Вінницький національний технічний університет
Факультет інформаційних технологій та комп'ютерної інженерії
Кафедра захисту інформації

**МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА**
на тему:
«СИСТЕМА ВИЯВЛЕННЯ КІБЕРЗАГРОЗ, ЩО СТВОРЮЮТЬСЯ СПІЛЬНОТАМИ В СОЦІАЛЬНИХ МЕРЕЖАХ. ЧАСТИНА 2. МОДУЛЬ МОНІТОРИНГУ ТА АНАЛІЗУ СПІЛЬНОТ ЗА ДОПОМОГОЮ ТЕОРІЇ ГРАФІВ.»

Виконав: студент 2 курсу групи 2БС-22м
спеціальності 125 Кібербезпека
_____ Вадим СТАДНІК

Керівник , к.т.н., професор каф. ЗІ
_____Наталія КОНДРАТЕНКО
«12» 12 2023 р.

Опонент: к. т. н., доцент каф. ПЗ
_____Олександр ТКАЧЕНКО
«13» 12 2023 р.

**Допущено до захисту**
Завідувач кафедри ЗІ
д. т. н., проф.
_____ Володимир ЛУЖЕЦЬКИЙ
«14» 12 2023 р.

Вінниця ВНТУ – 2023 року

Вінницький національний технічний університет
Факультет інформаційних технологій та комп'ютерної інженерії
Кафедра захисту інформації
Рівень вищої освіти II (магістерський)
Галузь знань – 12 Інформаційні технології
Спеціальність – 125 Кібербезпека
Освітньо-професійна програма – Безпека інформаційних і комунікаційних систем

ЗАТВЕРДЖУЮ
Зав. кафедри ЗІ, д. т. н., проф.
_____Володимир ЛУЖЕЦЬКИЙ
19.09 2023 року

**ЗАВДАННЯ**
**НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ**

Стадніку Вадиму Леонідовичу

1. Тема роботи: «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів.» керівник роботи: Кондратенко Наталія Романівна, к.т.н., професор, затверджені наказом ВНТУ №247 від 18.09.2023.
2. Строк подання студентом роботи – 13 грудня 2023 року
3. Вихідні дані до роботи:
   - засіб повинен запускатись в операційних системах Windows 10 та Linux;
   - засіб повинен надавати достовірність визначення спільнот у графі;
   - засіб повинен надавати перелік пов'язаних із визначеною спільнотою загроз.
4. Зміст текстової частини: Вступ. 1. Огляд основних методів аналізу соціальних мереж для виявлення кіберзагроз. 2. Методи моніторингу та аналізу спільнот у соціальних мережах за допомогою теорії графів. 3. Програмна реалізація 4. Економічне обґрунтування. Висновки. Перелік використаних джерел. Додатки.
5. Перелік ілюстративного матеріалу: Актуальність, мета та задачі МКР (плакат А4). Схема алгоритму роботи програмного засобу (плакат, А4). Схема алгоритму Краскала (плакат, А4). Інтерфейс програмного засобу (плакат, А4). Сучасні напрямки моніторингу мереж на основі теорії графів(плакат, А4). Системний підхід до виявлення кіберзагроз, що утворюються спільнотами соціальних мереж на основі теорії графів (плакат, А4). Структура спілкування мережі (плакат, А4). Приклад побудованої мережі (плакат, А4). Структура інтервальної нечіткої моделі(плакат, А4). Моделі соціальних графів (плакат, А4) Блок схема

алгоритму пошукі графа в глибину( плакат, А4).Блок схема алгоритму Дейкстр
плакат, А4).

6. Консультанти розділів роботи

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|---|---|---|---|
| | | завдання видав | завдання прийняв |
| 1 | Наталія КОНДРАТЕНКО., к.т.н., професор. кафедри ЗІ | 19.09 | 25.09 |
| 2 | Наталія КОНДРАТЕНКО., к.т.н., професор. кафедри ЗІ | 19.09 | 29.09 |
| 3 | Наталія КОНДРАТЕНКО., к.т.н., професор. кафедри ЗІ | 19.09 | 26.10 |
| 4 | Ольга РАТУШНЯК, к. т. н., доц. каф. ЕВПМ | 19.09 | 27.11 |

7. Дата видачі завдання – 1 вересня 2023 року

КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів магістерської кваліфікаційної роботи | Строк виконання етапів роботи | Приміт |
|---|---|---|---|
| 1 | Аналіз завдання. Вступ | 01.09.2023 – 10.09.2023 | |
| 2 | Аналіз літературних джерел за напрямком магістерської кваліфікаційної роботи | 10.09.2023 – 15.09.2023 | |
| 3 | Науково-технічне обгрунтування | 16.09.2023 – 22.09.2023 | |
| 4 | Розробка технічного завдання | 23.09.2023 – 29.09.2023 | |
| 5 | Розробка рішень | 30.09.2023 – 12.10.2023 | |
| 6 | Практична реалізація, моделювання, експериментування, результати | 14.10.2023 – 10.11.2023 | |
| 7 | Розробка розділу економічного обгрунтування доцільності розробки | 11.11.23 – 17.11.2023 | |
| 8 | Аналіз виконання ТЗ, висновки | 18.11.2023 – 24.11.2023 | |
| 9 | Оформлення пояснювальної записки | 25.11.2023 – 30.11.2023 | |
| 10 | Попередній захист та доопрацювання МКР | 28.11.2023 – 01.12.2023 | |
| 11 | Перевірка магістерської роботи на наявність плагіату | 02.12.2023 – 10.12.2023 | |
| 12 | Представлення МКР до захисту | 11.12.2023 – 14.12.2023 | |
| 13 | Захист МКР | 14.12.2023 – 21.12.2023 | |

Студент_____Вадим СТАДНІК

Керівник роботи_____Наталія КОНДРАТЕНКО

# АНОТАЦІЯ

Магістерська кваліфікаційна робота складається з 100 сторінок формату А4, на яких є 15 рисунків, 8 таблиці, 37 формул, список використаних джерел містить 26 найменуваннь.

Магістерська кваліфікаційна робота присвячена розробці системи виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Ця дослідницька робота складається з двох частин, і дана частина є другою, де акцент здійснюється на модулі моніторингу та аналізу спільнот, використовуючи теорію графів.

Дослідження цієї частини роботи спрямоване на розвиток конкретних методик та алгоритмів для виявлення кіберзагроз у соціальних мережах, які можуть бути корисні для організацій, що працюють у сфері кібербезпеки, а також для правоохоронних органів та владних структур.

Результати цього дослідження сприятимуть підвищенню захищеності інтернет-спільнот від кіберзагроз, що виникають у соціальних мережах.

## ABSTRACT

The master's qualification thesis consists of 100 pages of A4 format, on which there are 25 figures, 10 tables, 37 formulas, the list of used sources contains 26 names.

The master's thesis is devoted to the development of a system for detecting cyberthreats created by communities in social networks. This research paper consists of two parts, and this part is the second one, which focuses on the community monitoring and analysis module using graph theory.

The research of this part of the work is aimed at the development of specific methods and algorithms for detecting cyber threats in social networks, which can be useful for organizations working in the field of cyber security, as well as for law enforcement agencies and power structures.

The results of this study will contribute to increasing the security of Internet communities against cyber threats arising in social networks.

**ЗМІСТ**

# ВСТУП

В сучасному цифровому світі, соціальні мережі відіграють важливу роль у житті суспільства. Вони стали не тільки платформами для спілкування, але й джерелом великої кількості інформації, що створюється та розповсюджується спільнотами користувачів. Однак, разом з цими можливостями, соціальні мережі стали місцем, де виникають кіберзагрози, що створюються спільнотами з різними цілями, включаючи деструктивні.

Сучасні кіберзагрози можуть приймати форму дезінформації, кібербулінгу, фішингу, атак на ідентичність, та багатьох інших видів атак, які можуть завдати шкоди індивідуальним користувачам, підприємствам, та навіть суспільству в цілому.

Такі загрози вимагають надзвичайної уваги та заходів щодо їх виявлення та запобігання. В цьому контексті, розробка системи виявлення кіберзагроз, що створюються спільнотами в соціальних мережах, набуває важливого значення. Вже існують певні підходи до виявлення кіберзагроз в мережах, але вони часто обмежені визначенням індивідуальних загроз, не звертаючи увагу на загрози, що формуються в спільнотах.

