

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Факультет систем управління літальних апаратів

Кафедра комп'ютерних наук та інформаційних технологій

Пояснювальна записка
до кваліфікаційної роботи
(тип кваліфікаційної роботи)

_____ магістра

(освітній ступінь)

на тему «Дослідження та візуалізація графів соціальних мереж»

ХАІ.302.366.22О.122. 9652063 ПЗ

Виконав: студент 6 курсу групи № 366
Спеціальність 122 «Комп'ютерні науки»

(код та найменування)

Освітня програма Комп'ютеризація
обробки інформації та управління

_____ Калюжний В.С.

(прізвище й ініціали студента)

Керівник: _____ проф. каф. 302, д.т.н.

_____ Прохоров О.В.

(прізвище й ініціали)

Рецензент: _____ Гамаюн І.П.

(прізвище й ініціали)

Харків – 2022

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М.Є. Жуковського
«Харківський авіаційний інститут»

Факультет систем управління літальних апаратів
 Кафедра комп'ютерних наук та інформаційних технологій
 Рівень вищої освіти другий (магістерський)
 Галузь знань 12 «Інформаційні технології»
 (код та найменування)
 Спеціальність 122 «Комп'ютерні науки»
 Освітня програма Комп'ютеризація обробки інформації та управління

ЗАТВЕРДЖУЮ
Завідувач кафедри
Олег ФЕДОРОВИЧ

(підпис) (ім'я та прізвище)
 «01» листопада 2022 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Калюжному Владиславу Сергійовичу
 (прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи Дослідження та візуалізація графів соціальних мереж

керівник кваліфікаційної роботи Прохоров О.В., д.т.н., проф. каф. 302,

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом Університету № 1602-уч. від «17» листопада 20 22 року

2. Термін подання студентом кваліфікаційної роботи 15.12.2022

3. Вихідні дані до проекту (роботи) дані з соціальних мереж Facebook та Instagram, бібліотека NetworkX

4. Зміст пояснювальної записки (перелік завдань, які потрібно розв'язати) аналіз задач дослідження соціальних мереж; дослідження структури соціальних графів; аналіз алгоритмів для розбиття користувачів на співтовариства та розрахунку різних метрик; розробка та реалізація програмного забезпечення додатку для аналізу і візуалізації графів соціальних мереж; здійснення програмної реалізації та проведення експериментальних досліджень

5. Перелік графічного матеріалу аналіз особливостей дослідження соціальних мереж; ціль та задачі дослідження; методи та алгоритми візуалізації; інструменти для аналізу і візуалізації соціальних мереж; метрики дослідження

графів соціальних мереж; статистика по набору даних з Facebook; аналіз соціальної мережі Facebook

6. Консультанти розділів кваліфікаційної роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
з основної частини	Прохоров О.В. проф. каф. 302		
з спеціального розділу	Прохоров О.В. проф. каф. 302		
з економіки та управління	Малєєва Ю. А. доц. каф. 302		

Нормоконтроль _____ Олександр ПРОХОРОВ «__» _____ 20__ р.
(підпис) (імя та прізвище)

7. Дата видачі завдання «_06_» _вересня_ 2022 р.

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів кваліфікаційної роботи	Примітка
1	АНАЛІЗ ОСОБЛИВОСТЕЙ ДОСЛІДЖЕННЯ СОЦІАЛЬНИХ МЕРЕЖ ТА WEB-ГРАФИ	06.09 – 20.09	
2	АНАЛІЗ МЕТОДІВ ТА АЛГОРИТМІВ ВІЗУАЛІЗАЦІЇ ГРАФІВ СОЦІАЛЬНИХ МЕРЕЖ	21.09 – 28.09	
3	МЕТРИКИ ДОСЛІДЖЕННЯ ГРАФІВ СОЦІАЛЬНИХ МЕРЕЖ	29.09 – 20.10	
4	РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ АНАЛІЗУ ТА ВІЗУАЛІЗАЦІЇ ГРАФІВ СОЦІАЛЬНИХ МЕРЕЖ	21.10 – 28.10	
5	Проведення експериментальних досліджень	29.10 – 20.11	
6	Економічні розрахунки	20.11 – 25.11	

