

УДК 004.048

ТЕОРЕТИЧНІ ДОСЛІДЖЕННЯ ЩОДО ФОРМУВАННЯ ОПТИМАЛЬНОЇ ТОПОЛОГІЇ ГЕННОЇ РЕГУЛЯТОРНОЇ МЕРЕЖІ

І. М. Лях

*Ужгородський національний університет,
пл. Народна, 3, Ужгород, 88000, Україна*

Подано результати теоретичних досліджень щодо формування оптимальної топології генної регуляторної мережі, яка може бути наведена у вигляді орієнтованого або неорієнтованого графа. Розглянуто топологічні параметри, комбінація яких визначає топологію мережі. Показано, що характер зміни різних топологічних параметрів у процесі розрядження мережі може суперечити один одному, що ускладнює процес оптимізації топології мережі. Визначені шляхи оптимізації топології генної регуляторної мережі, які передбачають формування структури і топології мережі за групою одиночних критеріїв, таких як коефіцієнт кластеризації мережі, щільність мережі, коефіцієнт централізації та гетерогенність із застосуванням комплексного критерію, який містить окремі критерії як компоненти на першому етапі. При цьому структура та топологія мережі формується шляхом оптимізації відповідного коефіцієнта трешолдінгу, що визначає наявність або відсутність зв'язків між відповідними вузлами. На другому етапі здійснюється оцінка відповідності мережі ступеневому закону шляхом аналізу відповідних розподілених параметрів.

***Ключові слова:** експресія генів, генна регуляторна мережа, топологічні параметри, оптимізація топології мережі.*

Постановка проблеми. Генна регуляторна мережа (ГРМ) — мережа взаємодії молекулярних елементів (протеїнів, метаболітів, генів тощо), що визначають біологічні властивості об'єкта, які досліджуються. У цьому випадку виникає необхідність у вирішенні проблеми оптимізації топології мережі з подальшою валідацією реконструйованої мережі шляхом застосування тестових експериментальних даних і оцінки відповідності значенням цільових функцій. Очевидно, що топологія мережі визначається кількістю генів/протеїнів, що містяться у вузлах мережі, та кількістю і характером зв'язків між відповідними вузлами мережі. Реконструкція ГРМ на основі великої кількості генів є проблематичною, оскільки у цьому випадку виникає проблема оптимізації топології мережі та оцінки її адекватності. У наш час існують і постійно поповнюються бази даних генних мереж різних організмів [1], що містять інформацію щодо цільових генів, які дають змогу ідентифікувати стан відповідного біологічного організму. Але у будь-якому випадку цільові гени мають складний характер прямих та непрямих взаємодій з іншими вузлами мережі, що визначають відповідний стан

цільових генів. Очевидно, що повнозв'язна мережа (має максимальну кількість зв'язків між вузлами мережі) не є ефективною за рахунок складності її топології, що визначає зменшення чутливості мережі (реакції цільових вузлів на зміну параметрів інших вузлів мережі). Зменшення кількості зв'язків між вузлами мережі, яка знаходиться нижче відповідного порогу, також погіршує її ефективність через зменшення точності оцінки стану відповідного об'єкта, оскільки при цьому можуть бути втрачені важливі зв'язки, що визначають стан цільових генів. На рис. 1 зображено формальну схему процесу формування топології ГРМ [2].

