. У [5-7] опубліковано декілька обзорів, які стосуються методів реконструкції та моделювання ГРМ. Основна ідея реконструкції ГРМ полягає у використанні експериментальних даних експресій генів для отримання моделей шляхом оцінки та аналізу зав'язків між молекулярними об'єктами. Однак слід зазначити, що даний процес є дуже складним внаслідок того, що дана проблема має комбінаторний характер, з одного боку, і експериментальні дані у багатьох випадках неповні та неточні, з іншого боку. Крім того, наявність великої кількості параметрів, змінних та обмежень викликає необхідність застосування чисельних та обчислюваних методів.

Сучасні технології отримання даних експресій генів мають тенденцію охоплювати максимальну кількість змінних системи [8-11]. Наприклад, технологія мікрочіпів ДНК дозволяє вимірювати експресію десятків тисяч генів одночасно, тобто кожний об'єкт, що досліджується, характеризується вектором експресій генів довжиною десятки тисяч одиниць. Використання повного набору параметрів дозволяє підвищити точність моделі, але при цьому ускладнюється процес пошуку адекватної моделі, оскільки розмірність простору пошуку експоненціально зростає зі збільшенням кількості параметрів моделі. Вочевидь, що у такому випадку гени можуть бути згруповані у модулі або кластери в залежності від функціональної подібності профілів їх експресій. Технологія кластеризації профілів експресій генів дозволяє вирішити проблему зменшення кількості генів [12-15]. Моделі, що отримані з використанням кластерів взаємно координуваних генів, легше інтерпретуються в процесі подальшої валідації моделі з використанням додаткової інформації щодо експресії генів. Але при цьому виникає проблема вибору кількості кластерів та рівню деталізації процесу кластеризації даних. Таким чином, можна зробити висновок, що ГРМ представляють складну систему взаємодіючих елементів, у якій прості невеликі компоненти виконують комплексне колективне завдання. При цьому слід зазначити, що на даний час не існує ефективної технології реконструкції генної регуляторної мережі, яка здатна з високим ступенем ймовірності передбачити характер подальшого розвитку біологічного організму на генному рівні.

**Метою статті** є проведення моделювання щодо оптимізації топології генної регуляторної мережі на основі комплексного застосування одиночних топологічних параметрів мережі та функції бажаності Харрінгтона.

### Викладення основного матеріалу дослідження

ГРМ представляє собою орієнтований або неорієнтований граф, дуги якого можуть бути зваженими (при наявності ваги, що визначає силу зв'язку) або незваженими. Тому для визначення параметрів, що визначають топологію мережі, може бути використана теорія графів. У процесі моделювання розраховувалися наступні топологічні параметри мережі [16]:

- *Кількість вузлів* мережі визначає загальну кількість генів, що мають зв'язки між собою.
- *Ступенем вузла мережі* або його *зв'язністю* називають сумарну вагу зв'язків (дуг), що з'єднують даний вузол із сусідніми вузлами:

$$k_i = \sum_{j=1, i \neq j}^{n_i} \omega_{ij}, \quad (1)$$

де  $n_i$  – кількість вузлів  $i$ -го гена,  $\omega_{ij}$  – вага дуги, що з'єднує гени  $i$  та  $j$ .

- *Середній ступінь* або *середня зв'язність* мережі визначається як середнє значення ступенів усіх вузлів мережі:

$$k_{\text{сеп}} = \frac{1}{n} \sum_{i=1}^n k_i. \quad (2)$$

- *Максимальний ступінь* визначає максимальне значення елементів вектору зв'язностей всіх вузлів мережі:

$$k_{\text{max}} = \max(\overline{k_1}, \overline{k_i}). \quad (3)$$

- *Щільність мережі* визначається як відношення кількості зважених зв'язків між вузлами мережі до максимально можливої кількості зв'язків між вузлами даної мережі:

$$Dens = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \omega_{ij}}{0.5 \times n(n-1)}. \quad (4)$$

- *Коефіцієнт кластеризації вузла* визначає ймовірність того, що найближчі сусіди даного вузла зв'язані між собою безпосередньо. *Коефіцієнт кластеризації мережі* визначається як середнє значення коефіцієнтів кластеризації всіх її вузлів:

$$Cl = \frac{1}{n} \sum_{i=1}^n \frac{e_i}{0.5 \times k_i(k_i - 1)}, \quad (5)$$

де  $n$  – кількість генів у мережі,  $e_i$  – кількість реальних зв'язків вузла  $i$  із сусідніми вузлами,  $k_i$  – кількість сусідів гена  $i$ , враховуючи сам ген, які можуть скласти повноцінний кластер.