Студент _____ Владислав КАЛЮЖНИЙ
(підпис) (імя та прізвище)

Керівник кваліфікаційної роботи _____ Олександр ПРОХОРОВ
(підпис) (імя та прізвище)

УДК 004.8

РЕФЕРАТ

Дослідження та візуалізація графів соціальних мереж – Калюжний Владислав, кваліфікаційна робота магістра, Харків, Національний аерокосмічний університет ім. М. Є. Жуковського «ХАІ», кількість сторінок, враховуючи додатки 85, кількість таблиць 8, кількість рисунків 37, кількість джерел літератури 15.

КЛЮЧОВІ СЛОВА: соціальні мережі, графи, вузли, метрики, спільноти, візуалізація, кластеризація.

ЦІЛЬ: аналіз та дослідження графів зв'язків користувачів соціальної мережі.

ОБ'ЄКТ ДОСЛІДЖЕННЯ: процеси формування та аналізу спільнот користувачів соціальних мереж.

ПРЕДМЕТ ДОСЛІДЖЕННЯ: методи та засоби аналізу і візуалізації графів соціальних мереж.

ЗАДАЧІ: провести аналіз задач дослідження соціальних мереж; дослідження структури соціальних графів; аналіз алгоритмів для розбиття користувачів на співтовариства та розрахунку різних метрик; розробити та реалізувати програмне забезпечення додатку для аналізу і візуалізації графів соціальних мереж; здійснити програмну реалізацію та провести експериментальні дослідження.

МЕТОДИ ДОСЛІДЖЕННЯ: теорія графів, методи розрахунку метрик графів, методи кластеризації, методи візуалізації графів, Python, фреймворк NetworkX.

ОТРИМАНІ РЕЗУЛЬТАТИ: розроблено програмне забезпечення, що дозволяє користувачам проводити та систематизувати дані про їх акаунти у соціальних мережах з вирішенням завдань аналізу і візуалізації соціальних графів та формування спільнот.

UDC 004.8

ABSTRACT

Research and visualization of social network graphs – Vladislav Kalyuzhny, Master Thesis, Kharkiv, National Aerospace University “KhAI”, amount of pages 85, amount of tables 8, amount of figures 37, amount of literature’s sources 15.

KEYWORDS: social networks, graphs, nodes, metrics, communities, visualization, clustering.

PURPOSE OF RESEARCH: analysis and research of graphs of connections of social network users.

OBJECT OF RESEARCH: processes of formation and analysis of communities of users of social networks.

SUBJECT OF RESEARCH: methods and means of analysis and visualization of graphs of social networks.

TASKS: conduct an analysis of social network research problems; research on the structure of social graphs; analysis of algorithms for dividing users into communities and calculating various metrics; develop and implement application software for analysis and visualization of graphs of social networks; implement program implementation and conduct experimental research.

METHODS OF RESEARCH: graph theory, graph metric calculation methods, claterization methods, graph visualization methods, Python, NetworkX framework.

RESULTS: developed software that allows users to organize and systematize data about their accounts in social networks, solving the tasks of analyzing and visualizing social graphs and forming communities.

