ЖУРНАЛ НАНО- ТА ЕЛЕКТРОННОЇ ФІЗИКИ *JOURNAL OF NANO- AND ELECTRONIC PHYSICS*
Том **4** № 2, 02027(4сс) (2012) Vol. **4** No 2, 02027(4pp) (2012)

Кореляційні ефекти в біологічних мережах

А.А. Багдасарян*, В.Н. Борисюк

Сумський державний університет, вул. Римського-Корсакова, 2, 40007 Суми, Україна

(Одержано 13.12.2011, у відредагованій формі – 27.05.2012, опубліковано online 04.07.2012)

Проведено огляд теорії складних мереж та розглянуто їх класифікацію за основними статистичними характеристиками. Для матриці суміжності реально існуючої нейронної мережі розраховані найкоротші відстані для кожної пари вузлів мережі, а також розподіли вузлів по ступеням та коефіцієнти кластеризації. Проведено порівняння основних статистичних параметрів з випадковою мережею, виходячи з чого зроблені висновки про кореляційні явища в біологічній системі.

Ключові слова: складна мережа, кореляція, коефіцієнт кластерності, матриця суміжності.

PACS numbers: 02.10.Ox, 89.75.Hc

1. ВСТУП

В останніх дослідженнях, теорію мереж почали використовувати як інструмент для опису та теоретичного обґрунтування реальних складних систем. Ця обставина обумовлена наявністю великої кількості об’єктів, що мають мережеву будову: Інтернет, екологічні системи, нейронні мережі, мережі наукового співробітництва, соціальні мережі, мережі цитування та ін. Особливість будови таких систем у порівнянні з регулярними мережами полягає у наявності так званих габів (від англійського hub – ядро). Це невелика кількість вузлів з великою кількістю зв’язків.

Структура мережі вивчалась за допомогою математичної теорії графів, яка в основному зосереджувала свою увагу на вивченні графів невеликих розмірів. Саме в останні роки ми стали свідками появи нового руху в науково-дослідній мережі, тобто зміщення уваги від вивчення одного невеликого графу та властивостей окремих вершин та ребер, до розгляду статистичних властивостей зв’язків графів. Цей новий підхід був обумовлений в значній мірі появою комп’ютерів і комунікаційних мереж, які дозволяють нам збирати й аналізувати дані набагато більше, ніж було можливо раніше. Все це дозволило отримати велику кількість дивних та інтригуючих властивостей (різні закони розподілу вузлів по кількості зв’язків, нові підходи отримання уявлення про структуру мереж). Тому на сьогодні вивченню складних мереж приділяють особливу увагу та присвячують цілі журнали чи секції журналів.

2. ОСНОВНІ ПОНЯТТЯ З ТЕОРІЇ СКЛАДНИХ МЕРЕЖ

Мережею називається сукупність вузлів, що поєднані зв’язками (див. рис. 1). У сучасній теорії мереж число зв’язків вузла (у теорії графів ребра та вершини відповідно) називається ступенем вузла (degree).

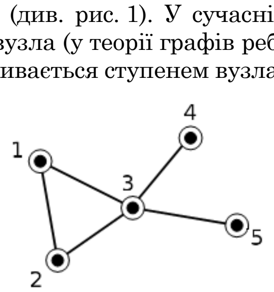


Рис. 1 – Граф, який має $N = 5$ вершини та $M = 5$ ребер

Згідно рис. 1 вершини 1, 2 мають ступінь два ($k_1 = k_2 = 2$), вершини 4, 5 відповідно ступінь один ($k_4 = k_5 = 1$), а вершина 3 – степінь 4 ($k_4 = 4$). Поняття ступінь є локальною характеристикою графа. Цілісну ж характеристику графа визначають двома поняттями – шлях і петля (цикл). Шлях – це послідовність суміжних вузлів і зв'язків між цими вузлами, коли вузли не повторюються. Петлею називається шлях, коли початковий і кінцевий вузли збігаються. Мережі без циклів називаються деревами. Число вузлів N (зване розміром мережі) і число зв'язків L в деревах пов'язані простим співвідношенням $N = L - 1$.