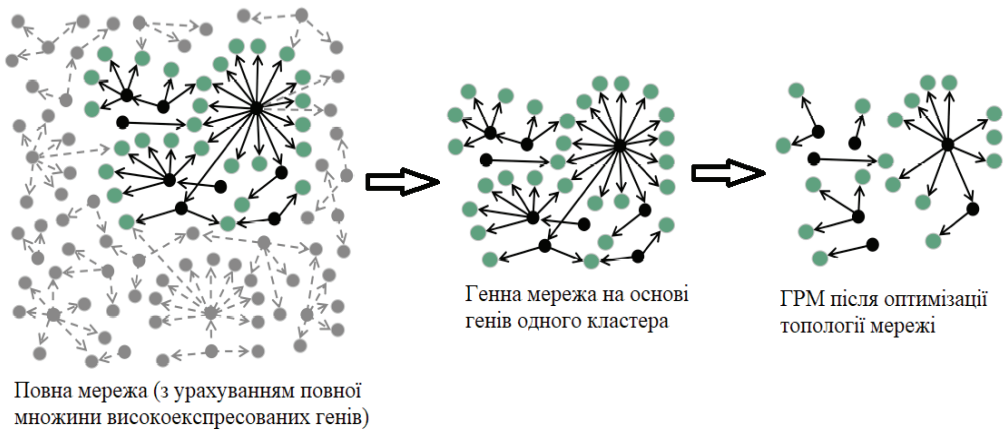


Рис. 1. Формальна схема процесу формування топології ГРМ

Отже, процес реконструкції ГРМ передбачає наявність таких етапів:

- Формування підмножини/підмножин інформативних взаємно-експресованих генів із застосуванням методів.
- Формування та оптимізація топології генної мережі шляхом кількісної оцінки параметрів, що визначають топологію мережі.
- Визначення цільових генів на основі аналізу даних з урахуванням хвороби, що досліджується, із застосуванням відповідних баз даних.
- Валідація ГРМ шляхом оцінки її адекватності при застосуванні тестових даних експресії генів і співставлення прогнозованого та реального станів об'єктів, що досліджуються.

Приклад реконструйованої ГРМ зображено на рис. 2 [3].

Аналіз наявних природних та штучних ГРМ дає змогу зробити висновок, що топологія мережі може бути представлена як орієнтований або неорієнтований граф, при цьому оптимізація топології мережі може бути виконана із застосуванням теорії графів, тоді як параметри можуть бути застосовані як одиничні характеристики топології мережі, так і статистичні характеристики, що визначають розподіл відповідних параметрів ГРМ.