ЗМІСТ

ВСТУП.....	8
1 АНАЛІЗ ОСОБЛИВОСТЕЙ ДОСЛІДЖЕННЯ СОЦІАЛЬНИХ МЕРЕЖ ТА WEB-ГРАФИ	10
1.1 Задачі та особливості дослідження соціальних мереж	10
1.2 Графи і візуалізація	17
Висновки за розділом.....	21
2 АНАЛІЗ МЕТОДІВ ТА АЛГОРИТМІВ ВІЗУАЛІЗАЦІЇ ГРАФІВ СОЦІАЛЬНИХ МЕРЕЖ.....	23
2.1 Методи та алгоритми візуалізації	23
2.1.1 Метод зв'язування ребер.....	23
2.1.2 Створення джгутів ребер на основі власної геометрії ребер	27
2.1.3 Інші приклади	28
2.2 Графи у сучасних сервісах	29
Висновки за розділом.....	35
3 МЕТРИКИ ДОСЛІДЖЕННЯ ГРАФІВ СОЦІАЛЬНИХ МЕРЕЖ	36
3.1 Аналіз метрик графів соціальних мереж	36
3.2 Опис алгоритму візуалізації Force Atlas 2	39
Висновки за розділом.....	41
4 РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ АНАЛІЗУ ТА ВІЗУАЛІЗАЦІЇ ГРАФІВ СОЦІАЛЬНИХ МЕРЕЖ	42
4.1 Опис даних	42
4.2 Аналіз мережі Facebook.....	42
4.3 Основні топологічні атрибути	45
4.4 Міри центральності	49
4.5 Ефекти кластеризації	61
4.6 Мости.....	63
4.7 Асортативність	64
4.8 Спільноти мережі	65
Висновки за розділом.....	67
5 ЕКОНОМІЧНИЙ РОЗДІЛ.....	68
5.1 Мета розділу	68
5.2 Розрахунок собівартості та визначення тривалості створення програмного продукту (перший варіант)	68

5.3 Розрахунок собівартості та визначення тривалості створення програмного продукту (другий варіант).....	72
5.4 Висновки за розділом	75
ВИСНОВКИ.....	76
ПЕРЕЛІК ПОСИЛАНЬ	77
ДОДАТОК А. Графічні матеріали	79

ВСТУП

Актуальність теми дослідження.

Багатьом із нас важко уявити життя без соціальних мереж. Але реалії цих акаунтів залишаються невідомими. Просування в Instagram такі, що неважливо скільки у тебе передплатників, важливо який у тебе реальний охоплення. У акаунта, де 6 тисяч передплатників і активна аудиторія, може бути охоплення до 3 тисяч на кожен пост, і таке ж охоплення може бути в якомусь паблік-акаунті з 100-тисячною аудиторією. Низьке охоплення для блогерів означає, що доведеться знижувати вартість рекламних інтеграцій, для компаній це втрата аудиторії і клієнтів. Один із способів збільшити залученість і підвищити охоплення, це чистка недійсних акаунтів (ботів) в Instagram. На сьогодні зібрати фейки, боти, комерційні і «мертві» акаунти в окремий список дуже складно. Запустити блокування з цього списку і зняття блокування, щоб не закривати доступ назавжди теж не з легких задач. Тому для очищення акаунта Instagram від недійсних та неактивних акаунтів, потрібен якісний додаток, який би фільтрував та аналізував акаунт-сторінки популярних соціальних мереж. Методи виділення спільнот у соціальних графах активно досліджуються останнім часом, і візуалізація графа всього лише один із практичних аспектів цього завдання. Тому розробка додатку для аналізу бази користувачів соціальної мережі та візуалізація результатів у вигляді графів є актуальною.

Мета дослідження – аналіз та дослідження графів зв'язків користувачів соціальної мережі.

Для досягнення мети необхідно виконати наступні **завдання**:

1. Провести аналіз задач дослідження соціальних мереж.
2. Дослідження структури соціальних графів.
3. Аналіз алгоритмів для розбиття користувачів на співтовариства та розрахунку різних метрик.
4. Розробити та реалізувати програмне забезпечення додатку для аналізу і візуалізації графів соціальних мереж.
5. Здійснити програмну реалізацію та провести експериментальні дослідження.

Об'єкт дослідження – процеси формування та аналізу спільнот користувачів соціальних мереж.

Предмет дослідження – методи та засоби аналізу і візуалізації графів соціальних мереж.

Практичне значення отриманих результатів. Розроблено програмне забезпечення, що дозволяє користувачам проводити та систематизувати дані

про їх акаунти у соціальних мережах з вирішенням завдань аналізу і візуалізації соціальних графів та формування спільнот.

Структура та обсяг кваліфікаційної роботи магістра. Кваліфікаційна робота магістра складається зі вступу, 5 розділів, висновків, додатків та переліку посилань. Повний обсяг кваліфікаційної роботи магістра становить 85 сторінок, в тому числі: 37 рисунків; 8 таблиць; 1 додаток на 6 сторінках; перелік посилань з 15 найменувань на 2 сторінках.