- *Коефіцієнт централізації мережі* визначає ступінь її близькості до топології зірки. Даний коефіцієнт розраховується за формулою:

$$Centr = \frac{n}{n-2} \left( \frac{k_{\text{max}}}{n-1} - Dens \right). \quad (6)$$

- *Гетерогенність мережі* визначає ступінь неоднорідності топології мережі та виражається через дисперсію і середнє значення середнього ступеня вузлів за формулою:

$$G = \frac{\sqrt{\text{var}(k_{\text{cep}})}}{\text{mean}(k_{\text{cep}})} . \quad (7)$$

Параметри (1) – (7) дозволяють зробити попередню оцінку топології моделі ГРМ. При незмінній кількості вузлів менші значення щільності та кластеризації мережі й більше значення гетерогенності свідчить про більш високу якість топології мережі за рахунок її спрощення і різноманітності при незмінній кількості вузлів. Значення коефіцієнта централізації свідчить про ступень близькості топології мережі до зіркоподібної структури. При цьому слід відзначити, що оптимальна структура мережі відповідає мінімальним значенням щільності вузлів і коефіцієнта кластеризації та максимальним значенням коефіцієнтів централізації та гетерогенності. Для прийняття остаточного рішення по вибору структури мережі запропоновано комплексний критерій на основі функції бажаності Харрінгтона, використання якої передбачає перетворення шкал топологічних параметрів у лінійну шкалу безрозмірного показника  $Y$ , значення якого змінюється від -2 до 5. Приватні бажаності для кожного значення показника  $Y$  розраховуються у відповідності до формули:

$$d = \exp(-\exp(-Y)) . \quad (8)$$

Шкала бажаності має інтервал від 0 до 1. Значення  $d = 0$  відповідає абсолютно неможливій топології з точки зору даного критерію, а значення  $d = 1$  – найкращій топології. Вибір оцінок на шкалі бажаності 0,63 і 0,37 пояснюється зручністю розрахунків:  $0,63=1-1/e$ , а  $0,37=1/e$ . Значення 0,37 зазвичай відповідає межі допустимих значень. Узагальнений індекс бажаності Харрінгтона як середнє геометричне значення усіх приватних бажаностей:

$$GDI = \sqrt[n]{\prod_{i=1}^n d_i} . \quad (9)$$

Максимальне значення індексу бажаності Харрінгтона відповідає коефіцієнту трешолдінгу, який дозволяє отримати оптимальну структуру генної мережі на основі комплексного аналізу її топологічних параметрів.

Процес реконструкції ГРМ на основі кореляційного аналізу передбачає розрахунок коефіцієнтів парної кореляції між профілями експресій генів, що досліджуються. Оскільки у випадку аналізу матриці експресій генів вектори профілів являють собою послідовності раціональних чисел, доцільним є використання метода Пірсона для розрахунку парної кореляції між відповідними профілями. Коефіцієнт парної кореляції у випадку його значущості представляє силу взаємозв'язку між відповідними вузлами мережі. Вага дуги дорівнює коефіцієнту кореляції пари профілів експресій генів, між якими оцінюється зв'язок. Топологія мережі у даному випадку визначається значенням трешолдінгового коефіцієнта  $\tau$ , що визначає порогове значення наявності зв'язку між парою генів мережі. Ваговий коефіцієнт дуги, що з'єднує відповідні гени, визначається наступним чином:

$$\omega(g_a, g_b) = \begin{cases} 0, & \text{якщо } r(g_a, g_b) < \tau; \\ r(g_a, g_b), & \text{якщо } r(g_a, g_b) \geq \tau. \end{cases} \quad (10)$$

Моделювання процесу реконструкції генної мережі на основі профілів експресій генів було проведено у програмному середовищі *CytoScape* [17] з використанням профілів експресій генів даних *moe430* бази даних *ArrayExpress*[18], що містять

інформацію про експресію генів мезенхімальних клітин двох типів: нервового гребня та мезодерми. Дані отримані шляхом ДНК-мікрочіпових експериментів. Матриця вихідних даних містила 147 генів та 20 умов визначення експресій відповідних генів. Практична реалізація процесу моделювання передбачала наступні етапи:

Етап I. Постановка проблеми. Формування даних.

1. Формування вихідних даних у вигляді матриці, де рядки представляють гени, а стовпці – умови проведення ДНК-мікрочіпового експерименту.

Етап II. Приблизна оцінка інтервалу зміни трешолдінгового коефіцієнта.

2. Завдання приблизного інтервалу та кроку зміни коефіцієнта трешолдінгу.

Ініціалізація початкового значення трешолдінгового коефіцієнта:  $\tau = \tau_{\min}$ .