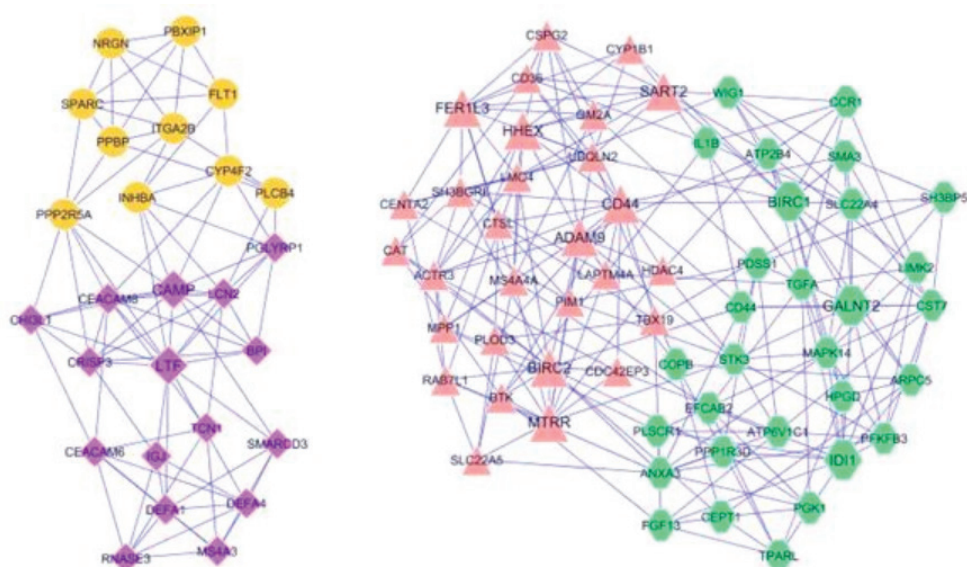


Рис. 2. Приклад реконструйованої генної регуляторної мережі

Аналіз останніх досліджень та публікацій. Вирішенню проблеми реконструкції генних регуляторних мереж на сьогодні присвячено велику кількість наукових праць [4–6]. Одним із ключових елементів ГРМ є транскрипційні фактори (ТФ). Оскільки клітини біологічного організму диференціюються на різні типи відповідно до функцій, які вони виконують, ТФ регулюють клітинну ідентичність, враховуючи тип відповідної клітини. Крім того, ТФ також контролюють зміни в експресії генів як реакцію на сигнали навколишнього середовища. ГРМ ідентифікує та фіксує механізми взаємодій між ТФ та генами, створюючи умови для проведення досліджень щодо складу та функцій тканин організму у контексті здоров'я та захворювань [7]. На рис. 3 зображений механізм взаємодії елементів у ГРМ [4]. Як можна побачити, ГРМ моделює регуляторний вплив ТФ на рівень експресії цільових генів. Кожен вузол в GRN представляє ТФ або ген, кожен край ребра відповідає регуляторному зв'язку між TF і цільовим геном. TF (жовтий вузол) зв'язується з послідовностями ДНК у промоторній області цільового гена (синій вузол) і має активуючий або інгібуючий регуляторний вплив на його транскрипцію (рис. 3 (а)). Ребро спрямоване від ТФ до цільового гена (рис. 3 (б)). Край може мати знак, що вказує, чи є ця взаємодія активуючою (+) чи інгібуючою (-).

Експериментальною основою для реконструкції ГРМ є масив експресії генів, отриманий шляхом ДНК-мікрочіпових експериментів або шляхом застосування більш сучасного і точного методу, заснованого на секвенуванні молекул РНК [8]. Якщо дані, що отримані шляхом застосування ДНК-мікрочіпових експериментів, дають змогу оцінити середні значення експресії генів у гетерогенній популяції типів клітин без урахування біологічних сигналів у профілях експресії генів окремих клітин, то застосування методу одноклітинного секвенування молекул

РНК (scRNA-seq) [9] дає змогу оцінити експресію генів в окремих клітинах без необхідності виділення кожного типу клітин. Однак потрібно зазначити, що обробка даних, отриманих шляхом секвенування одноклітинних молекул РНК, і реконструкція на їх основі ГРМ потребують великих обчислювальних ресурсів, оскільки необхідно врахувати стохастичні варіації експресії генів від клітини до клітини, зміни в експресії генів, що виникають на різних стадіях клітинного циклу, високу розрідженість через недостатню чутливість у секвенуванні транскриптів в окремих клітинах для генів з низькою експресією. Однак із різким зростанням обчислювальних можливостей метод реконструкції ГРМ на основі даних експресії генів, отриманих шляхом секвенування молекул РНК, стає більш популярним.

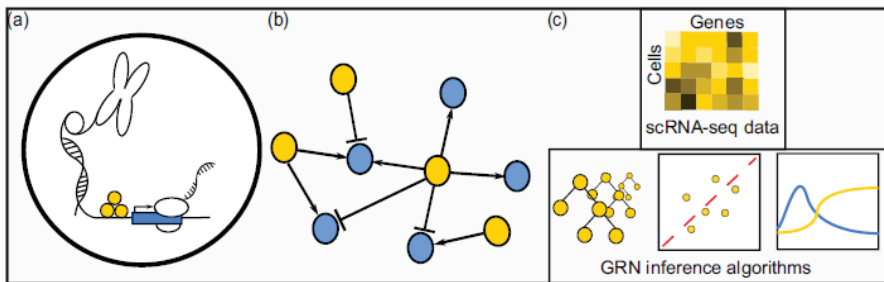


Рис. 3. Модель процесу реконструкції ГРМ на основі даних ланцюгів одноклітинних молекул РНК [4]

На цей час існує кілька підходів, що застосовуються для реконструкції ГРМ на основі даних експресії генів. У будь-якому випадку вхідні дані представляються як матриця, що кількісно відображає рівень експресії кожного гена клітини, що досліджуються (рис. 3 (c)). Результатом алгоритму реконструкції ГРМ є мережа прогнозованих зв'язків генів і ТФ (рис. 3 (b)). Кожне ребро може мати вагу, що представляє силу відповідного зв'язку. У публікації [10] автори навели огляд наявних методів реконструкції ГРМ на основі даних експресії генів. Однак потрібно зазначити, що проблема оптимізації топології ГРМ на цей час немає однозначного рішення.

Мета статті — теоретичні дослідження щодо оптимізації топології генної регуляторної мережі на основі комплексного застосування одиничних та розподілених параметрів мережі.

