

РЕЦЕНЗІЯ

**доктора технічних наук, доцента,
доцента кафедри комп'ютерних наук
навчально-наукового інституту фізико-технічних та комп'ютерних наук
Чернівецького національного університету імені Юрія Федьковича**

УГРИНА Дмитра Ілліча

**на дисертаційну роботу Кириченко Оксани Леонідівни
«Дослідження статистичних характеристик складних мереж методами
інтелектуального аналізу даних», яка подана на здобуття наукового
ступеня доктора філософії за спеціальністю 121 – «Інженерія програмного
забезпечення», галузь знань 12 «Інформаційні технології»**

Актуальність дисертаційного дослідження

Більшість об'єктів природи і суспільства сьогодні можна представити у вигляді складних мереж. Форма мережі притаманна багатьом системам, зокрема WWW, Інтернету, транспортним, нейронним, інформаційним, телекомунікаційним, соціальним мережам, мережі цитування та багатьом іншим. Топологічні властивості таких мереж розглядаються незалежно від їхньої фізичної природи, істотно визначають їх функціонування, відіграють провідну роль як предмет дослідження складних мереж. Вивчення статистичних характеристик складних мереж сприяє кращому розумінню організації таких мереж, їх використанню в різних компаніях та галузях, адже сфера застосування таких мереж досить широка – економіка, менеджмент, соціологія, біологія, медицина, криміналістика, кібербезпека. Для проведення теоретичних і практичних досліджень, які сьогодні здійснюються в різних галузях, що описуються складними мережами, застосовують методи інтелектуального аналізу даних. Дисертаційна робота Кириченко Оксани Леонідівни присвячена дослідженням реальних підмереж веб-простору; розробці інформаційної технології для збирання статистичних даних, проведення статистичного та кластерного аналізу цих сегментів; визначення оптимальної кількості кластерів відомими методами; розробці нового методу визначення оптимальної кількості кластерів у складних мережах та перевірці адекватності запропонованого методу визначення оптимальної кількості кластерів за допомогою методів моделювання. Зважаючи на це, рецензована

кваліфікаційна робота Кириченко Оксани Леонідівни є повноцінним науковим дослідженням, актуальність тематики якого не викликає сумнівів.

Зв'язок роботи з державними програмами, планами, темами

Дисертаційне дослідження Кириченко О.Л. відповідає тематиці науково-дослідних робіт «Математичне та програмне забезпечення обчислювальних систем» (Державний реєстраційний номер 011U007046) та «Дослідження, моделювання та розробка програмного забезпечення складних динамічних систем» (Державний реєстраційний номер 0121U109232), які виконувались на кафедрі програмного забезпечення комп'ютерних систем навчально-наукового інституту фізико-технічних та комп'ютерних наук Чернівецького національного університету імені Юрія Федьковича. Тема дисертації відповідає науковому напрямку кафедри. Дослідження та результати, які викладено в дисертації, виконано відповідно до пріоритетних напрямків науково-дослідних робіт кафедри.

Ступінь обґрунтованості наукових положень, висновків, рекомендацій, сформульованих у дисертації

Аналіз основних положень, висновків та результати дисертації свідчать, що автор володіє широким методологічним апаратом, який застосовувався при проведенні досліджень (класичні математичні методи, статистичні методи, кластерний аналіз, теорія матриць).

Ступінь обґрунтованості наукових положень значною мірою зумовлюється логікою і структурою дослідження. Достовірність і обґрунтованість результатів роботи забезпечується:

- науковістю методологічного інструментарію дослідження на філософському, загальнонауковому, конкретно-науковому, а також технологічному та технічному рівнях, а відтак використанням ефективних, релевантних меті та завданням підходів і методів дослідження, їх виваженням та коректним застосуванням;

- достатньою та відповідною тематиці роботи джерельною базою дослідження (список використаних джерел налічує 252 найменування);
- системністю здійсненого дослідження зон веб-простору, інтелектуального аналізу статистичних даних сегментів веб-простору, порівняльного аналізу алгоритмів знаходження оптимальної кількості кластерів, побудови відповідного критерію їх оцінки за обґрунтованим авторкою алгоритмом перевірки належності елементів (вершин графу) до одного кластеру;
- належною апробацією результатів дослідження на міжнародних та всеукраїнських науково-практичних конференціях (15 тез конференцій), участю у роботі Міжнародної літньої школи ECODAM (м. Ясси, Румунія, 2019 р.), доповіддю на Міжнародному науковому семінарі «Advances & Challenges in Computing (A2C)» (Тернопіль, 2022р.);
- впровадженням результатів дисертаційної роботи у навчальний процес відділу комп'ютерних технологій навчально-наукового інституту фізико-технічних та комп'ютерних наук Чернівецького національного університету імені Юрія Федьковича;
- успішним впровадженням результатів дослідження у роботі компаніями «Квант Азимут» та «Qlicks B.V.».