В даній дослідницькій роботі, ми розглядаємо систему виявлення кіберзагроз, створюваних самими спільнотами у соціальних мережах, та розвиваємо модуль моніторингу та аналізу, який використовує теорію графів для ідентифікації аномалій та потенційних загроз. Дослідження в цій галузі є важливим кроком у напрямку забезпечення кібербезпеки в соціальних мережах та захисту користувачів від кіберзагроз, що можуть виникати у цьому віртуальному середовищі.

**Актуальність.** Сучасне суспільство переживає період інтенсивного росту використання соціальних мереж як засобу комунікації, інформаційного обміну та взаємодії. З цим ростом збільшується і кількість кіберзагроз, що виникають в цьому цифровому середовищі, та наносять шкоду як окремим користувачам, так і громадській безпеці в цілому.Зокрема, на соціальних мережах активно

поширюються фейкові новини, дезінформація, агітація, атаки на ідентичність, кібербулінг та інші форми загроз. Важливим аспектом є те, що часто ці загрози не створюються окремими індивідами, але групами користувачів або спільнотами, які мають різні мотиви та цілі. Це ставить перед нами завдання розробити ефективну систему виявлення та відстеження кіберзагроз, що створюються самими спільнотами в соціальних мережах. Враховуючи складну графову структуру зв'язків між користувачами та спільнотами, теорія графів виявляється потужним інструментом для аналізу таких взаємодій та ідентифікації аномалій.

**Актуальність** цієї теми полягає в тому, що зростаюча кількість людей використовує соціальні мережі для комунікації та інформаційного обміну, і важливо забезпечити їхню безпеку та захист від потенційних кіберзагроз.

**Об’єктом** дослідження є система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах, зокрема модуль моніторингу та аналізу цих спільнот з використанням теорії графів.

**Предметом** дослідження є виявлення та аналіз кіберзагроз, які створюються спільнотами в соціальних мережах з використанням модуля моніторингу та аналізу, розробленого на основі теорії графів

**Метою** магістерської кваліфікаційної роботи є розробка та вдосконалення системи виявлення кіберзагроз, що створюються спільнотами в соціальних мережах, зокрема шляхом розробки та впровадження модуля моніторингу та аналізу на основі теорії графів. Робота націлена на вдосконалення засобів і методів виявлення кіберзагроз та на розробку інструменту, який допоможе в ідентифікації підозрілих активностей у соціальних мережах та забезпечить підвищену рівень кібербезпеки в цьому онлайн середовищі. Для досягнення мети необхідно виконати наступні завдання:

- Розробка модуля моніторингу;
- Використання теорії графів та виявлення аномалій;
- Практична реалізація , експерименти та валідація;
- Аналіз результатів , формулювання висновків та рекомендацій;

**Методи дослідження.** Здійснення аналізу графів соціальних мереж засобами теорії графів, включаючи обчислення метрик центральності (Edge Betweenness) для визначення ключових ребер, які можуть бути критичними для безпеки спільноти.

# 1 ОГЛЯД ОСНОВНИХ МЕТОДІВ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ ДЛЯ ВИЯВЛЕННЯ КІБЕРЗАГРОЗ

## 1.1 Основні типи соціальних мереж та методи їх генерації.

Розвиток соціальних мереж формує нову культуру та впливає на суспільство. Ці платформи не лише дозволяють спілкуватися та здобувати корисну інформацію, а й сприяють самореалізації людини. Основна мета їх використання - сприяти соціалізації та прагненню до спілкування з іншими користувачами. Кожна функція соціальної мережі взаємодіє з комунікацією і, нарешті, перетворюється на процес взаємодії між користувачами. В інтернеті існує безліч різноманітних соціальних мереж, які можна класифікувати за різними критеріями.

Соціальні мережі можна розділити за доступністю на три типи: закриті, відкриті та змішані. У більшості випадків сучасні соціальні мережі відкриті для всіх користувачів, хоча деякі проекти з причини своєї бізнес-моделі розраховані на обмежену аудиторію і починали як закриті. Соціальні мережі, які поєднують обидва підходи, розвиваються менш успішно: їх головна мета - досягнення популярності, але користувачі виявляють маленький інтерес через наявні бар'єри та обмеження, що призводить до їх неактивного використання.

Соціальні мережі класифікуються за їхнім географічним розташуванням на кілька типів: глобальні, національні, регіональні та ті, що не мають прив'язки до певного регіону. Щодо напрямку, вони можуть бути особистими, професійними або тематичними. Особисті мережі призначені для підтримки і розширення існуючих зв'язків і пошуку нових знайомств. Професійні спрямовані на кар'єрний зріст і розвиток у відповідній сфері. Тематичні збирають аудиторію за спільними інтересами, такими як музика, хобі та інше.

Більш детальна класифікація соціальних мереж включає наступне:Соціальні платформи для спілкування:

Це мережі, які сприяють обміну повідомленнями та інформацією між користувачами, такі як "Twitter", "Facebook" і подібні. Спочатку цей тип мереж

запропонував можливість створення особистих міні-сайтів, які пізніше перетворилися на відомі профілі.

Мережі для обміну медіа-контентом: Це платформи, де користувачі можуть ділитися відео та фотографіями, такі як "Instagram", "YouTube".

Мережі для групових дискусій: Цей тип мереж базується на обміні знаннями та дискусіях у групах користувачів, такі як "Quora", "Reddit"

.Платформи для авторських записів: Ці сервіси створені для блогів, де користувачі можуть публікувати текстовий та медійний контент, такі як "Blogger", "Twitter".

Соціальні закладки: Це платформи, де користувачі зберігають контент у своїй особистій бібліотеці, до якої можуть підписуватися інші користувачі, такі як "Pinterest", "Flipboard"

.Мережі за інтересами: Це платформи, що дозволяють знаходити осіб із подібними інтересами та спільно цікавитися чимось, наприклад, "Goodreads", "Friendster".

В цілому можна зробити висновок, що класифікація видів соціальних мереж є умовною, і до цього дня триває ділення на види, типи і категорії. Соціальні мережі стали важливою складовою сучасного інтернет-середовища, і різноманітність їх видів дає користувачам можливість вибирати той формат спілкування та взаємодії, який найкраще відповідає їхнім потребам та інтересам[1].

Рисунок 1.1 – соціальні мережі.

Методи генерації соціальних мереж включають різні підходи для створення моделей мереж, які відображають реальні соціальні взаємодії між людьми.

Нижче наведено кілька основних методів:

- Моделі на основі Ваттса-Строгатца (Small-World Models): Цей метод передбачає створення мережі, де кожен вузол має лише небагато найближчих сусідів та декілька довгих зв'язків. Це відображає характеристики "малих світів" реальних соціальних мереж.
- Моделі на основі безмасштабності (Scale-Free Models): Ці моделі передбачають створення мережі, де деякі вузли мають значно більше зв'язків, ніж інші. Це відображає "закон Барабаши-Альберта" і характерний розподіл ступенів вузлів у реальних соціальних мережах.
- Моделі на основі агентів (Agent-Based Models): Ці моделі використовують агентів, які імітують поведінку реальних осіб у мережі. Агенти можуть створювати та руйнувати зв'язки на основі певних правил, що дозволяє вивчати динаміку мережі.
- Моделі на основі реальних даних (Empirical Models): Інший спосіб генерації соціальних мереж полягає в використанні реальних даних з різних джерел, таких як соціальні мережі, дослідження або опитування. Це дає можливість створювати мережі, які точно відображають реальні взаємодії.
- Моделі на основі стохастичних процесів (Stochastic Models): Деякі методи використовують стохастичні процеси, щоб імітувати зв'язки між користувачами в мережі. Це дозволяє враховувати випадковість та еволюцію мережі.

Вибір конкретного методу для генерації соціальної мережі повинен бути обґрунтованим і залежати від мети вашого дослідження та того, які аспекти або характеристики мережі ви бажаєте відтворити. Кожен метод має свої переваги та обмеження, і важливо враховувати ці аспекти при розробці моделі.

Наприклад, якщо ваша мета - вивчення мережі для аналізу впливових користувачів, то ви можете обрати метод генерації, який надає більше уваги структурі мережі та центральності вузлів. Однак, якщо ваше дослідження спрямоване на аналіз динаміки взаємодій, то моделі на основі агентів або стохастичних процесів можуть бути більш відповідними[2].

Важливо враховувати, що кожен метод має свій контекст та сферу застосування, і вибір повинен бути здійснений на основі вашого конкретного дослідницького питання та мети дослідження. Додатковий аналіз переваг та обмежень кожного методу допоможе вам визначити найкращий підхід для вашого дослідження соціальних мереж.