3. Реконструкція ГРМ, топологія якої відповідає завданому значенню трешолдінгового коефіцієнта.

4. Розрахунок топологічних параметрів отриманої генної регуляторної мережі.

5. Якщо значення коефіцієнта трешолдінгу менше за максимальне, збільшення даного значення на  $d\tau$  (крок зміни коефіцієнта трешолдінгу) та перехід на крок 3 даної процедури. У протилежному випадку, побудова діаграм залежності отриманих топологічних параметрів від значення трешолдінгового коефіцієнта.

6. Аналіз отриманих результатів, визначення нового, більш вузького інтервалу та меншого кроку зміни значення трешолдінгового коефіцієнта.

Етап III. Визначення оптимального значення коефіцієнта трешолдінгу.

7. Реконструкція генної регуляторної мережі в межах нового інтервалу зміни значень трешолдінгового коефіцієнта. Розрахунок параметрів оцінки топології генної регуляторної мережі на кожному кроці зміни значення трешолдінгового коефіцієнта.

8. Побудова діаграм зміни значень топологічних параметрів у залежності від коефіцієнта трешолдінгу. Аналіз отриманих результатів. Визначення оптимального значення трешолдінгового коефіцієнта.

Етап IV. Реконструкція генної регуляторної мережі.

9. Реконструкція ГРМ із застосування оптимального значення коефіцієнта трешолдінгу.

На рис. 1 зображені діаграми зміни одиночних топологічних параметрів при зміні коефіцієнта трешолдінгу від 0,45 до 0,55 з кроком 0,01. При цьому, у випадку наявності декількох зв'язаних між собою, але відокремлених від основної мережі генів, виділялася мережа генів, що мають найбільшу кількість взаємозв'язаних вузлів. Підмережа з декількома вузлами видалялася з мережі.

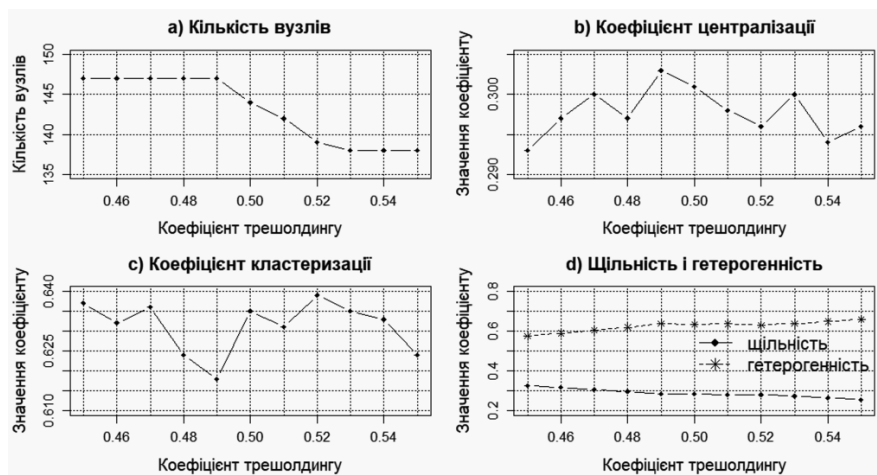


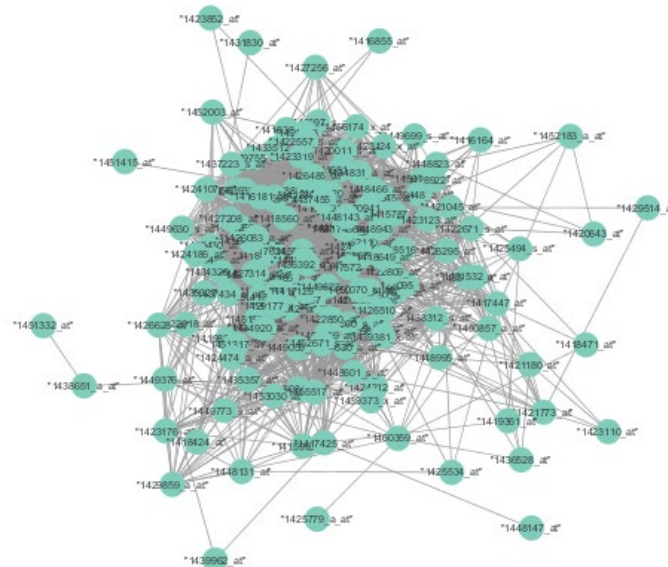
Рис. 1. Діаграми зміни одиночних параметрів оцінки топології генної мережі в інтервалі зміни коефіцієнта трешолдінгу від 0,45 до 0,55 з кроком 0,01