Результати дослідження достатньо повно відображено в 22 публікаціях авторки: 1 стаття – в журналі, що індексується у наукометричній базі SCOPUS, 5 статей – у фахових виданнях України, 1 – розділ монографії міжнародного видання, 15 – тези конференцій.

Структура дисертації

Дисертація має традиційну структуру – анотація, вступ, чотири розділи, висновки, перелік використаних джерел, чотири додатки, список публікацій дисертантки (22 публікації). Обсяг основного тексту відповідає встановленим вимогам. Повний обсяг роботи становить 229 сторінок.

Актуальність задачі дослідження, мета, завдання, предмет, об'єкт та методи дослідження, наукова новизна, теоретичне, практичне значення отриманих результатів та структура дисертаційної роботи обґрунтовані у вступі.

У *першому розділі* проведено огляд літературних джерел, які стосуються теми дисертаційної роботи та пов'язані із дослідженням складних мереж. Суттєва увага приділяється основним моделям складних мереж, аналізу реальних мереж. Здійснено класифікацію та огляд методів інтелектуального аналізу даних.

У *другому розділі* дисертаційного дослідження проведено огляд існуючих програмних засобів для збирання інформації у веб-просторі. Наведено опис розробленого власного спеціалізованого програмного забезпечення – кроулера. Його особливістю є вбудований аналітичний модуль, функціональним призначенням якого є проведення статистичного та кластерного аналізу веб-графу.

Третій розділ присвячений дослідженню статистичних характеристик українського (edu.ua), ізраїльського (ac.il) та польського (edu.pl) сегментів веб-простору. Для вказаних сегментів отримано кластерну структуру, здійснено їх порівняльний аналіз. Описано алгоритми знаходження оптимальної кількості кластерів методом «ліктя» та k-core decomposition, знайдено оптимальну кількість кластерів та проведено їх порівняльний аналіз. Для проведення наведених вище досліджень використовувалась розроблена дисертанткою інформаційна технологія.

Питання кластеризації в графі розглянуто у *четвертому розділі*. Проведено аналіз спектральних властивостей стохастичної матриці. Отримано теоретичні результати про збіжність власних значень матриці за умов, накладених на елементи матриці суміжності; встановлено асимптотичні властивості спектру стохастичної матриці. Для оцінки оптимальної кількості кластерів побудовано критерій, обчислення якого ґрунтується на власних значеннях стохастичної матриці. Проведено моделювання та порівняння з класичними результатами.

Основні результати дисертаційного дослідження зазначені у *висновках*.

Укладені дисертанткою *додатки* (діаграма основних класів кроулера та їх опис, лістинг частини коду програми, акти та довідки про впровадження) логічно доповнюють та унаочнюють зміст і результати дослідження.

Виходячи з аналізу основної частини дисертації, можна зробити висновок, що дисертація є завершеним науковим кваліфікаційним дослідженням, у якій Кириченко О.Л. вирішила поставлені завдання та досягла мети дисертаційної роботи.

Наукова новизна

Отримані наукові результати дисертантки розв'язують актуальну науково-прикладну задачу. Заслуговують на особливу увагу такі висновки дисертантки:

- вперше розроблено спеціалізоване програмне забезпечення – кроулер з вбудованим аналітичним модулем для проведення статистичного та кластерного аналізу складних мереж;
- для збирання, обробки, збереження та проведення статистичного та кластерного аналізу інформації у складних мережах розроблено інформаційну технологію;
- з використанням інформаційної технології вперше проведено дослідження статистичних особливостей, кластерної структури та здійснено порівняльний аналіз освітянських зон веб-простору (edu.ua, edu.pl, ac.il);
- проведено аналіз основних властивостей спектру стохастичної матриці графу, що дозволило провести спектральний аналіз власних значень матриці переходу для випадку однотипного зв'язку для всіх елементів;
- зроблено висновок про оптимальну кількість кластерів для різних прикладних задач на основі припущень про однотипність зв'язків у кластері;
- побудовано критерій оцінки оптимальної кількості кластерів, обчислення якого ґрунтується на власних значеннях стохастичної матриці;

- отримано теоретичні результати про асимптотичні властивості спектру стохастичної матриці;
- розроблено алгоритм перевірки належності елементів до одного кластеру;
- виконано моделювання мережі зв'язків, що розподілені за законом Пуассона та визначено оптимальну кількість кластерів, здійснено порівняльний аналіз із методом «ліктя» та k-core decomposition.