### 1.2.Класифікація кіберзагрози у соціальних мережах та методи їх виявлення.

У сучасному світі, коли використання соціальних мереж стало невід'ємною частиною нашого повсякденного життя, виникає серйозна необхідність у забезпеченні кібербезпеки в цих середовищах. Соціальні мережі стають не лише місцем спілкування та обміну інформацією, але і сферою, в якій зростає кількість кіберзагроз.

Для ефективного захисту користувачів соціальних мереж важливо розуміти різновиди цих кіберзагроз і розробляти методи їх виявлення та запобігання. У цьому контексті, класифікація кіберзагроз та розробка відповідних заходів безпеки відіграють важливу роль у забезпеченні цифрової безпеки користувачів соціальних мереж.

Ці загрози можуть включати атаки, які базуються на соціальній інженерії, яка спрямована на обман користувачів та використання їхньої довіри. Зокрема, такі загрози можуть включати розміщення приманок, фішінг, веб-атаки, витік конфіденційної інформації та компрометацію працівників. Нижче ми розберемо конкретно кожну із загроз[3].

Фішинг - цей вид кіберзагрози, при якому атакуючий намагається обманути людину, щоб вона надала особисті чи конфіденційні дані, такі як

паролі, номери кредитних карток, інформацію про обліковий запис тощо. Зазвичай фішери видають себе за легітимних суб'єктів, такі як банки, соціальні мережі, електронні поштові служби або інші довірені організації.

Ось приклади видів фішингу:

1. Фішинг електронною поштою (email phishing): Атакуючий відправляє листи, які схожі на листи від легітимних організацій, і просить отримувача ввести свої особисті дані на піддельні веб-сайти. Наприклад, атаки від імені банків, соціальних мереж чи поштових служб.
2. Соціальний фішинг (social phishing): Атакуючий намагається зламати обліковий запис шляхом отримання особистої інформації про користувача через соціальні мережі. Наприклад, хакер може використовувати інформацію з профілю для вгамування або обманування користувача.
3. Фішинг на мобільних пристроях (mobile phishing): Атакуючі надсилають шкідливі додатки або посилання через SMS або месенджери, які можуть викликати виток інформації з мобільних пристроїв.
4. Фішинг через фізичну пошту (vishing): Атакуючий може здійснювати фішингові атаки за допомогою телефонних викликів, намагаючись здійснити шахрайські схеми або отримати інформацію через голосову комунікацію.
5. Фішинг через соціальні мережи (social media phishing): Атакуючий створює фальшиві профілі в соціальних мережах, які намагаються встановити контакт із потенційними жертвами та отримати конфіденційну інформацію

Фішинг є серйозною кіберзагрозою, і наслідки для користувачів можуть бути значною та негативною впливати на їхнє життя та фінанси.

Прикладом цього може бути втрата особистої інформації: Фішери можуть отримати доступ до особистих даних, таких як паролі, номери кредитних карток, адреси, соціальні номери тощо. Це може призвести до ідентичності злочинця, крадіжки грошей чи інших видів фінансового шахрайства також якщо фішери отримають доступ до фінансових облікових записів або інформації про кредитні

картки, користувач може стати жертвою фінансової крадіжки, ізоляції чи різних видів шахрайства[4].

Втрата контролю над обліковим записом: Коли фішери отримають доступ до облікового запису на соціальній мережі, поштовому сервісі або іншому важливому ресурсі, користувач може втратити контроль над своїм обліковим записом. Це може призвести до публікації непристойного контенту або зловмисних дій в ім'я користувача.

Поширення спаму: Фішери можуть використовувати компрометований обліковий запис для відправки спаму та небажаних повідомлень іншим користувачам. Це може призвести до негативного впливу на репутацію користувача та його стосунків з іншими користувачами.

Зловживання особистою інформацією: Якщо фішери отримають доступ до особистих даних, вони можуть використовувати цю інформацію для шантажу, загроз або інших злочинних дій, що порушують приватність користувача.

Втрата часу та зусиль: Відновлення втраченого облікового запису та виправлення наслідків фішингу може зайняти значну кількість часу та зусиль.

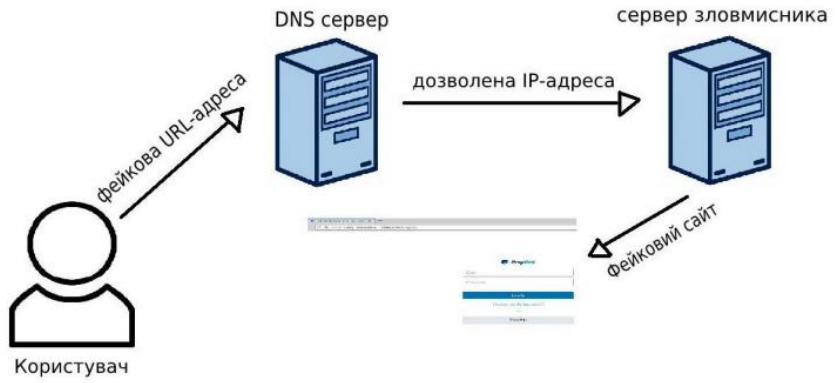



Рисунок 1.2 – Схема традиційної фішингової атаки

Спам в соціальних мережах включає в себе небажані повідомлення або вміст, який надсилається користувачам без їхньої згоди та часто містить рекламу, недоречний або небажаний вміст. Це може бути реклама товарів, послуг, сайтів, політична пропаганда, шахрайські схеми, ланцюгові листи, віруси та інші

неавторизовані або небажані повідомлення.

Ці ознаки допомагають визначити спам та розрізнити його від легітимних повідомлень на соціальних мережах:

Масовість: Спам часто відправляється великою кількістю користувачів, щоб максимізувати його поширення.

Небажаний вміст: Спам часто містить небажаний вміст, який користувачі не запитували і не бажають отримувати.

Агресивність: Спам може бути агресивним, намагаючись привернути увагу користувача, іноді навіть обманюючи його.

Порушення правил соціальної мережі: Спам часто порушує правила користування соціальною мережею, включаючи обмеження на надсилання небажаних повідомлень.

Вміст для дорослих: Спам часто містить матеріали для дорослих або образливий вміст.

Основні застереження та заходи безпеки, щоб уберегти себе від спаму у соціальних мережах:

Якщо ви отримали спам-повідомлення від незнайомого або сумнівного джерела, не відкривайте його та особливо не переходьте за посиланнями.

Більшість соціальних мереж мають можливість повідомити про спам. Використайте цю функцію, щоб сповістити адміністрацію мережі про небажаний вміст.

Якщо ви отримуєте спам-повідомлення від конкретного користувача, ви можете заблокувати цього користувача, щоб запобігти отриманню подібних повідомлень в майбутньому.

Не реагуйте на спам-повідомлення, оскільки це може підтвердити відправнику, що ваш обліковий запис активний, і він продовжить надсилати спам.

Переконайтеся, що ваші налаштування конфіденційності на соціальній мережі встановлені на максимальний рівень, щоб зменшити можливість отримання спаму[5].

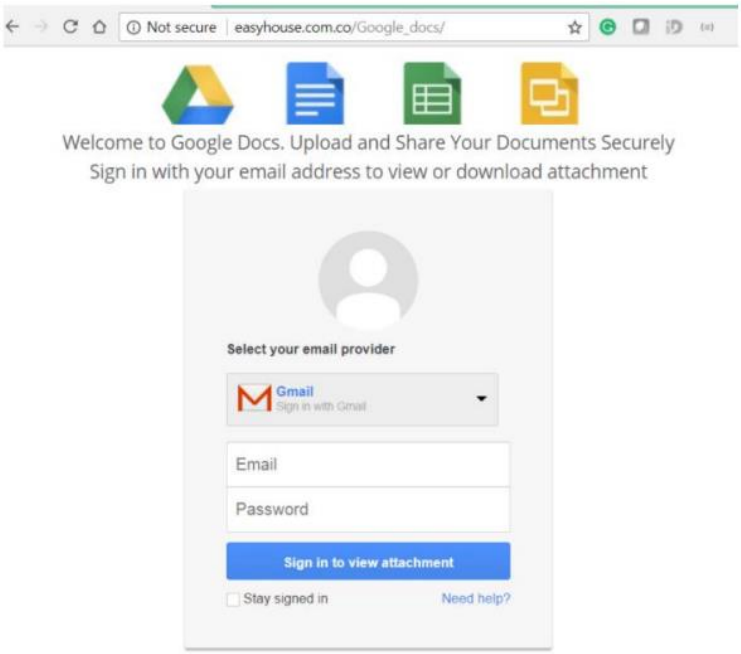



Рисунок 1.3 – Імітація сторінки Google Docs

Кібербулінг - це форма цифрового насильства або тиску, при якій особа (кібербулі) використовує Інтернет, соціальні мережі, повідомлення і інші електронні засоби для образ, погроз, шахрайства, приниження, обліку або інших негативних дій в напрямку іншої особи (жертви).

