ВИСНОВОК

про наукову новизну, теоретичне та практичне значення результатів дисертації Кириченко Оксани Леонідівни на тему: “Дослідження статистичних характеристик складних мереж методами інтелектуального аналізу даних” на здобуття ступеня доктора філософії за спеціальністю 121 – Інженерія програмного забезпечення з галузі знань 12 – Інформаційні технології

1. Обґрунтування вибору теми дослідження та її зв’язок із планами наукових робіт Університету.

Характерною рисою сучасного етапу розвитку суспільства є накопичення та обробка величезних обсягів інформації, для чого найчастіше використовують методи інтелектуального аналізу даних. Сьогодні ми живемо і працюємо в інформаційному просторі, який, без сумніву, являє собою складну мережу з притаманними таким структурам особливостями та зв’язками. Дослідження цих статистичних особливостей сприяє кращому розумінню організації таких мереж, їх використанню в різних кампаніях та інших галузях аналізу суспільства. Статистичною основою досліджень служить адекватна сегментація складних мереж на організовані за різними критеріями структури меншої розмірності, які дозволяють спростити аналіз великих обсягів даних, дослідити структуру мережі в цілому та її кластерів зокрема.

Дослідженням складних мереж займалися та займаються багато видатних українських та закордонних науковців, серед яких слід відзначити роботи M.E.J. Newman, P. Erdos, A. Renyi, D. J. Watts, S. H. Strogatz, A.-L. Barabasi, R. Albert, S. Dorogovtsev, Д.В. Ланде, І.Ю. Субач , О. Головач, К. Олємской та інші. У роботах зазначених авторів розроблено класичні моделі складних мереж та отримано суміжні теоретичні результати, що дозволяють точніше описувати реальні дані.

Основну увагу дисертаційної роботи присвячено нез’ясованим на сьогодні питанням аналізу складних мереж: оптимальний вибір кластерів складної мережі, вивчення статистичних особливостей підмереж (наприклад,

окремих зон веб-мережі), технології збирання та аналізу інформації про такі мережі за допомогою спеціалізованого програмного забезпечення тощо.

Метою дисертаційного дослідження є проведення статистичного та кластерного аналізу окремих зон веб-простору на основі програмного забезпечення власної розробки та нових підходів для визначення оптимальної кількості кластерів складних мереж.

Методи дослідження. В основу дисертаційної роботи покладені методи інтелектуального аналізу даних, а саме: кластерний аналіз; класичні математичні методи (побудова графової структури мережі), статистичні методи (обчислення статистичних характеристик мережі); теорія матриць (побудова стохастичних матриць та дослідження їх спектральних характеристик).

Об'єктом дослідження є складні мережі та процес збирання й обробки інформації.

Предметом дослідження є моделі, методи та інформаційна технологія збирання і обробки статистичної інформації, отриманої з веб-простору.

2. Формулювання наукового завдання, нове розв'язання якого отримано в дисертації.

Наукові завдання, розв'язання яких отримано у дисертації, полягають у наступному: розробці програмного забезпечення з урахуванням сучасних засобів та методів розробки, яке включає аналітичний модуль інтелектуальної обробки інформації; розробці інформаційної технології збирання та обробки інформації у складних мережах; порівняльному дослідженні, за допомогою розробленої технології, статистичних характеристик та кластерної структури окремих зон веб-простору; розробці нового методу визначення оптимальної кількості кластерів у складних мережах з використанням нового підходу до теорії дослідження графових структур; дослідженні адекватності запропонованого методу визначення оптимальної кількості кластерів за допомогою методів моделювання.

3. Наукові положення, розроблені особисто дисертантом, та їх новизна:

- на основі сучасних методів розробки програмного забезпечення розробила кроулер, який, на відміну від існуючих, містить аналітичний модуль інтелектуальної обробки інформації, підтримує контейнеризацію, роботу в багатопотоковому режимі, легко масштабується, і може бути використаний для аналізу споріднених мереж;
- розробила інформаційну технологію збирання та інтелектуального аналізу даних у складних мережах;
- виконала порівняльні дослідження статистичних характеристик та кластерної структури українського освітянського сегменту edu.ua, польської підмережі edu.pl та ізраїльської академічної зони ac.il;
- сформулювала та довела теореми та леми, які дозволяють оцінювати розподіл власних значень випадкових матриць та переносити класичні результати на матриці із слабко корельованими елементами; на основі розширення границь теорії дослідження графових структур розроблено алгоритм перевірки належності елементів до одного кластеру, що надає можливість визначити оптимальну кластерну структуру;
- на основі отриманих теоретичних результатів виконала порівняльний аналіз з класичними методами кластеризації, а саме: методом «ліктя», k-core decomposition та методом силуету. У результаті проведених досліджень, побудовано критерій оцінки оптимальної кількості кластерів k_{opt}, обчислення якого ґрунтується на власних значеннях стохастичної матриці P, а саме

$$\hat{k}_{opt} = \#\{\lambda_i(P): Re(\lambda_i(P)) > \max|Im(\lambda_i(P))|\}.$$