Існують різні типи зв’язків – орієнтовані та неорієнтовані. В залежності від цього розглядають вхідний (in-degree) та вихідний (out-degree) ступені відповідно. Більш повну інформацію про мережу дає нам матриця суміжності (adjacency matrix). Мережа з N вузлів описується квадратною матрицею суміжності a розмірності $N \times N$, в якій ненульові елементи матриці відповідають наявності зв’язків між відповідними вузлами. Її елементи a_{ij} дорівнюють 1, якщо вузли i та j з’єднані між собою, та 0, якщо ці вузли не з’єднані [1]. Для неспрямованих мереж недіагональний елемент a_{ij} дорівнює кількості зв’язків між вузлами i та j, тобто така матриця є симетричною. Діагональний же елемент a_{ij} дорівнює 0. Виходячи з цього всього, степінь k_i має вигляд [2]:

$$k_i = \sum_j a_{ij}. \qquad (1)$$

Чисельні розрахунки з матрицями суміжності для мереж великих розмірів вимагають значних ресурсів пам’яті комп’ютера. Але так як більшість реальних мереж є розрідженими, то часто розрахунки можна спростити. Розріджена мережа представляє собою мережу, в якій кількість зв’язків значно менше, ніж у повнозв’язній ($\langle k_i \rangle << N$). Таким чином більшість елементів матриці суміжності реальних мереж дорівнює 0.

Щоб охарактеризувати “лінійний розмір” мережі використовують поняття середнього $\langle l \rangle$ та максимального l_{max} найкоротших шляхів. Середній коротший шлях для мережі з N вузлів визначається наступним чином:

$$\langle l \rangle = \frac{2}{N(N-1)} \sum_{i>j} l_{ij}, \qquad (2)$$

* artemsumdu@ukr.net

де l_{ij} – довжина найкоротшого шляху між . Максимальний найкоротший шлях відповідно буде дорівнювати найбільшому значенню l_{ij} даної мережі.

Однією з локальних характеристик мережі є кластеризація. Вона характеризує степінь взаємозв’язку між двома сусідніми вузлами. Наприклад, якщо вузол A з’єднаний з вузлом – B, а вузол B – з вузлом C, то існує велика ймовірність, що вузли A і C також зв’язані. Тобто коефіцієнт кластеризації дає нам імовірність того, що два найближчих сусіди цього вузла самі є найближчими сусідами. Іншими словами, якщо вузол i має q_i найближчих сусідів з числом зв’язків l_j, то коефіцієнт кластеризації дорівнює [3-5]:

$$C_j(q_j) = \frac{2t_j}{q_j(q_j-1)}, \qquad (3)$$

де t_j є сумарна кількість трикутників – циклів довжини 3, прикріплених до вузла j, а $q_i(q_i-1)/2$ – максимально можливе число трикутників. Середнє значення коефіцієнта кластеризації має вигляд:

$$\langle C \rangle = \frac{1}{N}\sum_i C_i \qquad (4)$$

Таким чином коефіцієнт кластеризації характеризує статистику тріад у мережі. З означення (3) випливає, що якщо між сусідніми вузлами є зв’язок, то коефіцієнт кластеризації дорівнює 1. Якщо ж між ними немає зв’язку (відсутні цикли), то $C_j = 0$.

Однією з найважливіших статистичних характеристик мережі є розподіл ступенів вузлів. Знання цієї характеристики достатньо для розуміння властивостей мережі та процесів, які протікають у ній [4-6]. Розподілом ступенів вузлів є ймовірність $P(k)$ того, що вузол i має ступінь $k_i = k$. Дослідження показали, що мають місце різні $P(k)$, відповідно і мережі демонструють різну поведінку. Деякі, найчастіше спостережувані, приклади розподілу ступенів вузлів мережі, а саме розподіл Пуассона:

$$P(k) = e^{-\langle k \rangle}\frac{\langle k \rangle^k}{k!}, \qquad (5)$$

степеневий розподіл:

$$P(k) \sim 1/k^{\gamma}, k \neq 0, \gamma > 0, \qquad (6)$$

експоненціальний розподіл:

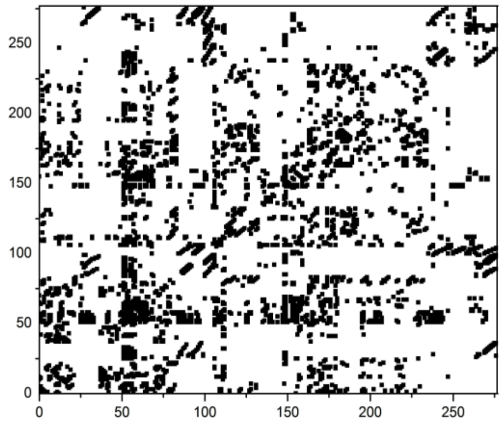
$$P(k) \sim \exp\left(-\frac{k}{\langle k \rangle}\right). \qquad (7)$$

Принципова відмінність між цими розподілами полягає у тому, що розподіли Пуассона та експоненціальний характеризуються різним масштабом. Тому мережі, які підкоряються степеневому розподілу називають безмасштабними. Ще однією відмінністю є “швидкість” спадання розподілу вузлів за кількістю зв’язків. Як видно з (6) ступеневий розподіл є одним з прикладів повільно спадаючих розподілів. Саме більшість реально існуючих мереж мають степеневий розподіл вузлів і в цих мережах габи (вузли з великою кількістю зв’язків) складають помітну частку від всіх вузлів, і саме вони визначають багато важливих властивостей цих мереж.