Практичне значення одержаних результатів

Не викликає сумнівів практичне значення дослідження, яке полягає:

- у розробці кроулера та створенні інформаційної технології для знаходження статистичних характеристик і проведення кластерного аналізу зон веб-простору;
- розробці нового алгоритму знаходження оптимальної кількості кластерів.

Матеріали дисертаційного дослідження використовуються під час викладання навчальних дисциплін «Технології машинного навчання», «Методи та засоби кластерного аналізу», «Проектування інформаційних систем» у Чернівецькому національному університеті імені Юрія Федьковича.

Про практичне значення результатів дослідження свідчать і схвальні відгуки в довідках впровадження, надані компанією «Квант Азимут» (використання архітектури аналітичного модуля при розробці програмного забезпечення) та компанією «Qlicks B.V.» (для проведення кластеризації клієнтів при здійсненні маркетингових досліджень).

Дискусійні положення та зауваження до змісту дисертаційного дослідження

Загалом позитивно оцінюючи виконане дисертаційне дослідження Кириченко Оксани Леонідівни за його актуальність, наукову новизну, хотілося б висловити окремі побажання та звернути увагу на дискусійні положення, що потребують пояснення:

- 1) Не заперечуючи важливість викладеного матеріалу в розділі 3, окремі наукові аспекти потребують більш ґрунтовного пояснення (вибір спектрального алгоритму РС).
- 2) При описі інформаційної технології у п. 3.3 (розділу 3) не зазначено знаходження оптимальної кількості кластерів методом “ліктя” .
- 3) У третьому розділі дисертантка зазначає, що глибина зондування мережі призводить лише до кількісних змін, а якісні характеристики мережі при цьому залишаються майже незмінними. На мою думку, варто було більш детально розкрити цей аспект досліджень.
- 4) Є окремі неточності в оформленні роботи (так таблиця 2.6.1. і таблиця 3.5.1 ідентичні).

Вищенаведені зауваження не мають принципового характеру і не впливають на загальну позитивну оцінку роботи Кириченко О.Л. У цілому науковий рівень дисертації високий. Новизна, достовірність, наукове та практичне значення отриманих результатів не викликають сумнівів.

Загальний висновок

У дисертації вирішено актуальну наукову задачу, усі основні положення та найбільш важливі результати дисертаційного дослідження Кириченко Оксани Леонідівни, що винесені на захист, опубліковані у потрібному обсязі, пройшли відповідну апробацію на міжнародних та всеукраїнських науково-практичних конференціях, наукових семінарах. Вимоги щодо кількості та якості публікацій виконано. Порухень академічної доброчесності не виявлено. Структура роботи дозволяє зробити висновок про достатню репрезентативність та повноту розкриття теми.

Дисертація Кириченко Оксани Леонідівни «Дослідження статистичних характеристик складних мереж методами інтелектуального аналізу даних», подана на здобуття наукового ступеня доктора філософії у галузі знань 12 – «Інформаційні технології» за спеціальністю 121 – «Інженерія програмного забезпечення», є завершеним, самостійним дослідженням, що має наукову новизну, теоретичне та практичне значення, цілком відповідає пунктам 6, 7, 8, 9

«Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України №44 від 12 січня 2022 року (зі змінами, внесеними згідно з Постановою Кабінету Міністрів України №341 від 21 березня 2022 року та №502 від 19 травня 2023 року), а також «Вимогам до оформлення дисертації», затверджених Наказом Міністерства освіти і науки України №40 від 12 січня 2017 року.

Вважаю, що Кириченко Оксана Леонідівна заслуговує на присудження їй наукового ступеня доктора філософії у галузі знань 12 – «Інформаційні технології» за спеціальністю 121 – «Інженерія програмного забезпечення».

Рецензент –

доктор технічних наук, доцент,
доцент кафедри комп'ютерних наук
навчально-наукового інституту
фізико-технічних та комп'ютерних наук
Чернівецького національного університету
імені Юрія Федьковича

Дмитро УГРИН