- з використанням методу Монте – Карло встановила, що у ряді випадків, запропонований спектральний метод визначення кількості кластерів дає більш точні оціночні значення кількості кластерів у графі в порівнянні з відомими методами («ліктя», k-core decomposition та методом силуету), що задається стохастичною матрицею P чи матрицею суміжності A. Крім того, запропонований новий метод є менш чутливим до наявності кластерів різної розмірності.

На всіх етапах досліджень нових задач і розробки нових методів дисертантка брала активну участь. Нею здійснено ґрунтовний аналіз отриманих результатів та сформовано висновки до кожного розділу дисертації.

4. Обґрунтованість і достовірність наукових положень, висновків і рекомендацій, які захищаються.

Достовірність наукових положень та висновків обґрунтовані тим, що для розробки спеціалізованого програмного забезпечення використані сучасні методи та середовища розробки; програмне забезпечення протестоване стандартними методами тестування; отримані за допомогою розробленої інформаційної технології дані не суперечать загальноприйнятим міркуванням і принципам, а теоретичні положення, розроблені дисертанткою із застосуванням добре апробованих теоретичних методів, перевірені та добре корелюють із класичними методами визначення оптимальної кількості кластерів у складних мережах.

Дисертація складається із вступу, чотирьох розділів, висновків до розділів, загальних висновків, переліку використаних джерел та чотирьох додатків. Дисертаційна робота є самостійною науковою працею, має теоретичне та прикладне значення. Усі висновки, рекомендації та положення, що характеризують наукову новизну дослідження, сформульовані особисто авторкою дисертації.

Публікації (одна стаття у виданні, що індексується у наукометричній базі даних SCOPUS, три статті у наукових фахових виданнях України) у наукових журналах за темою дисертації висвітлюють проблематику та основні положення наукової роботи. Наукові праці (3), які додатково відображають результати дисертації, підтверджують високий фаховий рівень дисертантки. Апробація основних наукових результатів відбулася у формі доповідей на 15 Всеукраїнських та Міжнародних науково-практичних конференціях.

5. Рівень теоретичної підготовки здобувачки та рівень її обізнаності з результатами наукових досліджень інших науковців високий. Це випливає з

великої кількості літературних джерел, наведених у дисертації, так і з доповіді матеріалів дисертації.

Особистий внесок здобувачки у розв'язанні конкретних наукових завдань також досить високий. Зокрема:

- дисертантка брала участь в обговоренні постановки задачі, визначенні мети роботи, виборі методів досліджень та у підготовці матеріалів до публікації у наукових журналах, представленні на наукових конференціях із доповіддю отриманих результатів;
- займалась розробкою, тестуванням та вдосконаленням програмного забезпечення;
- виконувала основні розрахунки та аналізувала отримані результати;
- виконувала моделювання мережі зв'язків, що розподілені за законом Пуассона та визначенням оптимальної кількості кластерів;
- займалась питаннями застосування теоретичного результату до визначення оптимальної кількості кластерів складних мереж, при дослідженнях структури веб-простору та Grid system.

6. Наукове та практичне значення роботи.

Важливе наукове значення дисертаційної роботи полягає в тому, що результати теоретичних досліджень, а саме розвитку теорії графових досліджень, сформульовані та доведені леми і теореми, можуть використовуватися для подальших досліджень у цій галузі, адже на сьогодні немає універсального способу проведення процесу кластеризації у складних мережах. Цю проблему вирішує дисертантка, яка довела, що розроблений новий метод визначення оптимальної кількості кластерів за певних умов дає кращі результати порівняно із існуючими класичними методами.

Практичне значення дисертаційної роботи полягає в тому, що розроблені авторкою кроулер, інформаційна технологія та метод визначення оптимальної кількості кластерів можуть використовуватися для подальшого вивчення складних мереж.