Аналіз випадків кібербулінгу та його вплив на жертв:

1. Психологічний тиск: Жертва кібербулінгу може відчувати психологічний тиск, страх, тривогу та невпевненість через образи та погрози, які надходять через Інтернет. Це може призвести до депресії, тривожних розладів і інших психічних проблем.
2. Сприйняття у школі або на роботі: Якщо кібербулінг відбувається в освітньому або робочому середовищі, жертва може стикатися з проблемами у школі або на роботі, включаючи зниження навчальної або професійної успішності.
3. Соціальна ізоляція: Через кібербулінг жертва може почувати себе відірваною від своїх друзів та родини. Це може призвести до соціальної ізоляції та втрати соціальної підтримки.

4. Фізичні наслідки: В деяких випадках кібербулінг може вплинути на фізичне здоров'я жертви через стрес, неспокій та тривогу.

Для того щоб запобігти кібербулінгу насамперед потрібно знати що це таке потрібно знати про його наслідки, це зможе допомогти користувачам уникнути втрат та стати менш схильними до цієї форми насильства.

Соціальні мережі та інші онлайн-платформи надають можливості для налаштування конфіденційності та блокування користувачів.

Повідомлення про кібербулінг: Інтернет-платформи повинні надавати можливість користувачам повідомляти про кібербулінг і блокувати агресивних користувачів[6].

Законодавство і правоохоронні органи: Багато країн мають законодавство, яке криміналізує кібербулінг. Жертви можуть звертатися до правоохоронних органів для захисту.

Підтримка та консультування: Жертви кібербулінгу можуть отримувати підтримку від психологів, консультантів, а також від родини та друзів..

Фейкові новини, або "fake news," - це дезінформація, яка поширюється через різні засоби масової інформації, особливо через соціальні мережі, з метою створення обманливого враження про події або інформацію.

Ось опис феномену фейкових новин на соціальних мережах:

Соціальні мережі дозволяють великій кількості користувачів легко ділитися інформацією, незалежно від її достовірності. Якщо фейкова новина вирішить припасти до глибин публічності, вона може швидко стати вірусною та розповсюдитися по всьому світу.

Фейкові новини можуть вплинути на думку громадськості і вибір громадян під час виборів або на інші значущі питання. Вони можуть бути використані для маніпулювання настроями та поглядами суспільства.

Декілька прикладів, які демонструють вплив дезінформації на суспільство:

1. Вибори та політика: Фейкові новини можуть впливати на результати політичних виборів, викликаючи підозру у виборців і сприяючи розпалюванню політичних конфліктів. Наприклад, у 2016 році багато

фейкових новин поширилися під час президентських виборів у США, що вплинули на голосування та перекрутили обговорення головних тем.

2. Здоров'я і безпека: Дезінформація може мати серйозний вплив на громадське здоров'я. Наприклад, неправдиві статті про пандемію COVID-19 можуть призвести до недостатнього розуміння та дотримання заходів безпеки, що може призвести до подальшого поширення вірусу.
3. Економіка: Фейкові новини можуть впливати на фінансові ринки. Реклама неправдивої інформації про компанії може призвести до спаду цін акцій або їх підвищення, що може призвести до великих втрат для інвесторів.


Рисунок 1.4 – Ознаки фішингових атак

Злам аккаунтів і крадіжка ідентичності - це серйозні проблеми в цифровому світі. Вони можуть призвести до втрати особистої інформації, приватності і навіть негативних наслідків для жертв. Далі ми розглянемо детальніше про ці поняття та засоби захисту!

Методи злому аккаунтів:

1. Перехоплення паролів: Зловмисники можуть спробувати вгадати пароль або використовувати "брутфорс" - метод спроби всіх можливих паролів - для отримання доступу до аккаунту.

2. Соціальна інженерія: Зловмисники можуть намагатися отримати інформацію від жертв, використовуючи маніпуляцію або фальшиві обіцянки.
3. Фішинг: Зловмисники можуть надсилати фальшиві повідомлення або посилання, які виглядають, як вони від інтернет-провайдерів, банків або соціальних мереж, з метою обману жертви і злому її аккаунту.

Наслідком цього може бути: втрата особистої інформації, крадіжка ідентичності, фінансові втрати та інше.

Злом аккаунту може призвести до втрати особистих даних, таких як паролі, фотографії, адреси електронної пошти тощо. Зловмисники можуть використовувати сторонній аккаунт для поширення дезінформації або крадіжки ідентичності жертви. Якщо зловмисники мають доступ до фінансових акаунтів або кредитних карток через злам аккаунтів, це може призвести до фінансових втрат[7].

Для того щоб захистити особисту інформацію у соціальних мережах потрібно дотримуватися простих правил:

Потрібно використовувати складні паролі, які складаються з букв, цифр і символів, і змінюйте їх регулярно. Не використовуйте один і той самий пароль для кількох акаунтів.

Більшість соціальних мереж дає можливість користувачам використовувати двофакторну аутентифікація (2FA): Ввімкніть 2FA на всіх своїх аккаунтах, де це можливо. Це додасть додатковий рівень безпеки.

Будьте обережні з посиланнями та повідомленнями не клікайте на підозрілі посилання або відповідайте на підозрілі повідомлення. Перевіряйте, чи вони відповідають домену офіційного веб-сайту.

Також всі соціальні мережі мають параметри приватності, які дозволяють обмежити, хто бачить вашу інформацію і дописи. Налаштуйте їх на рівні, який вам зручний.

Та найголовніше, будьте обережні з особистою інформацією не надавайте її невідомим особам або публічно.

Моніторинг активності користувачів та виявлення ненормальної активності є важливими аспектами для забезпечення безпеки та правильної роботи онлайн-систем, таких як веб-сайти, соціальні мережі, електронні комерційні платформи та інші. Для цього використовується аналітика, яка допомагає виявляти зміни в користувацькій активності, які можуть вказувати на аномалії або потенційні загрози.

Ось декілька основних способів, які використовуються для виявлення ненормальної активності:

1. Порівняльний аналіз: Системи можуть порівнювати активність користувача зі стандартами або з нормами, що є для цього типу аккаунту звичайними. Якщо активність значно виходить за рамки звичайних показників, то це може бути ознакою ненормальної активності.
2. Виявлення надзвичайних подій: Системи можуть виявляти надзвичайні події, такі як надто велика кількість неуспішних спроб входу в систему, різке збільшення відправлених запитів чи несподівані зміни в профілі користувача.
3. Машинне навчання: Використання алгоритмів машинного навчання для навчання системи розпізнавати аномальну активність на основі історичних даних. Такі алгоритми можуть автоматично виявляти нові шаблони аномальної активності.
4. Аналіз ризику: Оцінка ризику для кожної активності користувача на основі різних факторів, таких як місцезнаходження, тип пристрою, інші признаки. Якщо ризик виявляється надто високим, система може ввести додаткові заходи безпеки, такі як двофакторну аутентифікацію або блокування аккаунту.
5. Поведінкова аналітика: Вивчення поведінки користувачів на основі їхніх дій, щоб виявити аномалії. Наприклад, якщо зазвичай користувач реєструється і публікує пост в соціальній мережі лише один раз на тиждень, то публікація декількох постів за годину може вказувати на аномалію.

6. Аналіз вмісту: Виявлення аномалій на основі аналізу текстового або візуального вмісту, таких як фільтрація спаму, образливого вмісту або виявлення несправжніх новин.

Машинне навчання та аналіз тексту можуть бути важливими інструментами для виявлення фейкових новин та агресивного мовлення в великих обсягах текстової інформації. Нижче розглянемо, як це працює.