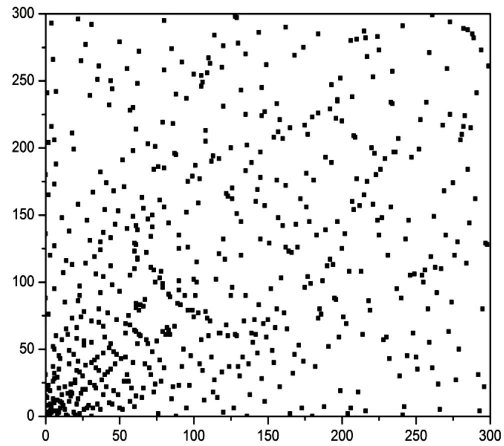
3. ДОСЛІДЖУВАНА МЕРЕЖА

Одним із найпоширеніших типів мереж є, так звані, нейронні мережі. Головною їх відмінністю від складних мереж є властивості, які не настільки визначаються топологією мереж, як вагою зв’язків та станом вузлів. Найбільш вивченим представником складних мереж є Caenorhabditis elegans – вільноживуча нематода (круглий черв'як) довжиною близько 1 мм. Саме завдяки вивченню нейронної мережі C. Elegans, де можливість повного аналізу існує завдяки своїй простоті, можна глибше зрозуміти більш складні нейронні мережі [7]. Щоб спростити вивчення складної мережі, необхідно знайти відповідні параметри, які відрізняють просту мережу від випадкової. В якості випадкової мережі ми взяли мережу Барабаші-Альберт (при малих N вона представляє собою випадкову мережу [4]). Тому ми почали з візуального подання матриць суміжності двох мереж.

На рис. 2 представлені матриці суміжності нейронної мережі та мережі Барабаші-Альберт. Кожна точка в матриці відповідає зв’язку в мережі. Коли точка існує в позиції (x, y), то це означає, що існує зв'язок від вузла x до y. З наведених вище матриць видно, що випадкова мережа рівномірно розподіляється, в той час як мережа C. Elegans, ні. Дивлячись на мережу C. Elegans, можна помітити, що є видимі



a



b

Рис. 2 – Матриця суміжності нейронної мережі (279 вузлів) (a), мережа Барабаші-Альберт (300 вузлів) (b) лінії на графіку, це відбувається, коли існує у мережі. Центри існують у будь-якій природній мережі, тобто є завжди ті вузли, які мають більше вхідних чи вихідних зв’язків. Є також вузли, які не мають багато зв'язків – це ті рядки, які майже повністю пусті.

Розглянемо тепер більш складну матрицю, а саме матрицю довжин зв’язків.

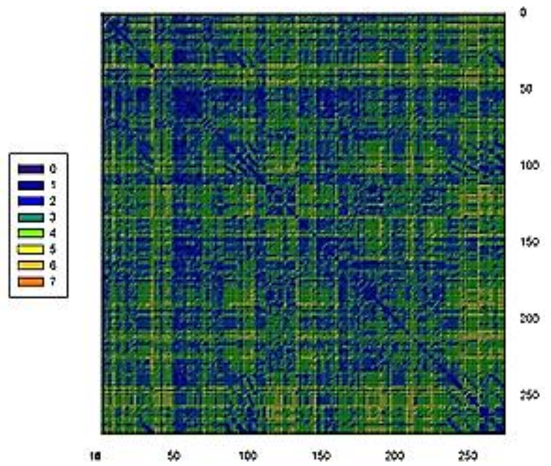


Рис. 3 – Матриця відстаней нейронної мережі C. Elegans ($\langle l \rangle = 2.79$, $l_d = 7$)