Виклад основного матеріалу дослідження. До одиничних параметрів мережі потрібно зарахувати такі характеристики:

- **Кількість вузлів мережі** — це параметр, що визначає загальну кількість генів/протеїнів/метаболітів, що складають мережу. Водночас потрібно зазначити, що загальна кількість експресованих генів у геномі біологічного організму становить приблизно 25000, при цьому приблизно половина з них має низький рівень експресії для всієї групи об'єктів, що досліджуються. Ці гени можуть бути видалені без суттєвої втрати інформації. Крім того, як показує аналіз (рис. 1), велика кількість генів може бути не зв'язаною з іншими генами

мережі або зв'язана попарно, що призводить до створення великої кількості неінформативних малих ГРМ з погляду хвороби, що досліджується. На етапі оптимізації топології мережі ці мінімережі також можуть бути видалені як неінформативні. Отже, кількість вузлів мережі визначає загальну кількість вузлів, що прямо або непрямо взаємно зв'язані між собою. Водночас потрібно зазначити, що процес реконструкції ГРМ з подальшою оптимізацією топології може призвести до виникнення декількох ГРМ, як показано на рис. 2. У цьому випадку кількість вузлів визначається для кожної мережі окремо.

- **Зв'язність (ступінь) вузла мережі** — визначає сумарну вагу зав'язків, що напряму з'єднують цей вузол із сусідніми вузлами мережі:

$$k_s = \sum_{i=1}^{n_s} w_{si}, \quad (1)$$

де n_s — кількість вузлів, що напряму зв'язані з вузлом s ; w_{si} — вага дуги, що зв'язує вузли s та i .

- **Середня зв'язність (ступінь) мережі** — визначається як середнє арифметичне значення зв'язності всіх вузлів мережі:

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

- **Максимальна зв'язність** — максимальне значення зв'язності всіх вузлів мережі:

$$k_{\text{max}} = \max(k_i), i = \overline{1, n}. \quad (3)$$

Потрібно зазначити, що зв'язність є одним із ключових параметрів, що визначає складність мережі. Аналіз характеру розподілу значення зв'язності від кількості вузлів, що мають відповідне значення зв'язності, дає змогу проаналізувати топологію мережі та визначити шляхи її оптимізації.

- **Щільність мережі** — відношення сумарної кількості ваг зав'язків мережі до максимально можливої кількості ваг зав'язків між вузлами мережі:

$$\text{Dens} = \frac{2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n \omega_{ij}}{n \cdot (n-1)}. \quad (4)$$

Потрібно зазначити, що значення вагового коефіцієнта ω_{ij} змінюється від 0 (відсутність зв'язку) до 1 (зв'язок максимальний). Значення щільності змінюється від 0 (вузли мережі не мають зав'язків) до 1 (повнозв'язна мережа). Зазвичай при реконструкції ГРМ із застосуванням відповідного алгоритму наявність або відсутність зав'язків визначається коефіцієнтом трешолдінгу, значення якого може варіюватися в обмеженому діапазоні. Збільшення значення коефіцієнта трешолдінгу сприяє зростанню ступеня розрідженості мережі, що зі свого боку може призвести до втрати важливих зав'язків між вузлами мережі. Зменшення значення трешолдінгового коефіцієнта призводить до наявності у мережі великої кількості зайвих зав'язків, що ускладнює інтерпретацію мережі.

- **Коефіцієнт кластеризації вузла** — визначається відношенням кількості реальних зав'язків у кластері, що містить відповідний вузол (центр кластера) до максимально можливої кількості зав'язків між вузлами кластера:

$$Clust_i = \frac{2 \cdot N_i}{n_i \cdot (n_i - 1)}, \quad (5)$$

де N_i — визначає кількість реальних зв'язків у i -му кластері; n_i — кількість вузлів, що становлять i -й кластер.

- **Коефіцієнт кластеризації мережі** — визначається як середнє арифметичне значення коефіцієнтів кластеризації всіх вузлів мережі:

$$Clust = \frac{1}{n} \sum_{i=1}^n Clust_i. \quad (6)$$

Потрібно зазначити, що цей параметр також (як і щільність) є критерієм повнозв'язності мережі. Його значення змінюється від 0 (відсутність зв'язків між вузлами) до 1 (мережа є повнозв'язною).