Виявлення фейкових новин:

Машинне навчання може використовуватися для визначення тональності тексту. Фейкові новини часто мають емоційно забарвлені заголовки або вміст, які можуть бути виявлені за допомогою алгоритмів аналізу емоційного виразу. Машинне навчання може аналізувати стиль публікації та виявляти аномалії. Фейкові новини можуть використовувати незвичайні стилістичні рішення або відмінності від стандартів новинарського письма. Аналіз тексту може включати перевірку авторства та джерела інформації. Машинне навчання може допомогти визначити, чи є джерело достовірним або має історію розповсюдження фейкових новин. Машинне навчання також може використовуватися для перевірки фактів, представлених в новинах. Це може включати перевірку фактичних помилок, порівняння із схожими новинами та використання відкритих джерел даних для підтвердження інформації[8].

Виявлення агресивного мовлення: машинне навчання може аналізувати текст на предмет агресивних, образливих або загрозливих виразів. Сентиментний аналіз може визначити загальний тон тексту та виділити агресивну лексику. Для виявлення агресивного мовлення в коментарях на соціальних мережах або форумах можуть використовуватися алгоритми класифікації тексту. Токсичні коментарі можуть бути виявлені на підставі образливого мовлення та загроз. Машинне навчання може визначати агресивні моделі взаємодії між користувачами, наприклад, визначаючи шаблони булінгу або цілеспрямованих атак.

Виявлення аномалій - це важлива складова систем безпеки та моніторингу в багатьох сферах, включаючи кібербезпеку, фінанси, медицину та багато інших

галузей. Використовуючи методи виявлення аномалій, можна виявити незвичайні дії користувачів або відхилення від звичайних моделей поведінки. Ось деякі методи та підходи до виявлення аномалій для реагування на незвичайні дії користувачів:

Методи статистичного виявлення аномалій:

1. Засновані на відхиленнях: Ці методи використовують статистичні метрики, такі як середнє значення і стандартне відхилення, для визначення, які значення є аномальними на основі їхнього відхилення від середнього.
2. Засновані на розподілі: Ці методи використовують інформацію про розподіл даних і визначають аномалії на основі того, наскільки вони відхиляються від очікуваного розподілу.
3. Навчання з учителем: Використовується для навчання моделі на основі історичних даних, включаючи як нормальні, так і аномальні випадки. Після навчання модель може визначати, чи є нові дані аномаліями.
4. Навчання без учителя: Це методи, в яких модель навчається на основі нормальних даних без явної інформації про аномалії. Це дозволяє виявляти незвичайні відхилення від очікуваної поведінки.
5. Нейронні мережі із вирівнюваннями: Мережі, такі як вирівнювання автоенкодерів або глибокі згорткові мережі, можуть використовуватися для виявлення аномалій у структурованих або зображеннях.
6. Виявлення аномалій на основі послідовностей: Ці методи можуть використовувати рекурентні нейронні мережі для виявлення аномалій в послідовних даних, таких як часові ряди або текст.
7. Використання комбінації різних методів, включаючи як статистичні, так і машинне навчання, може підвищити точність виявлення аномалій.

Для виявлення аномалій у діях користувачів, такі системи можуть аналізувати історічні дані та реальний потік інформації, виявляючи незвичайні відхилення. Це може бути корисно в кібербезпеці для виявлення зловмисних атак, в фінансовому секторі для виявлення шахрайств або в медицині для виявлення аномальних медичних станів.

Спам і фішинг - дві зовсім різні, але однаково небезпечні загрози в онлайн-світі. Ці атаки можуть призвести до втрати особистої інформації, фінансових втрат і порушення приватності користувачів. Щоб боротися з цими проблемами, існують різні технології та підходи[9].

Спам - це не просто надокучливі листи в електронній пошті, але і важлива проблема в онлайн-світі.

1. Боротьба на рівні поштових серверів: Використання чорних списків, аналіз заголовків та вмісту листів для виявлення спаму.
2. Фільтри в клієнтських програмах: Популярні поштові клієнти мають вбудовані фільтри для виділення спаму.
3. Використання CAPTCHA: Захист від автоматизованих атак. Фішинг - це хибний імітація легітимних джерел для обману користувачів.
4. Детектори фішингу: Виявлення підозрілих листів та веб-сайтів, що намагаються видати себе за довірені джерела.
5. Використання цифрових підписів та сертифікатів: Для перевірки автентичності відправника.
6. Освіта користувачів: Навчання користувачів розпізнавати фішингові атаки.

Боротьба зі спамом і фішингом - це постійний процес, який вимагає комбінації технологій та освіти.

Незалежно від того, наскільки розвинуті технології захисту, важливо пам'ятати, що основою є уважність і пильність користувачів в онлайн-світі. Виявлення кіберзагроз на соціальних мережах є критичним завданням для забезпечення безпеки користувачів та захисту їхньої приватності. Сучасні розробки та програми використовують різні методи та технології для виявлення кіберзагроз на соціальних мережах.

Ось деякі з них:

Виявлення фейкових акаунтів та ботів:

1. Botometer: Ця програма використовує машинне навчання для виявлення ботів на Twitter. Вона аналізує активність акаунта, його мережу зв'язків та інші параметри для виявлення підозрілих активностей.
2. Socedo: Спеціальна платформа для виявлення ботів та фейкових акаунтів на Twitter та інших соціальних мережах. иВиявлення фейкових новин та дезінформації:
3. NewsGuard: Розширення для веб-переглядачів, яке надає рейтинги новинарським джерелам на основі їхньої достовірності.
4. FactCheck.org і Snopes: Веб-сайти, які спеціалізуються на перевірці фактів та виявленні фейкових новин. Аналіз тональності та агресивного мовлення:
5. Perspective API: Google розробив цей інструмент для виявлення токсичних коментарів та агресивного мовлення в реальному часі.
6. Hatebase: База даних, яка визначає та моніторить слова та вирази, що використовуються для вираження ненависті в онлайн-спільнотах. Виявлення аномалій та вторгнень:
7. Darktrace: Платформа для виявлення аномалій в мережах та інформаційних системах. Вона використовує штучний інтелект для пошуку відхилень в звичайній активності.
8. Splunk: Платформа для моніторингу та аналізу журналів інформаційних систем для виявлення вторгнень та аномалій. Засоби для аналізу глибокого контенту:
9. Jigsaw's Perspective API: Інструмент від Google для аналізу вмісту та виявлення образливого мовлення, загроз та іншої небажаної активності в коментарях та вмісті соціальних мереж.
10. OpenAI's GPT-3: Може використовуватися для аналізу текстового вмісту та перевірки на наявність образливого чи засудженного мовлення.

Ці інструменти та програми представляють сучасні розробки у сфері кібербезпеки та виявлення кіберзагроз на соціальних мережах. Вони допомагають покращити безпеку і забезпечити більший контроль над тим, що

публікується та обговорюється в цих мережах.

Ось декілька прикладів успішних ініціатив на популярних соціальних мережах:

Facebook - Виявлення ботів і фейкових акаунтів: Facebook вдосконалив алгоритми виявлення підозрілих активностей і фейкових акаунтів. Вони регулярно видаляють десятки тисяч фейкових акаунтів, які використовуються для розповсюдження спаму та дезінформації. Автоматичне виявлення токсичних коментарів: Facebook застосовує машинне навчання для виявлення та блокування образливих та агресивних коментарів.

Twitter - Посилення боротьби з ботами: Twitter зробив кроки для виявлення та призупинення активності ботів, які сприяють розповсюдженню спаму та фейкових новин. Перевірка відомостей: Twitter розглядає інформацію, що розповсюджується у популярних темах, для виявлення неправдивих чи маніпулятивних даних.

Instagram - Захист від булінгу та непристойних вмістів: Instagram використовує автоматичні фільтри та машинне навчання для виявлення та блокування образливого вмісту та незаконних дій. Захист від фішингу: Instagram впровадив перевірку безпеки облікового запису, вимагаючи двоетапну аутентифікацію для деяких дій.

LinkedIn - Виявлення неправдивих облікових записів: LinkedIn використовує алгоритми для виявлення фейкових профілів та спаму в мережі. Зміцнення безпеки повідомлень: Використання шифрування та автентифікації для захисту особистих даних користувачів.

Ці приклади показують, що соціальні мережі активно вживають заходи для виявлення і запобігання кіберзагрозам. Вони постійно вдосконалюють свої технології та співпрацюють зі спеціалізованими командами безпеки, щоб забезпечити безпеку та конфіденційність своїх користувачів. Однак боротьба з кіберзагрозами є постійним процесом, і важливо, щоб користувачі також були пильні та виявляли підозрілу активність.