Запропоновані дисертанткою підходи до архітектури аналітичного модуля вже використовуються у компанії “Квант Азимут” для розробки власного програмного забезпечення та в компанії «Qlicks B.V.» – для проведення сегментації клієнтів на різні категорії, які потім ефективно використовуються для персоналізованих маркетингових кампаній і стратегій та передбачення поведінки клієнтів на основі аналізу покупок, історії пошуку або профілів в соціальних мережах.

7. Повнота викладу матеріалів дисертації в публікаціях та особистий внесок здобувачки в публікації.

Особистий внесок здобувачки в публікації такий, який вказаний у пункті 5 цього висновку.

Результати перевірки тексту дисертації з використанням антиплагіатної системи UNICHECK показав на 8,24% схожості з джерелами з Інтернету. Робота відповідає принципам академічної доброчесності.

Основні положення і висновки дисертаційної роботи викладені у 7 наукових працях. Зокрема, одна з них індексована у наукометричній базі Scopus, 3 наукові праці – у виданнях, що включені до переліку наукових фахових видань України. Кількість наукових праць, які додатково відображають наукові результати дисертації – 3. Матеріали дисертації доповідалися і опубліковані у 15 тезах доповідей на Міжнародних та Всеукраїнських наукових конференціях.

Результати дисертації повною мірою викладені в зазначених публікаціях.

Наукові праці у періодичних наукових виданнях, проіндексованих у наукометричних базах даних Scopus:

1. Kyrychenko O., Ostapov S., Kanovsky I. Investigation of the certain internet domain statistical characteristics / Статистичні характеристики деяких зон інтернету та їх дослідження. *Eastern-European Journal of Enterprise Technologies*. 2013. Vol. 6, no. 12(66). P. 91–96. (Scopus).

Наукові праці у виданнях, включених до переліку наукових фахових видань України:

2. Кириченко О.Л., Малик І.В., Остапов С.Е. Стохастичні моделі в задачах штучного інтелекту. *Вісник Київського національного університету імені Тараса Шевченка. Серія фізико-математичні науки.* 2021. № 2. С. 53–57.
3. Kyrychenko O. Information technology for statistical cluster analysis of information in complex networks. *Computer Systems and Information Technologies*. 2022. No 4. P. 47–51.
4. Кириченко О. Особливості архітектури програмного забезпечення для збору та аналізу статистичної інформації в глобальній мережі. *Information Technology: Computer Science, Software Engineering and Cyber Security*. 2023. № 2. С. 107–112.

Наукові праці, які додатково відображають наукові результати дисертації:

5. Kyrychenko O.**,** Ostapov S., Kanovsky I. Comparison of Statistical Characteristics of Certain Internet Subdomains. Chapter 1. Monograph. *Internet in the information society. Insights on the information systems, structures and applications* / ed. by M. Rostanski, P. Pikiewicz. Scientific Publishing of the Academy of Business in Dabrowa Gornicza : Wydawnictwo Naukowe, 2014. 138 p. ISBN: 978-83-62897-91-9.
6. Кириченко О.Л., Малик І.В., Остапов С.Е. Аналіз кластерної структури Інтернет-мереж на основі випадкових матриць. *Проблеми керування та інформатики*. 2022. №1. С. 37–46.
7. Кириченко О. Л., Kanovsky I., Остапов С. Е. Програмне забезпечення для дослідження статистичних характеристик глобальної мережі WWW. *Системи обробки інформації*. 2013. Том 2. Вип. 3. С. 99–104.

Наукові праці, які засвідчують апробацію матеріалів дисертації:

8. Кириченко О. Л., Остапов С. Е. Статистичні характеристики деякої частини українського домену глобальної мережі WWW. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2012)* : тези доповідей Всеукр. наук.-практ. конф., м. Чернівці, 3–5 травня 2012 р. Чернівці : Золоті литаври, 2012. С. 93–94.