- **Коефіцієнт централізації** — визначає ступінь близькості мережі до зіркоподібної топології і розраховується за формулою:

$$Centr = \frac{n}{n-2} \left(\frac{k_{\max}}{n-1} - Dens \right). \quad (7)$$

Як було зазначено вище, однією із особливостей ГРМ є наявність генів концентраторів, які є ключовими при ідентифікації відповідної хвороби. Коефіцієнт концентрації є одним із параметрів, який визначає наявність або відсутність таких генів. Якщо мережа повнозв'язна, вузли-концентратори відсутні, і значення коефіцієнта централізації дорівнює 0. Більше значення цього параметра відповідає більш високому ступеню подібності мережі до зіркоподібної топології і, як наслідок, більшій здатності ідентифікувати цільові гени (гени концентратори).

- **Гетерогенність мережі** — як і коефіцієнт централізації, цей параметр виражає ступінь неоднорідності топології мережі і розраховується через дисперсію та середнє арифметичне значення зв'язностей вузлів мережі таким чином

$$Geter = \frac{\sqrt{\text{var}(k_i)}}{\text{mean}(k_i)}, k = \overline{1, n}. \quad (8)$$

Більш високе значення цього параметра свідчить про більшу різницю значень зв'язності вузлів мережі і, як наслідок, про більший ступінь неоднорідності мережі. Повнозв'язна ГРМ має нульове значення гетерогенності.

- **Топологічний коефіцієнт** вузлів мережі визначається за формулою:

$$T_p = \frac{\frac{1}{n} \sum_{i=1}^n L(p, i)}{k_p}, \quad (9)$$

де k_p — кількість сусідів p -го вузла; n — кількість вузлів, що мають спільних сусідів із вузлом p ; $L(p, i)$ — кількість сусідів i -го гена, що мають хоча б одного спільного сусіда з геном p .

Аналіз вищенаведених одиночних критеріїв оцінки топології ГРМ дає змогу зробити висновок, що характер їх зміни у процесі розрядження мережі може суперечити один одному. Так, якщо значення коефіцієнтів кластеризації мережі та щільності мережі зменшується у процесі формування оптимальної топології мережі шляхом видалення зв'язків з низькою вагою (шляхом застосування відповідного трешолдінгового коефіцієнта), то значення гетерогенності та коефіцієнта

централізації зростають. У цьому випадку виникає необхідність у пошуку компромісного рішення щодо значення коефіцієнта трешолдінгу, враховуючи оптимальні значення кількості вузлів мережі та її зв'язності.

У табл. 1 наведені значення одиночних параметрів ГРМ та підмережі, яка містить більшу кількість генів для мережі, що зображена на рис. 4. Ця мережа є однією із генних мереж пацієнтів, що досліджуються на рак простати і входить до атласу геному раку (TCGA — The Cancer Genome Atlas Program) [11]. Як зображено на рис. 4, вона містить основну мережу і малі підмережі, які зв'язують декілька генів і можуть бути видалені, оскільки не є визначальними для ідентифікації відповідної хвороби.

Таблиця 1

Генна регуляторна мережа пацієнтів, що досліджуються на рак простати

Кількість вузлів	Кількість зв'язків	Коефіцієнт кластеризації	Щільність мережі	Гетерогенність	Коефіцієнт централізації
2713	3087	0.102	0.001	1.370	0.037
2563	2814	0.096	0.00089	1.389	0.042