Витік інформації: Витік інформації відбувається, коли конфіденційні дані

ненавмисно або навмисно потрапляють до сторонніх осіб або організацій. Це може статися через надмірний обмін інформацією в соціальних мережах, коли люди розкривають особисту інформацію, яка може бути використана для крадіжки особистих даних або цілеспрямованих атак. Витік інформації також може статися через неправильне поводження з конфіденційними документами, неналежну утилізацію фізичних чи цифрових записів або крадіжку пристроїв, що містять конфіденційну інформацію.

Компрометація співробітників: Компрометація співробітників передбачає використання довіри до них для отримання несанкціонованого доступу до систем компанії або конфіденційної інформації. Зловмисники можуть видавати себе за керівників або IT-персонал, обманом змушуючи працівників виконувати певні дії або розкривати конфіденційну інформацію. Ще один метод, коли зловмисники створюють правдоподібний сценарій або історію, щоб маніпулювати працівниками, змушуючи їх ділитися конфіденційними даними або надавати несанкціонований доступ.

Зниження загроз соціальної інженерії вимагає поєднання обізнаності користувачів, протоколів безпеки та технологічних заходів. Інформування людей про ризики, впровадження надійних механізмів автентифікації, регулярне оновлення паролів, проведення аудитів безпеки та підвищення обізнаності про безпечні онлайн-практики можуть допомогти захиститися від загроз соціальної інженерії.

Відсутність доступу до конкретних даних користувачів змушує нас використовувати альтернативні методи для побудови соціального графа. Ми маємо можливість аналізувати лише публічні сторінки у соціальній мережі. Кожна з цих сторінок може містити публікації, які, у свою чергу, можуть мати коментарі. Кожен коментар може також містити інші коментарі і так далі. До 2018 року Facebook API надавав можливість отримувати дані тільки про тих, хто коментував публікації у вигляді певних хешованих ідентифікаторів. Це означає, що ми можемо встановити, хто залишав коментарі, але не можемо точно визначити їх реальну особистість. Однак, якщо одна і та ж особа залишає

коментарі під різними публікаціями, ми можемо виявити цю спільну аудиторію, не розголошуючи їхню ідентичність. Аналізуючи такі дані для всіх публікацій, ми можемо побудувати множину, яка представляє схожість аудиторій в межах соціальної мережі. Ці дані можуть бути використані для ранжування відносин між різними публічними сторінками. Наш підхід базується на використанні методу data fusion для аналізу та об'єднання даних[10].

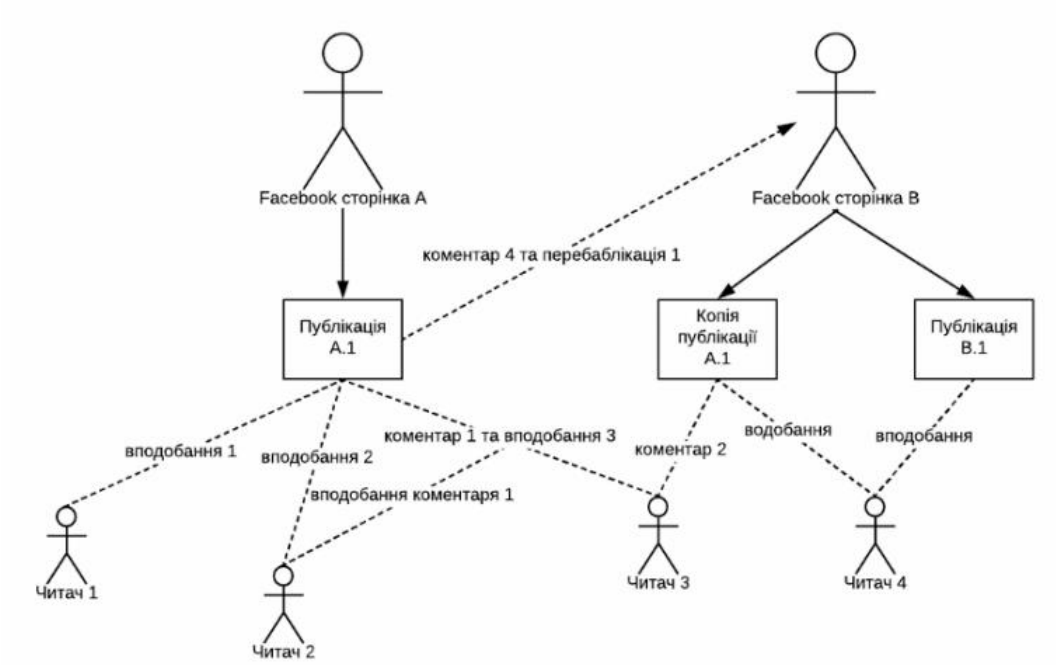



Рисунок 1.5 – Структура спілкування facebook.com

Попередній підхід до побудови соціального графу на основі коментарів став неможливим через зміни у Facebook API після скандалу з Cambridge Analytica у 2018 році. Тепер соціальна мережа не надає доступ до даних про користувачів, які коментують публікації. Тому ми повинні знайти інший шлях для визначення особистостей.

Наш новий підхід використовує аналіз семантики коментарів та наявність посилань у коментарях. Ми порівнюємо слова у коментарях для визначення їх схожості, а також перевіряємо, чи містять коментарі посилання на однакові домени. Кожній метриці призначається вага, і ці ваги можна налаштовувати.

Після визначення схожості коментарів, ми можемо відновити відносини

між користувачами. Хоча ми не знаємо точно, хто коментував публікації, ми можемо вказати, наскільки схожі ці користувачі. Це дозволяє нам встановлювати зв'язки на соціальному графі і проводити аналіз мережі.

Маючи доступ до історичних даних про існуючі соціальні графи, ми можемо порівнювати їх зі свіжоствореними графами для налагодження параметрів метрик. Для цього ми використовуємо різні метрики графів, які були описані в теоретичній частині.

Процес калібрування включає багато порівнянь між двома графами на основі цих метрик. Ми продовжуємо цей процес до тих пір, поки метрики старого архівного графа не стануть схожими на метрики поточного графа. Після завершення калібрування ми отримуємо структуровані соціальні графи, які можна використовувати для подальшого аналізу.

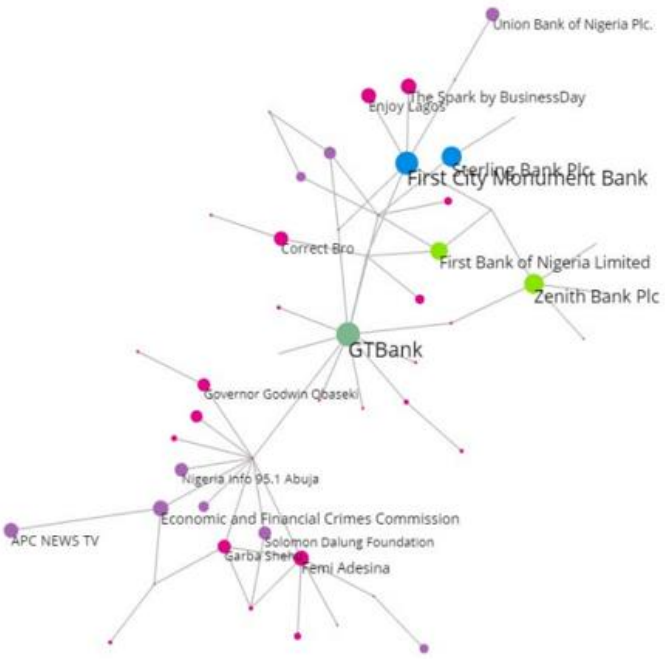



Рисунок 1.6 – приклад побудови соціального графу мережі

У цьому розділі було розглянуто важливі аспекти виявлення кіберзагроз у соціальних мережах[11]. Виявлення кіберзагроз є актуальним завданням, оскільки соціальні мережі стали неодмінною частиною сучасного життя та комунікаційного простору. Основні тези можуть бути сформульовані наступним

чином:

1. Типи соціальних мереж і їх генерація: Відкрито багато різних типів соціальних мереж, включаючи соціальні мережі загального споживання, професійні мережі та багато інших. Вони мають свої особливості та структуру, що впливає на методи аналізу та виявлення кіберзагроз.
2. Класифікація кіберзагроз у соціальних мережах: Різні види кіберзагроз можна класифікувати залежно від їх характеру та методів реалізації, такі як загрози соціальної інженерії. Загрози, які включають розміщення приманок, фішінг, витік інформації та інші, можуть спричиняти серйозні проблеми для користувачів соціальних мереж.
3. Методи виявлення кіберзагроз: Виявлення кіберзагроз у соціальних мережах використовує різноманітні методи, включаючи аналіз активності користувачів, моніторинг веб-атак, аналіз тексту та багато інших. Важливо підкреслити, що інтеграція різних методів може покращити ефективність виявлення кіберзагроз та підвищити безпеку соціальних мереж.