9. Кириченко О. Л., Остапов С. Е., Kanovsky I. Статистичні характеристики українських доменів edu.ua, net.ua глобальної www-мережі та їх дослідження. *Актуальні проблеми інформаційних технологій, економіки та права (ІТЕП–2013)* : тези доповідей Міжнар. наук.-практ. конф., м. Чернівці, 3–5 квітня 2013 р. Чернівці : Книги-ХХІ, 2013. С.84–85.
10. Кириченко О. Л., Kanovsky I., Остапов С. Е. Складні мережі та їх статистичні характеристики: аналіз деяких сегментів web-простору. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2013)* : тези доповідей Всеукр. наук.-практ. конф., м. Чернівці, 27–31 травня 2013 р. Чернівці : Видавничий дім «Родовід», 2013. С. 16–22.
11. Кириченко О. Л., Kanovsky I. Y., Остапов С. Е. Дослідження статистики складних мереж на прикладі українських та ізраїльського доменів www-простору. *Фізико-технологічні проблеми радіотехнічних пристроїв, засобів телекомунікацій, нано- та мікроелектроніки* : матеріали ІІІ-ої Міжнар. наук.-практ. конф., м. Чернівці, 24–26 жовтня 2013 р. Чернівці, 2013. С. 125–126.
12. Кириченко О. Л., Остапов С. Е. Статистичні характеристики польського домену Інтернет edu.pl. *Інтелектуальні технології в системному програмуванні (ІТСП-2014)* : збірн. наук. праць ІІІ Всеукр. наук.-практ. конф. молодих учених та студентів, м. Хмельницький, 23–25 квітня 2014 р. Хмельницький : Гонта А.С., 2014. С. 269–270.
13. Кириченко О. Л., Остапов С. Е., Кановський І. Я. Моделювання, дослідження та аналіз деяких сегментів web-простору. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2014)* : праці Міжнар. наук.-практ. конф., м. Чернівці, 27–30 травня 2014 р. Чернівці : Видавничий дім «Родовід», 2014. С. 174–176.
14. Кириченко О. Л., Остапов С. Е. Дослідження структури Веб-простору за допомогою кластерного аналізу. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2016)* : праці Міжнар. наук.-практ. конф., м. Чернівці, 21–24 травня 2016 р. Чернівці : Видавничий дім «Родовід», 2016. С. 112–114.

15. Кириченко О.Л., ОстаповС.Е., Кановський І.Я. Проведення оптимальної кластеризації структури веб-простору. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2017)* : праці Міжнар. наук.-практ. конф., м. Чернівці, 05–08 жовтня 2017 р. Чернівці : Видавничий дім «Родовід», 2017. С. 68–70.
16. Кириченко О. Л., Остапов С. Е., Кановський І. Я. Проведення оптимальної кластеризації структури деяких зон веб-простору за допомогою методу K-Core Decomposition. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2018)* : праці Міжнар. наук.-практ. конф., м. Чернівці, 11–14 жовтня 2018 р. Чернівці : Видавничий дім «Родовід», 2018. С. 48–50.
17. Кириченко О. Л., Остапов С. Е. Застосування методу k-Core Decomposition для проведення оптимальної кластеризації. *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я* : тези доп. XXVII Міжнар. наук.-практ. конф. MicroCAD-2019 (м. Харків, 15–17 травня 2019 р.) : у 4 ч. Ч. IV. / за ред. проф. Сокола Є. І. Харків : НТУ «ХПІ». С. 155.
18. Кириченко О. Л., Остапов С. Е., Кановський І. Я. Розробка інформаційної технології проведення статистично-кластерного аналізу інформації у складних мережах. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2019)* : праці Міжнар. наук.-практ. конф., м. Чернівці, 03–06 жовтня 2019 р. Чернівці : ЧНУ, 2019. С. 92–94.
19. Кириченко О. Л., Остапов С. Е. Інформаційна технологія для проведення досліджень у складних мережах. *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я* : тези доп. XXVIII Міжнар. наук.-практ. конф. MicroCAD-2020 (м. Харків, 28–30 жовтня 2020 р.) : у 5 ч. Ч. І. / за ред. проф. Сокола Є.І. Харків : НТУ «ХПІ». С. 324.
20. Кириченко О. Л., Малик І. В., Остапов С. Е. Кластеризація великих даних на основі спектрального аналізу матриці переходу. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2020)* : праці IX Міжнар. наук.-практ. конф., м. Чернівці, 28–31 жовтня 2020 р. Чернівці : ЧНУ, 2020. С. 83–84.

21. Кириченко О. Л., Малик І. В., Остапов С. Е. Асимптотичний розподіл власних значень матриці переходу. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2021)* : праці X Міжнар. наук.-практ. конф., м. Чернівці, 28–31 жовтня 2021 р. Чернівці : ЧНУ, 2021. С. 22–24.
22. Kyrychenko O. L., Knignitska T. V., Ostapov S. E. Stochastic models in artificial intelligence development. *Modern stochastics: theory and applications V* : Materials of the Intern. Conf., Kyiv, Ukraine, June 1–4, 2021. Kyiv, 2021. P. 35.