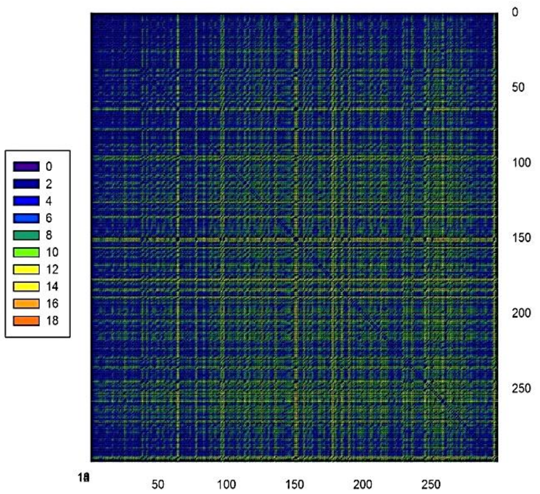


Рис. 4 – Матриця відстаней мережі Барабаші-Альберт ($\langle l \rangle = 5.41$, $l_d = 18$)

Аналізуючи рис. 4 та рис. 5, можна порівняти “лінійні розміри ” мереж, а тобто поняття середніх найкоротших шляхів та діаметр мережі. Обидві характеристики є глобальними і дають уявлення про мережу. Порівнюючи дані для обох мереж, можна відмітити, що малий середній шлях та діаметр вказують на те, що нейронна мережа представляє собою тісний світ. Локальною ж характеристикою мережі є середній коефіцієнт кластерності, який показаний на рис. 5.

Даний коефіцієнт містить інформацію про наявність у мережі циклів довжиною 3. Наявність циклів є специфічною формою кореляції в мережах, а значення цього коефіцієнта вказує на присутність кореляції, тоді як для мережі Барабаші-Альберт кореляція практично відсутня.

Раніше було показано [3], що в складних мережах можлива наявність кореляцій у зв’язках з різними

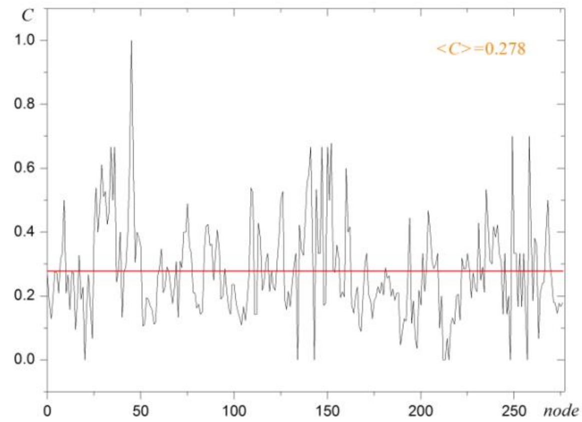


Рис. 5 – Середній коефіцієнт кластерності нейронної мережі

значеннями k. За характером кореляцій складні мережі поділяються на асортативні, де зв’язки встановлюються переважно між вузлами з однаковою кількістю зв’язків, та диасортативні, де вузли з великою кількістю зв’язків (габи) зв’язані з найменшим числом зв’язків. Іншими словами ця характеристика є мірою тенденції вузлів мережі виявитися з'єднаними з іншими вузлами з одним і тим же числом зв'язків, чи з різними ступенями. Яскравим прикладом асортативного змішування є соціальні мережі, біологічні ж загалом відносяться до диасортативного.

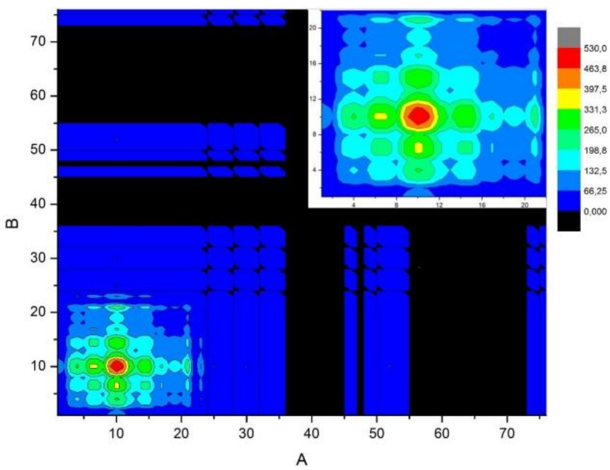


Рис. 6 – Кількість зв’язків між вузлами з різними ступенями (нейронна мережа)

Як видно з рис. 6 переважна кількість зв’язків знаходиться в області перетину вузлів, які мають степінь від 4 до 16. Але в мережі також присутні вузли з великою кількістю зв’язків, які з’єднані з вузлами з різними ступенями. Все це вказує, про те що дана мережа має змішану міру з’єднання.