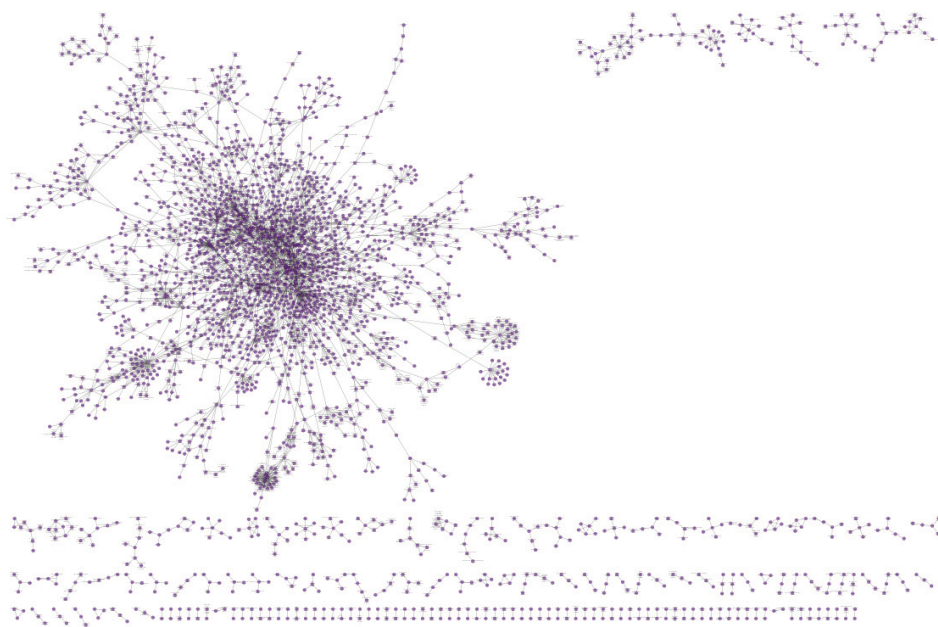


Рис. 4. Реконструйована ГРМ пацієнтів, що досліджуються на рак простати

Крім того, частина генів не мають зв'язків із сусідами, що також дає змогу видалити ці вузли як неінформативні. Перший рядок табл. 1 містить загальну статистику даних повної мережі, а другий рядок — статистичні дані для виділеної підмережі, що містить найбільшу кількість взаємозв'язаних генів.

На рис. 5 зображено характер розподілу коефіцієнта кластеризації вузлів мережі та топологічного коефіцієнта від зв'язності (ступеня) відповідних вузлів. У цьому випадку було досліджено повну ГРМ, що зображена на рис. 3. Аналіз отриманих діаграм дає змогу зробити висновок, що характер залежності більшості генів, що прямо або непрямо зв'язані між собою (основна мережа), є нелінійним, і значення обох параметрів різко зменшується на початковому етапі збільшення ступеня вузлів. Далі швидкість зміни значень цих параметрів суттєво зменшується при збільшенні значення зв'язності (ступеня). Значення як коефіцієнта кластеризації, так і топологічного коефіцієнта вузлів, що не входять в основну мережу, майже не змінюються при збільшенні ступеня вузлів.

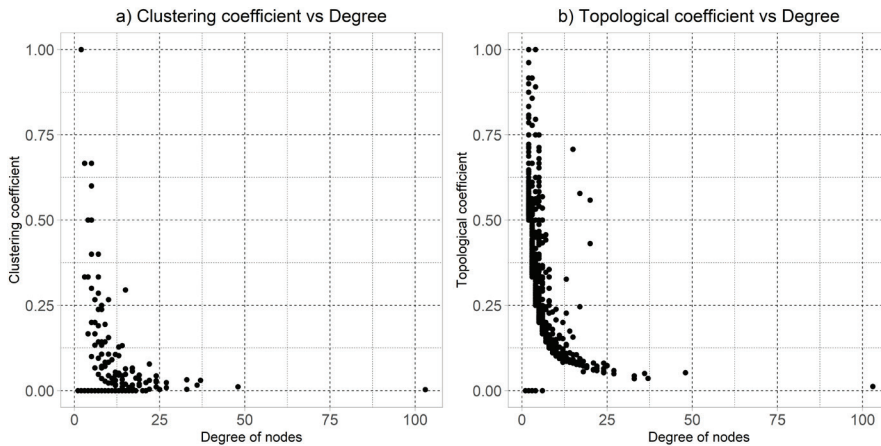


Рис. 5. Діаграми залежності коефіцієнта кластеризації та топологічного коефіцієнта вузлів ГРМ від значення зв'язності (ступеня) вузлів

Потрібно зазначити, що загальною характеристикою ГРМ багатьох біологічних організмів є їх безмасштабна топологія або розподіл вузлів за їх ступенем, яку можна описати такою функцією [12]:

$$P(k) = Ck^{-\alpha}, \quad (10)$$

де $P(k)$ — ймовірність, що випадково обраний вузол має зв'язність (ступінь) k ; α є ступеневим коефіцієнтом масштабування; константа C є дзета-функцією Рімана, яка нормалізує розподіл ймовірностей за ступеневим законом так, що

$$\sum_{k=1}^{\infty} P(k) = 1. \quad (11)$$

Як було зазначено вище, у ГРМ більшість вузлів мають порівняно невелику кількість взаємодій, що проявляється як нижчий ступінь цих вузлів, тоді як невелика кількість вузлів (концентратори) мають вищий ступінь.

