

РЕЦЕНЗІЯ

**доктора фізико-математичних наук, професора,
завідувача кафедри комп'ютерних наук
навчально-наукового інституту фізико-технічних та комп'ютерних наук
Чернівецького національного університету імені Юрія Федьковича
УШЕНКА Юрія Олександровича
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

Дисертаційна робота Кириченко Оксани Леонідівни присвячена дослідженню складних мереж, які сьогодні використовуються в багатьох сферах життя. Як правило, стрімкий розвиток мереж призводить до появи складних самоорганізованих систем, аналіз яких може надати дослідникам багато інформації про розвиток, сучасний стан та перспективи таких мереж. Для проведення досліджень, а саме – проведення статистичного та кластерного аналізу різних зон веб-простору, дисертанткою розроблена інформаційна технологія, яка містить в собі спеціалізоване програмне забезпечення (кроулер з вбудованим аналітичним модулем). Основні результати роботи отримано за допомогою застосування методів інтелектуального аналізу даних, методів кластерного та статистичного аналізу. Авторкою запропоновано новий метод знаходження оптимальної кількості кластерів у неструктурованих наборах даних, досліджено адекватність запропонованого підходу визначення оптимальної кількості кластерів за допомогою методів моделювання. Таким чином, кваліфікаційне дослідження Кириченко О.Л., безперечно, є актуальним.

Зв'язок роботи з державними програмами, планами, темами

Дисертаційне дослідження Кириченко О.Л. виконано на кафедрі програмного забезпечення комп'ютерних систем навчально-наукового інституту фізико-технічних та комп'ютерних наук Чернівецького національного університету імені Юрія Федьковича. Дослідження відповідають 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).

Тему дисертації затверджено рішенням Науково-технічної ради Чернівецького національного університету імені Юрія Федьковича (протокол №13 від 21 листопада 2017 року).

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

Наукові положення, висновки та результати дисертації обґрунтовані за допомогою загальноприйнятих методів інтелектуального аналізу та теорії стохастичних випадкових матриць графу для розв'язання визначених завдань. Результати роботи подані чітко, послідовно та логічно. В роботі проведено ґрунтовний огляд літературних джерел з тематики кваліфікаційного дослідження. Дисертаційне дослідження є самостійною науковою працею. Усі висновки, рекомендації та пропозиції, зокрема й ті, що характеризують наукову новизну, одержані авторкою особисто. За темою дисертаційного дослідження опубліковано 22 роботи. Основні результати опубліковані у чотирьох наукових статтях (1 з яких – в журналі, що індексується у наукометричній базі SCOPUS, 3 – в українських фахових виданнях). Додатково результати дисертації відображено в трьох наукових статтях (1 – в розділі монографії міжнародного видання, 2 – в українських фахових виданнях). Кваліфікаційне дослідження було представлено на 15 міжнародних та всеукраїнських наукових конференціях із публікацією тез чи матеріалів доповідей.

Структура дисертації

Дисертація складається зі вступу, чотирьох розділів, висновків до кожного розділу, загальних висновків, переліку використаних джерел (252 найменування), чотирьох додатків та списку публікацій авторки дисертаційного дослідження за темою дисертації (7 статей та 15 тез конференцій). Робота викладена на 229 сторінках, містить 12 таблиць та 43 рисунки. Основні результати дисертації у повній мірі відображені у публікаціях.

У вступі обґрунтовано актуальність задачі дослідження, виділено мету, завдання, предмет, об'єкт та методи дослідження, вказано наукову новизну, теоретичне, практичне значення отриманих результатів та структуру дисертаційної роботи.

У першому розділі наведені основні відомості з теорії складних мереж, проведено огляд та опис основних моделей складних мереж, розглянуто та проаналізовано приклади реальних складних 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. Описано розроблене власне програмне забезпечення (кроулер), яке містить вбудований аналітичний модуль для проведення інтелектуальної обробки зібраної інформації.

У третьому розділі наведено опис розробленої інформаційної технології, за допомогою якої проведено дослідження статистичних характеристик освітніх сегментів веб-простору (українського (edu.ua), ізраїльського (ac.il) та польського (edu.pl)). Отримано кластерну структуру вказаних вище сегментів веб-простору та здійснено їх порівняльний аналіз. Знайдено оптимальну кількість кластерів двома класичними методами (метод «ліктя» та k-core decomposition), проведено їх порівняльний аналіз.

У четвертому розділі розглянуто питання кластеризації в графі на основі матриці суміжності. Із врахуванням кластерної структури графу проведено аналіз спектральних властивостей стохастичної матриці. Доведено збіжність власних значень матриці за умов, накладених на елементи матриці суміжності. Встановлено факти про асимптотичні властивості спектру стохастичної матриці. У результаті проведених досліджень побудовано критерій оцінки оптимальної кількості кластерів, обчислення якого ґрунтується на власних значеннях стохастичної матриці. Проведено моделювання та порівняння з класичними результатами.