Як можна бачити з рис. 1, значення коефіцієнта трешолдінгу, що визначає структуру ГРМ, визначається чотирма топологічними параметрами: коефіцієнтами кластеризації, централізації і гетерогенності мережі та щільністю її вузлів. При цьому слід відзначити, що оптимальна структура мережі відповідає мінімальним значенням щільності вузлів і коефіцієнта кластеризації та максимальним значенням коефіцієнтів централізації та гетерогенності. На рис. 2 зображено діаграму залежності значення комплексного критерію від трешолдінгового коефіцієнта, значення якого змінювалися в інтервалі від 0,45 до 0,55 з кроком 0,01.



**Рис. 2. Діаграма розподілу індексу бажаності Харрінгтона при різних значеннях коефіцієнта трешолдінгу**

Аналіз отриманих результатів показує, що оптимальним за одиночними параметрами оцінки топології генної мережі є значення коефіцієнта трешолдінгу 0,49. У цьому випадку мережа містить 147 генів, коефіцієнт централізації досягає максимуму, а коефіцієнт кластеризації – мінімуму. Значення коефіцієнтів щільності та гетерогенності в інтервалі зміни трешолдінгового коефіцієнта від 0,45 до 0,49 монотонно змінюються у меншу та більшу сторони відповідно. В інтервалі від 0,49 до 0,51 швидкість зміни даних параметрів дорівнює нулю. Значення комплексного критерію, розрахованого на основі функції бажаності Харрінгтона, також досягає максимуму при значенні коефіцієнта трешолдінгу 0,49. На рис. 3 зображений результат реконструкції генної регуляторної мережі у випадку застосування алгоритму кореляційного виводу з коефіцієнтом трешолдінгу 0,49.



**Рис. 3. Результат реконструкції генної мережі при застосуванні алгоритму кореляційного виводу**

### Висновки

У статті представлені результати моделювання щодо оптимізації значення трешолдінгового коефіцієнта при реконструкції генної регуляторної мережі при застосуванні кореляційного алгоритму реконструкції мережі. Даний параметр визначає топологію мережі, що отримується, та її функціональні можливості в процесі подальшого моделювання. Отримані діаграми залежності значень топологічних параметрів від коефіцієнту трешолдінгу, аналіз яких показав, що більш якісна з точки зору топології є мережа з меншими значеннями середньої щільності вузлів і коефіцієнта кластеризації та більшими значеннями коефіцієнтів централізації й гетерогенності мережі. Остаточне рішення по формуванню мережі (фіксація значення коефіцієнта трешолдінгу) приймалося на основі максимуму узагальненого індексу бажаності Харрінгтона, який містив як компоненти відповідні топологічні параметри генної мережі.

### Список використаної літератури

1. D'haeseleer P., Wen X., Fuhrman S., Somogyi R. Linear modeling of mRNA expression levels during CNS development and injury. *Pacific Symposium on Biocomputing*. 1999. P. 41–52.
2. Liang S., Fuhrman S., Somogyi R. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. *Pacific Symposium on Biocomp*. 1998. P. 18–29.
3. Friedman N., Linial M., Nachman I., Pe'er D. Using Bayesian networks to analyse expression data. *Journal of Computational Biology*. 2000. №7(3-4). P. 601–620.
4. Chen T., He H. L., Church G. M. Modeling gene expression with differential equations. *Proceedings of the Pacific Symposium on Biocomputing*. 1999. P. 29–40.
5. Bansal M., Belcastro V., Ambesi-Impiomato A., di Bernardo D. How to infer gene networks from expression profiles. *Molecular Systems Biology*. 2007. №3(78). P.1–10.
6. Yaghoobi H., Haghipour S., Hamzeiy H., et al. A Review of Modeling Techniques for Genetic Regulatory Networks. *Journal of Medical Signals and Sensors*. 2012. №2(1). P. 61–70.
7. Emmert-Streib F., Dehmer M., Haibe-Kains B. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in cell and developmental biology*. 2014. P. 1–7.
8. Sonesson C., Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*. 2013. №14(91). P. 1–18.
9. Arnone M.I., Davidson E.H. The hardwiring of development: organization and function of genomic regulatory systems. *Development*. 1997. №124(10). P. 1851–1864.
10. Jeong H., Tombor B., Albert R., et al. The large-scale organization of metabolic networks. *Nature*. 2000. № 407(6804). P. 651–654.
11. Bork P., Jensen L.J., von Mering C., et al. Protein interaction networks from yeast to human. *Current Opinion in Structural Biology*. 2004. №14(3). P. 292–299.
12. Babichev S., Taif M. A, Lytvynenko V. Inductive model of data clustering based on the agglomerative hierarchical algorithm. *Proceeding of the 2016 IEEE First International Conference on Data Stream Mining and Processing (DSMP)*. 2016. P.19–22.
13. Babichev S., Taif M. A., Lytvynenko V., Korobchynskyi M. Objective clustering inductive technology of gene expression sequences features. *Communication in Computer and Information Science*. 2017. № 716. P. 359–372.