Отже, процес оптимізації топології мережі передбачає наявність двох етапів:

1. Формування структури і топології мережі за групою одиночних критеріїв, таких як коефіцієнт кластеризації мережі, щільність мережі, коефіцієнт централізації та гетерогенність із застосуванням комплексного критерію, який містить окремі критерії як компоненти. При цьому структура та топологія мережі формується

шляхом оптимізації відповідного коефіцієнта трешолдінгу, що визначає наявність або відсутність зв'язків між відповідними вузлами.

2. Оцінка відповідності мережі ступеневому закону (10) шляхом аналізу розподілених параметрів, що наведені на рис. 5.

Висновки. Наведено результати теоретичних досліджень щодо формування топології генної регуляторної мережі. Зроблений аналіз одиночних та розподілених топологічних параметрів мережі, що створює умови для формування узагальненого топологічного параметра, який може бути застосований як трешолдінговий коефіцієнт при формуванні оптимальної топології мережі. На основі аналізу біологічної генної мережі пацієнтів, що досліджуються на рак простати, побудовані діаграми залежності коефіцієнта кластеризації вузлів мережі та топологічного коефіцієнта від зв'язності (ступеня) вузлів, аналіз яких створює умови для оптимізації топології мережі шляхом зіставлення закону зміни відповідних параметрів реконструйованої та біологічної мереж.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Reg Network: an integrated data base of transcriptional and post-transcriptional regulatory networks in human and mouse / Liu Z. P., Wu C., Miao H., Wu H. *Database*. 2015. Pp. 1–12.
2. Ouma W. Z., Pogacar K., Grotewold E. Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties. *PLoS Computational Biol.* 2018. Vol. 14 (4), art. no. e1006098.
3. Gene Regulatory Network. URL: <https://www.ese.wustl.edu/~nehorai/research/genomic/grn.html>.
4. Akers K., Murali T. M. Gene regulatory network inference in single-cell biology. *Current Opinion in Systems Biology*. 2021. Vol. 26. Pp. 87–97.
5. Liu E., Li L., Cheng L. Gene regulatory network revive. *Encyclopedia of Bioinformatics and Computational Biology*. 2019. Vol. 2. Pp. 155–164.
6. Gene Regulatory Network Inference: Connecting Plant Biology and Mathematical Modeling / Van den Broeck L., Gordon M., Inzé D., Williams C., Sozzani R. *Frontiers in Genetics*. 2020. Vol. 11, art no 457.
7. Wagner A., Regev A., Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Natural Biotechnology*. 2016. Vol. 34. Pp. 1145–1160.
8. Marbach D., Costello J., Küffner R. et al. Wisdom of crowds for robust gene network inference. *Natural Methods*. 2012. Vol. 9. Pp. 796–804.
9. Tang F., Barbacioru C., Wang Y. et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Natural Methods*. 2009. Vol. 6. Pp. 377–382.
10. Akers K., Murali T. M. Gene regulatory network inference in single - cellbiology. *Current Opinion in Systems Biology*. 2021. Vol. 26. Pp. 87–97.
11. Chu Y., Corey D. R. RNA sequencing: Platform selection, experimental design, and data interpretation. *Nucleic Acid Therapeutics*. 2012. Vol. 22 (4). Pp. 271–274.
12. Techniques of DNA microarray data pre-processing based on the complex use of Bioconductor too Isand Shannonentropy / Babichev S., Durnyak B., Zhydetsky V., Pikh I., Senkivsky V. CEUR Workshop Proceedings. 2019. Vol. 2353. Pp. 365–377.