У висновках підсумовано основні результати дисертаційного дослідження.

Додатки містять акти та довідки про впровадження результатів роботи, діаграму основних класів кроулера та їх опис, лістинг частини коду програми.

Наукова новизна

Результати і висновки дисертації, які становлять наукову новизну, включають наступні аспекти:

- Проведено аналітичний огляд існуючих програмних засобів для збирання інформації у веб-просторі, наведено їх короткий опис, визначено їхні переваги та недоліки.
- Із застосуванням сучасних методів розробки програмного забезпечення вперше розроблено спеціалізоване програмне забезпечення – кроулер. Особливістю даного кроулера є наявність вбудованого аналітичного модуля для проведення статистичного та кластерного аналізу складних мереж.
- Розроблено інформаційну технологію, за допомогою якої вперше досліджено статистичні характеристики та кластерну структуру освітянських сегментів веб-простору: українського edu.ua, польської підмережі edu.pl та ізраїльської академічної зони ac.il. Проведено порівняльний аналіз статистичних характеристик та кластерної структури.
- Здійснено аналіз граничного розподілу власних значень матриці переходу на основі зв'язку викидів серед власних значень та оптимальної кількості кластерів. Зроблено висновок про оціночне значення оптимальної кількості кластерів для різних випадків (на основі припущень про однотипність зв'язків у кластері). Побудовано критерій оцінки оптимальної кількості кластерів, обчислення якого ґрунтується на власних значеннях стохастичної матриці.
- Сформульовано і доведено теореми та леми, які дозволяють оцінювати розподіли власних значень випадкових матриць та переносити класичні результати на матриці із слабо корельованими елементами. На основі отриманих теоретичних результатів розроблено алгоритм перевірки належності елементів (вершин графу) до одного кластеру.

- Проведено моделювання мережі зв'язків (графів), що будуються за законом Пуассона, оцінено оптимальну кількість кластерів та проведено порівняльний аналіз із класичними методами визначення оптимальної кількості кластерів (метод «ліктя», метод силуету, k-core decomposition).

Практичне значення одержаних результатів

Практичне значення одержаних результатів полягає у розробці кроулера, інформаційної технології та методу визначення оптимальної кількості кластерів. Результати дисертаційного дослідження використані при розробці програмного забезпечення компанією «Квант Азимут» (підходи до архітектури аналітичного модуля) та компанією «Qlicks B.V.» – для проведення сегментації клієнтів на різні категорії, які потім використовуються для персоналізованих маркетингових кампаній і стратегій та передбачення поведінки клієнтів на основі аналізу покупок, історії пошуку або профілей в соціальних мережах. Результати роботи впроваджені та використовуються в освітньому процесі відділу комп'ютерних технологій Навчально-наукового інституту фізико-технічних та комп'ютерних наук Чернівецького національного університету імені Юрія Федьковича при викладанні освітніх компонентів «Технології машинного навчання», «Методи та засоби кластерного аналізу», «Проектування інформаційних систем».

Дискусійні положення та зауваження до змісту дисертаційного дослідження

- 1) У другому розділі дисертаційного дослідження було б доцільно при описі роботи розробленого програмного забезпечення приділити більше уваги процесу його тестування.
- 2) На графіку рис. 3.4.2.1. третього розділу відображено розподіл ймовірності вузлів за ступенями по вихідних зв'язках (out degree) для зони edu.pl, але не проведено апроксимацію різних ділянок різними степеневими законами, як це зроблено на рисунках 3.4.1.1 та 3.4.3.1.
- 3) У дисертації присутні незначні стилістичні та граматичні описки.

Вказані зауваження не применшують наукового значення дисертаційного дослідження Оксани Леонідівни Кириченко та не впливають на висновок про належний рівень рецензованого дослідження, його загальну позитивну оцінку як самостійного і завершеного.

Загальний висновок

Актуальність дисертаційної роботи Оксани Леонідівни Кириченко незаперечна. Висновки й основні положення дисертації мають наукову новизну, високу теоретичну та практичну цінність. Під час дослідження тексту дисертаційної роботи порушень академічної доброчесності не виявлено.

Дисертаційне дослідження та публікації автора виконані на достатньо високому науковому рівні, наукова новизна, висновки, їх обґрунтованість, рівень виконання поставленого наукового завдання та володіння здобувачем методологією наукових досліджень відповідають вимогам, що висуваються до праць такого рівня.

Дисертаційна робота 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 завершеною науковою працею, яка за її актуальністю, науково-теоретичним рівнем, новизною розв'язання поставлених завдань, практичним значенням отриманих результатів, а також їх апробацією цілком відповідає пунктам 6-9 «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), 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 фізико-математичних наук, професор,
завідувач кафедри комп'ютерних наук
навчально-наукового інституту фізико-технічних
та комп'ютерних наук
Чернівецького національного університету
імені Юрія Федьковича

Юрій УШЕНКО