14. Babichev S., Lytvynenko V., Skvor J., Fiser J. Model of the objective clustering inductive technology of gene expression profiles based on SOTA and DBSCAN clustering algorithms. *Advances in Intelligent Systems and Computing*. 2018. №.689. P. 21–39.
15. Pontes B., Giráldez R., Aguilar-Ruiz J. S. Biclustering on expression data: A review. *Journal of Biomedical Informatics*. 2015. № 57. P. 163–180.
16. Assenov Y., Ramirez F., Schelhorn S.-E., et al. Computing topological parameters of biological networks. *Bioinformatics*. 2008. №24(2). P. 282–284.
17. Shannon P., Markiel A., et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003. P. 2498–2504.
18. Bhattacharjee V., Mukhopadhyay P., et al. Neural crest and mesoderm lineage-dependent gene expression in orofacial development. *Differentiation*, 2007. P. 128–139.

### References

1. D'haeseleer, P., Wen, X., Fuhrman, S., & Somogyi R. (1999). Linear modeling of mRNA expression levels during CNS development and injury. *Pacific Symposium on Biocomputing*. pp. 41–52.
2. Liang, S., Fuhrman, S., & Somogyi, R. (1998). REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. *Pacific Symposium on Biocomp*. pp. 18–29.
3. Friedman, N., Linial, M., Nachman, I., & Pe'er, D. (2000). Using Bayesian networks to analyse expression data. *Journal of Computational Biology*. 7(3-4), 601–620.
4. Chen, T., He, H. L., & Church, G. M. (1999). Modeling gene expression with differential equations. *Proceedings of the Pacific Symposium on Biocomputing*. pp. 29–40.
5. Bansal, M., Belcastro, V., Ambesi-Impimbatto, A., & di Bernardo, D. (2007). How to infer gene networks from expression profiles. *Molecular Systems Biology*. 3(78), 1–10.
6. Yaghoobi, H., Haghypour, S., Hamzeiy, H., et al. (2012). A Review of Modeling Techniques for Genetic Regulatory Networks. *Journal of Medical Signals and Sensors*. 2(1), 61–70.
7. Emmert-Streib, F., Dehmer, M., & Haibe-Kains, B. (2014). Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in cell and developmental biology*. pp. 1–7.
8. Sonesson, C., & Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*. 14(91), 1–18.
9. Arnone, M.I., & Davidson E.H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. *Development*. 124(10), 1851–1864.
10. Jeong, H., Tombor, B., Albert, R., et al. (2000). The large-scale organization of metabolic networks. *Nature*. 407(6804), 651–654.
11. Bork, P., Jensen, L.J., von Mering, C., et al. (2004). Protein interaction networks from yeast to human. *Current Opinion in Structural Biology*. 14(3), 292–299.
12. Babichev, S., Taif, M. A., & Lytvynenko V. (2016). Inductive model of data clustering based on the agglomerative hierarchical algorithm. *Proceeding of the 2016 IEEE First International Conference on Data Stream Mining and Processing (DSMP)*. pp.19–22.

13. Babichev, S., Taif, M. A., Lytvynenko, V., & Korobchynskiy, M. (2017). Objective clustering inductive technology of gene expression sequences features. *Communication in Computer and Information Science*. **716**, 359–372.
14. Babichev, S., Lytvynenko, V., Skvor, J., & Fiser, J. (2018). Model of the objective clustering inductive technology of gene expression profiles based on SOTA and DBSCAN clustering algorithms. *Advances in Intelligent Systems and Computing*. **689**, 21–39.
15. Pontes, B., Giráldez, R., & Aguilar-Ruiz, J. S. (2015). Biclustering on expression data: A review. *Journal of Biomedical Informatics*. **57**, 163–180.
16. Assenov, Y., Ramirez, F., Schelhorn, S.-E., et al. (2008). Computing topological parameters of biological networks. *Bioinformatics*. **24**(2), 282–284.
17. Shannon, P., Markiel, A., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* pp. 2498–2504.
18. Bhattacharjee, V., Mukhopadhyay, P., et al. (2007). Neural crest and mesoderm lineagedependent gene expression in orofacial development. *Differentiation*. pp. 128–139.