8. Апробація результатів.

Апробація матеріалів дисертації здійснювалася на таких конференціях:

1. Міжнародна науково-практична конференція «Актуальні проблеми інформаційних технологій, економіки та права», м. Чернівці, 3–5 квітня 2013 року;

2. III Міжнародна науково-практична конференція «Фізико-технологічні проблеми радіотехнічних пристроїв, засобів телекомунікацій, нано– та мікроелектроніки», м. Чернівці, 24–26 жовтня 2013 року;

3. III Всеукраїнська науково-практична конференція молодих учених та студентів «Інтелектуальні технології в системному програмуванні», м. Хмельницький, 23–25 квітня 2014 року;

4. XXVII та XXVIII Міжнародні науково-практичні конференції «Інформаційні технології: наука, техніка, технологія, освіта, здоров'я (MicroCAD)», MicroCAD-2019, м. Харків, 15–17 травня 2019 року; MicroCAD-2020, м. Харків, 28–30 жовтня 2020 року;

5. Міжнародні науково-практичні конференції «Проблеми інформатики та комп'ютерної техніки (ПІКТ)», м. Чернівці: 3–5 травня 2012 року; 27–31 травня 2013 року; 27–30 травня 2014 року; 21–24 травня 2016 року; 05–08 жовтня 2017 року; 11–14 жовтня 2018 року; 03–06 жовтня 2019 року; 28–31 жовтня 2020 року; 28–31 жовтня 2021 року;

6. Міжнародна конференція «Сучасна стохастика: теорія та застосування V», м. Київ, 1 – 4 червня 2021року.

Дисертантка брала участь у роботі Міжнародної літньої школи ECODAM (м. Ясси, Румунія, 2019 р.), де доповідались результати дисертаційного дослідження.

У вересні 2022 року було зроблено доповідь на Міжнародному науковому семінарі «Advances & Challenges in Computing (A2C)», де було представлено основні наукові результати, які увійшли у дисертаційне дослідження.

9. Оцінка мови і стилю дисертації.

Мова і стиль дисертації відповідають вимогам, що висуваються до праць такого рівня.

10. Відповідність змісту дисертації спеціальності з відповідної галузі знань, з якої вона подається до захисту.

Зміст дисертації відповідає чинним вимогам до оформлення дисертації, встановленим освітньо-науковою програмою «Інженерія програмного забезпечення» галузі знань 12 «Інформаційні технології», спеціальності 121 – Інженерія програмного забезпечення.

11. Дотримання нормативних вимог щодо оформлення дисертації.

Нормативні вимоги щодо оформлення дисертації дотримані повністю.

12. Рекомендації дисертації до захисту.

Дисертаційна робота Кириченко Оксани Леонідівни «Дослідження статистичних характеристик складних мереж методами інтелектуального аналізу даних», подана на здобуття ступеня доктора філософії у галузі знань 12 – Інформаційні технології за спеціальністю 121 – Інженерія програмного забезпечення, за її актуальністю, науково-технічним рівнем, новизною постановки та розв’язання проблем, практичним значенням отриманих результатів відповідає вимогам пунктів 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 від 12 січня 2022 р. №44 (із змінами, внесеними згідно з Постановою Кабінету Міністрів України №431 від 21.03.2022 р.); за результатами публічної презентації результатів дисертації та їх обговорення на засіданні кафедри програмного забезпечення комп'ютерних систем Навчально-наукового інституту фізико-технічних та комп'ютерних наук Чернівецького національного університету імені Юрія Федьковича 12 грудня 2023 року дисертацію Кириченко Оксани Леонідівни рекомендовано до захисту в разовій спеціалізованій вченій раді для здобуття ступеня доктора філософії.

Голова засідання, Юрій ДОБРОВОЛЬСЬКИЙ

Гарант освітньо-наукової програми
«Інженерія програмного забезпечення»,
доктор технічних наук, професор,
доцент кафедри програмного забезпечення
комп'ютерних систем
Чернівецького національного університету
імені Юрія Федьковича

Підпис Добровольського Ю.О.
засвідчую
Учений секретар Чернівецького національного університету імені Юрія Федьковича
Мирковська
"17" 12 2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ * ЧЕРНІВЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ЮРІЯ ФЕДЬКОВИЧА * Код 02071240