В цілому, виявлення кіберзагроз у соціальних мережах є складною задачею, але вкрай важливою для забезпечення безпеки користувачів та збереження довіри до цих платформ. Розглянуті методи та класифікації можуть бути використані для розробки систем виявлення кіберзагроз та забезпечення безпеки у соціальних мережах.

# 2. МЕТОДИ МОНІТОРИНГУ ТА АНАЛІЗУ СПІЛЬНОТ У СОЦІАЛЬНИХ МЕРЕЖАХ ЗА ДОПОМОГОЮ ТЕОРІЇ ГРАФІВ

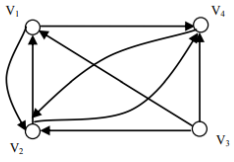
## 2.1. Графові моделі та їх характеристики матриці подання

Подання графів у вигляді малюнків - це зручний спосіб, але досить обмежений, особливо коли граф досить великий (з великою кількістю вершин). Для математичного опису графів зазвичай використовують матриці.Якщо ми маємо простий орграф G=(V, E), де V={v1, v2, …, vn} - це множина вершин, які впорядковані від v1 до vn, то можна створити квадратну матрицю A розмірністю nxn, де елементи цієї матриці визначаються за певним виразом.

$$a_{ij} = \left[\frac{1,\text{якщо } (V_i, V_j,) \epsilon E}{0, \text{ інакше}}\right] \quad (2.1)$$

Матриця суміжності (МС) графа G - це таблиця, де елементи можуть бути 0 або 1, що вказує на наявність або відсутність зв'язку між вершинами. Ця матриця може представляти граф як набір бітів або значень true/false. Якщо можна отримати одну МС з іншої, переставляючи її рядки і/або стовпці, то ці графи є еквівалентними[12].

Матричне представлення може використовуватися не лише для простих графів, але і для графів з багатьма зв'язками чи вагами ребер. Тут значення елементів можуть бути не тільки 0 та 1, а й числами, що вказують кількість зв'язків чи вагу ребра. Для неорієнтованих графів МС є симетричною.Якщо у графі є тільки окремі вершини без зв'язків (ізольовані вершини), МС буде містити лише 0. Якщо тільки петлі та інші зв'язки відсутні, або ж є тільки петлі, то МС буде мати вигляд одиничної матриці. На малюнку 2.4 показано простий напрямлений граф та його матрицю суміжності A.




| | v1 | v2 | v3 | v4 |
|---|---|---|---|---|
| v1 | 0 | 1 | 0 | 1 |
| v2 | 1 | 0 | 0 | 1 |
| v3 | 1 | 1 | 0 | 1 |
| v4 | 0 | 1 | 0 | 0 |

Рисунок 2.1 – приклад графа та його матриця суміжності

Розглянемо матриці, які представляють квадрати матриці суміжності A: A2, A3, ..., An, де елементи матриці позначені як a n ij. Давайте сконцентруємось на матриці A2. Елементи цієї матриці визначаються певною формулою

$$.a_{ij}^2 = \sum_{k=1}^{n} a_{ij} * a_{kj} \quad (2.2)$$

Коли в графі існують ребра між вершинами vi і vk, а також між vk і vj, отримуємо шлях довжиною 2 між вершинами vi і vj. Матриця A2 відображає кількість таких шляхів довжиною 2 між вершинами. За допомогою індукції або рекурсії можна показати, що матриця An відображає кількість елементарних шляхів довжиною n між вершинами vi і vj. Сумуючи матриці від A до An, матриця B показує, скільки існує елементарних шляхів з вершини vi до вершини vj і чи є вони взагалі. Якщо значення в матриці B не дорівнює нулю, то вершина vj є досяжною з вершини vi.

$$p_{ij} = \left[\frac{1, \text{якщо існує шлях з } V_i \text{ в } V_j,}{0, \text{ інакше}}\right] \quad (2.3)$$

Для визначення досяжності можна використовувати матрицю P, яку ще називають шляховою матрицею для графа G[13]. Щоб її отримати, використовують булеві матричні операції. Таблиця розміром 3 на 3 (див. рис. 2.2а) спочатку порожня, і в неї можна внести граф, зображений на рис. 2.2б. Після кожного кроку на вільному місці в цій таблиці з'являється кулька, яку можна перемістити відповідно лише по горизонталі або вертикалі. Наприклад, на першому кроці кулька з'явилася на клітинці 6 (див. рис. 2.2в). На другому кроці кулька з'явилася на клітинці 3, і перемістити її на клітинку 5 (див. рис. 2.2в). На малюнку 2.3 подані матриці суміжності для графів, зображених на рисунку 2.2б і 2.2г (для графів на рис. 2.2е і 2.2ж).

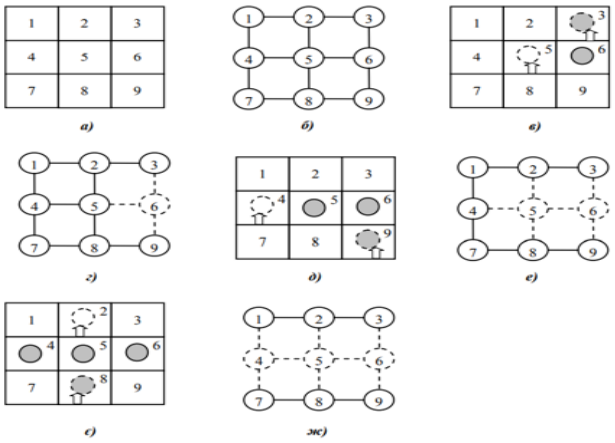


Рисунок 2.2 – Відображення стану таблиці графами

На рисунку 2.4 показано спосіб знаходження шляху від вершини vi до vj, використовуючи різні ступені матриці суміжності, позначені як An, де n - це степінь матриці A. Для визначення наявності шляху спочатку перевіряється значення конкретного елемента aij у матриці A.

| Граф на рис. | Шукані шляхи та елементи матриці | | Шляхові матриці | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $v_i \rightarrow v_j$ | $a_{ij}^n$ | $A^1$ | $A^2$ | $A^3$ | $A^4$ | $A^5$ | $A^6$ | $A^7$ | $A^8$ | $A^9$ |
| 2.5г | 3→5 | $a_{35}^n$ | 0 | 1 | | | | | | | |
| 2.5е | 9→4 | $a_{94}^n$ | 0 | 0 | 1 | | | | | | |
| 2.5ж | 8→2 | $a_{82}^n$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Рисунок 2.3 – Визначення досяжності вершини vj із вершини vi

Якщо aij = 1, це означає наявність шляху довжиною 1; якщо aij = 0, розраховується aij з використанням ступеня n = 2, 3, ..., N, доки не з'явиться ненульове значення. Це число вказує на кількість шляхів vi → vj довжиною n. Якщо всі aij для певного n рівні нулю, шляху не існує. Шлях можна відтворити, розглядаючи кожний вираз[14].

$$\sum - kj\ n\ aik\ *a \qquad (2.4)$$

У математичній галузі теорії графів теорема Кірхгофа або теорема Кірхгофа про матричне дерево, названа на честь Густава Кірхгофа , — це теорема про кількість охоплюючих дерев у графі , яка показує, що це число можна обчислити за поліноміальний час з визначника підматриці графа . Матриця Лапласа графа; зокрема, число дорівнює будь-якому кофактору матриці Лапласа. Теорема Кірхгофа є узагальненням формули Кейлі , яка визначає кількість остовних дерев у повному графі . Теорема Кірхгофа спирається на поняття лапласівської матриці графа, яка дорівнює різниці між матрицею ступенів графа (діагональна матриця зі ступенями вершин на діагоналях) і його матрицею суміжності ((0,1)-матриця з одиницями у місцях, що відповідають записам, де вершини є суміжними та 0 в іншому випадку). Для заданого зв'язного графа G з n позначеними вершинами нехай λ 1, λ 2, …, λn −1 є ненульовими власними значеннями його матриці Лапласа. Тоді кількість остовних дерев G дорівнює

$$t(G) = \frac{1}{n}\ \lambda_1\lambda_2\ldots\ldots\lambda_{n-1} \qquad (2.5)$$

Спочатку будеєм матрицю Лапласа Q для прикладу діамантового графа G (див.рис. 2.4):

$$Q = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}.$$

Рисунок 2.4 –матрицяі Лапласа Q

Далі матрицю Q *, видаливши будь-який рядок і будь-який стовпець з Q . Наприклад, видалення рядка 1 і стовпця 1 дає результат

$$Q^* = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

Рисунок 2.5 – Матриція Q

Визначник Q *, щоб отримати t ( G ) , яке дорівнює 8 для діамантового графіка. (Звернемо увагу, що t ( G ) є (1,1) - кофактором Q у цьому прикладі.)