REFERENCES

1. Liu, Z. P., Wu, C., Miao, H., & Wu, H. (2015). Reg Network: an integrated data base of transcriptional and post-transcriptional regulatory networks in human and mouse: Database, 1–12 (in English).
2. Ouma, W. Z., Pogacar, K., & Grotewold, E. (2018). Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties: PLoS Computational Biol, 14 (4), art. no. e1006098 (in English).
3. Gene Regulatory Network. Retrieved from <https://www.ese.wustl.edu/~nehorai/research/genomic/grn.html> (in English).
4. Akers, K., & Murali, T. M. (2021). Gene regulatory network inference in single-cell biology: Current Opinion in Systems Biology, 26, 87–97 (in English).
5. Liu, E., Li, L., & Cheng, L. (2019). Gene regulatory network revive: Encyclopedia of Bioinformatics and Computational Biology, 2, 155–164 (in English).
6. Van den Broeck L., Gordon, M., Inzé, D., Williams, C., & Sozzani, R. (2020). Gene Regulatory Network Inference: Connecting Plant Biology and Mathematical Modeling: Frontiers in Genetics, 11, art no 457 (in English).
7. Wagner, A., Regev, A., & Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics: Natural Biotechnology, 34, 1145–1160 (in English).
8. Marbach, D., Costello, J., & Küffner, R. et al. (2012). Wisdom of crowds for robust gene network inference: Natural Methods, 9, 796–804 (in English).
9. Tang, F., Barbacioru, C., Wang, Y. et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell: Natural Methods, 6, 377–382 (in English).
10. Akers, K., & Murali, T. M. (2021). Gene regulatory network inference in single – cellbiology: Current Opinion in Systems Biology, 26, 87–97 (in English).
11. Chu, Y., & Corey, D. R. (2012). RNA sequencing: Platform selection, experimental design, and data interpretation: *Nucleic Acid Therapeutics*, 22 (4), 271–274 (in English).
12. Babichev, S., Durnyak, B., Zhydetsky, V., Pikh, I., & Senkivsky, V. (2019). Techniques of DNA microarray data pre-processing based on the complex use of Bioconductor too l sand Shannonentropy. CEUR Workshop Proceedings, 2353, 365–377 (in English).

doi: 10.32403/0554-4866-2023-1-85-40-50

THEORETICAL STUDIES REGARDING THE FORMATION OF THE GENE REGULATORY NETWORK OPTIMAL TOPOLOGY

I. M. Liakh

*Uzhhorod National University,
3, Narodna Square, Uzhhorod, 88000, Ukraine
ihor.lyah@uzhnu.edu.ua*

This work presents the results of theoretical studies on the formation of the optimal topology of the gene regulatory network, which can be presented in the form of an oriented

or unoriented graph. The topological parameters, the totality of which determines the network topology, are explored. It is shown that the nature of changes of various topological parameters in the process of network discharge can contradict each other, which complicates the process of network topology optimization. In this case, there is a need to find a compromise solution regarding the value of the thresholding coefficient, considering the optimal values of the number of network nodes and its connectivity. Diagrams, that are created, determine the nature of the distribution of the clustering coefficient of network nodes and the topological coefficient of the connectivity (degree) of the corresponding nodes. In this case, the complete HRM of patients studied for prostate cancer and included in the cancer genome atlas (TCGA - The Cancer Genome Atlas Program) is researched. The analysis of the resulting diagrams allows one to conclude that the nature of the dependence of the majority of genes that are directly or indirectly connected to each other (the main network) is non-linear and the value of both parameters sharply decreases at the initial stage of increasing the degree of nodes, and then the rate of change of the values of these parameters decreases significantly, when the value of connectivity (degree) increases. The values of both the clustering coefficient and the topological coefficient of nodes, that are not included in the main network, almost do not change when the degree of nodes increases. The ways of optimizing the topology of the gene regulatory network are determined, which involves the formation of the structure and topology of the network according to a group of single criteria, such as the clustering coefficient networks, network density, centralization coefficient and heterogeneity using a complex criterion containing separate criteria as components in the first stage. At the same time, the structure and topology of the network are formed by optimizing the corresponding threshold coefficient, which determines the presence or absence of connections between the corresponding nodes. At the second stage, the conformity of the network to the power law is assessed by analyzing the corresponding distributed parameters.

Further research prospects include the development of information technology for the reconstruction of gene regulatory networks for various types of gene expression data.

Keywords: gene expression, gene regulatory network, topological parameters, network topology optimization.

Стаття надійшла до редакції 19.04.2023.

Received 19.04.2023.