Аналізуючи рис. 7, видно, що мережа Барабаші-Альберт має також змішану міру з’єднання, але вона є більш диасортативною ніж біологічна мережа. На це вказує колір в верхньому лівому та нижньому правому кутах. Це є не що інше як габ, тобто вузли з великою кількістю зв’язків. Більшість її зв’язків з’єднує вузли зі ступенем, що коливається від 1 до 6. Кількість же зв’язків, яка приходиться на цю область є набагато меншою ніж у біологічній мережі.

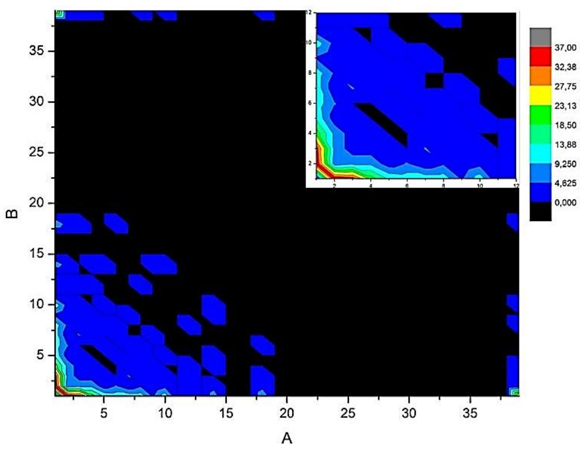


Рис. 7 – Кількість зв'язків між вузлами з різними ступенями (мережа Барабаші-Альберт)

4. ВИСНОВОК

У результаті чисельного аналізу проведено порівняння статистичних параметрів біологічної мережі з мережею Барабаші-Альберт. Основні характеристики вказують на присутність кореляційних явищ у біологічній мережі, а саме наявність значення коефіцієнта кластеризації. Тоді як для мережі Барабаші-Альберт кластеризація відсутня. Малі значення характеристик, які відповідають за "лінійний розмір" мережі вказують на наявність ефекту «тісного світу» (small world). Взагалі для реально існуючих мереж велика скорельованість та середній коефіцієнт кластерності є типовими.

Корреляционные эффекты в биологических сетях

А.А. Багдасарян, В.Н. Борисюк

Сумский государственный университет, ул. Римского-Корсакова, 2, 40007 Сумы, Украина

Проведен обзор теории сложных сетей и рассмотрена их классификация согласно основных статистических характеристик. Для матрицы смежности реально существующей нейронной сети рассчитаны кратчайшие расстояния для каждой пары узлов сети, а также распределения узлов по ступеням и коэффициенты кластеризации. Проведено сравнение основных статистических параметров со случайной сетью, исходя из чего сделаны выводы о корреляционных явлениях в биологической системе.

Ключевые слова: сложная сеть, корреляция, коэффициент кластерности, матрица смежности.

Correlation Effects in Biological Networks

A.A. Bagdasaryan, V.N. Borisyuk

Sumy State University, 2, Rimsky-Korsakov Str., 40007 Sumy, Ukraine

Review of the complex network theory is presented and classification of such networks in accordance with the main statistical characteristics is considered. For the adjacency matrix of a real neural network the shortest distances for each pair of nodes as well as the node degree distribution and cluster coefficients are calculated. Comparison of the main statistical parameters with the random network is performed, and based on this, the conclusions about the correlation phenomena in biological system are made.

Keywords: Complex network, Correlation, Cluster coefficient, Adjacency matrix.

СПИСОК ЛІТЕРАТУРИ

1. Ю. Головач, О. Олемской, К. Фон Фербер, Т. Головач, О. Мриглод, І. Олемской, В. Пальчиков, *Журнал фізичних досліджень* **10** No4, 247 (2006) (Yu. Holovatch, O. Olemskoi, C. von Ferber, T. Holovatch, O. Mryglod, I. Olemskoi, V. Palchykov, *Journal of Physical Studies* **10** No 4, 247 (2006)).
2. И.А. Евин, *Компьютерное исследование и моделирование* **2** No2, 121 (2010) (I.A. Yevin, *Computer Research and Modeling* **2** No2, 121 (2010)).
3. M.E.J. Newman, arXiv:cond-mat/0303516v1.
4. R. Albert, A.-L. Barabasi, arXiv.org:cond-mat/0106096v1.
5. S.N. Dorogovtsev, *Lectures on Complex Networks* (Oxford University Press: 2010).
6. M.E.J. Newman, A.-L. Barabasi, D.J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press: Princeton: 2006).
7. Y. Choe, B.H. McCormick, W. Koh, *Network connectivity analysis on the temporally augmented C. Elegans web: A pilot study*. Society of Neuroscience Abstracts 30:921.9 (2004).