Схема доведення:

(Наведене нижче доказ базується на формулі Коші-Біне . Елементарний індукційний аргумент для теореми Кірхгофа можна знайти на сторінці 654 Мура (2011). [1] )

Матриця Лапласа має властивість, що сума її записів у будь-якому рядку та будь-якому стовпці дорівнює 0. Таким чином, ми можемо перетворити будь-який мінор на будь-який інший мінор, додаючи рядки та стовпці, міняючи їх і множачи рядок або стовпець на −1. Таким чином, співмножники однакові з точністю до знака, і можна перевірити, що насправді вони мають однаковий знак.

Визначник мінору M 11 підраховує кількість остовних дерев. Нехай n — кількість вершин графа, а m — кількість його ребер. Матриця інцидентності E є матрицею розміром n на m , яка може бути визначена таким чином: припустимо, що ( i , j ) є k-м ребром графа, і що i < j . Тоді Eik = 1, Ejk = −1, а всі інші записи в стовпці k дорівнюють 0 (дивіться орієнтовану матрицю інцидентності , щоб зрозуміти цю модифіковану матрицю інцидентності E ). Для попереднього прикладу (з n = 4 і m = 5):

Лапласіан L можна розкласти на множники добутку матриці інцидентності та її транспонування, тобто L = EE [14]. Крім того, нехай F буде матрицею E з видаленим першим рядком, так що FF T = M 11.

Тепер формула Коші-Біне дозволяє нам писати

$$\det\left(\mathrm{M}_{11)}\right) = \sum_{\mathrm{S}} \det(\mathrm{Fs}) \det) \mathrm{F}_{\mathrm{S}}^{\mathrm{T}} ) = \sum_{\mathrm{S}} \det(\mathrm{Fs})^2 \qquad (2.6)$$

де S розташовується між підмножинами [ m ] розміром n − 1, а FS позначає матрицю ( n − 1) на ( n − 1), стовпці якої є стовпцями F з індексом у S . Тоді кожен S визначає n − 1 ребер початкового графа, і можна показати, що ці ребра індукують остовне дерево тоді і тільки тоді, коли визначник FS дорівнює +1 або −1, і що вони не індукують остовне дерево, якщо і тільки якщо визначник дорівнює 0. На цьому доведено.

Як вирішити:

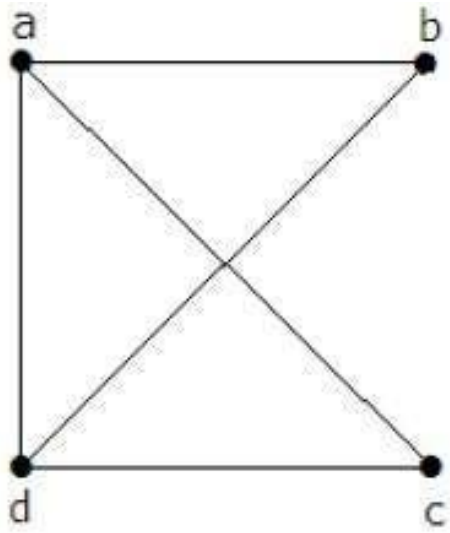



Рисунок 2.6 – матриця a-c

Матриця "A" має бути заповнена так: якщо між двома вершинами є ребро, то воно має бути задано як "1", інакше як "0". Для обчислення центральності посередності (betweenness centrality) вершин у графі потрібно знайти кількість найкоротших шляхів між всіма парами вершин та кількість таких шляхів, які проходять через кожну вершину.

Спочатку знайдемо всі можливі найкоротші шляхи між парами вершин:

0 -> 1: [0, 1] 0 -> 2: [0, 2] 0 -> 3: [0, 3] 1 -> 2: [1, 2] 1 -> 3: [1, 0, 3], [1, 3] 2 -> 3: [2, 0, 3], [2, 1, 3], [2, 3]

Тепер для кожної вершини знайдемо кількість шляхів, які проходять через неї:

Для вершини 0:

Немає шляхів через 0 для 1-2, 1-3, 2-3.

Для вершини 1:

Шляхи через 1: 0-3.

Для вершини 2:

Шляхи через 2: 0-3, 1-3.

Для вершини 3:

Шляхи через 3: 1-0-3, 2-0-3, 2-1-3.

Отже, кількість шляхів, які проходять через кожну вершину:

Для вершини 0: 0

Для вершини 1: 1

Для вершини 2: 2

Для вершини 3: 3

Це показує кількість найкоротших шляхів, які проходять через кожну вершину[15]. Тепер можемо обчислити значення центральності посередності для кожної вершини за формулою, де кількість шляхів, які проходять через вершину v, ділимо на загальну кількість можливих шляхів між всіма парами вершин:

Де CB(v) - міра центральності посередності для вершини v, а )σst(v) - кількість найкоротших шляхів між s та t, які проходять через вершину v, а σst - загальна кількість найкоротших шляхів між s та t.

У цьому випадку маємо:

Для вершини 0: CB (0)=0

Для вершини 1:  CB (1)=1/6

Для вершини 2: CB(2)=2/6

Для вершини 3:  CB (3)=3/6

Алгоритм Краскала потребує роботи з зваженими ребрами. Граф не має ваг, тому необхідно додатково задати ваги ребрам. Припустимо, що ваги ребер у графі виглядають наступним чином:

0 -- 1 (вага 1)

0 -- 2 (вага 1)

0 -- 3 (вага 1)

1 -- 3 (вага 1)

2 -- 3 (вага 1)

Тепер можемо застосувати алгоритм Краскала:

Сортування ребер за вагою: Ребра сортуються за їх вагою у зростаючому порядку.

Початок з пустого кістякового дерева: Починаємо з порожнього кістякового дерева.

Поступове додавання ребер: Додаємо ребра по черзі в порядку зростання їх ваги. У разі, якщо додаване ребро не утворює цикл разом з вже доданими раніше ребрами, воно включається до кістякового дерева.

Візуально це можна представити таким чином:

Початковий граф:

0 -- 1
| / |
|/ |
2 – 3

Після застосування алгоритму Краскала:

0 -- 1
|
2 – 3

Отже, мінімальне кістякове дерево, знайдене алгоритмом Краскала для даного графа, буде виглядати так: вершини 0, 1, 2, 3 пов'язані між собою ребрами так, що утворюється мінімальне кістякове дерево без утворення циклів із заданими ребрами[16].

Виявляйте слабкі місця у своїй мережі за допомогою Betweenness Centrality

Кожен процес або мережа має дуже важливі ресурси, про які вам потрібно особливо піклуватися — будь то важлива частина даних, розробник-єдиноріг, якого ви найняли, чи дорога частина інфраструктури. Отже, кожен набір даних графа має дуже важливі вузли. Деякі вузли є важливими, тому що вони мають

вирішальне значення для успішної роботи вашої системи, але деякі (іноді ті самі) вузли важливі, тому що, якщо вони вийдуть з ладу, вони можуть спричинити хаос.

Щоб дізнатися, які вузли в мережі є важливими на основі топологічної структури мережі та, отже, мають відношення до її успіху чи краху, запустіть аналіз центральності вузлів у вашій графовій базі даних. Заходи аналізу центральності

Аналіз центральності можна виконати за допомогою вимірювань, які перевіряють ступінь вузла або короткі шляхи .

Ступінь — це кількість зв’язків, які має певний вузол. Коли важливі лише зв’язки певного вузла, аналіз виконується з використанням ступеня центральності . Але, якщо ступінь оточуючих вузлів також включено в рівняння, аналіз виконується за допомогою центральності власного вектора .

Як і в реальному житті, інколи важливо мати багато друзів, але іноді важливо мати друзів, у яких є багато інших друзів (політики зламали це).

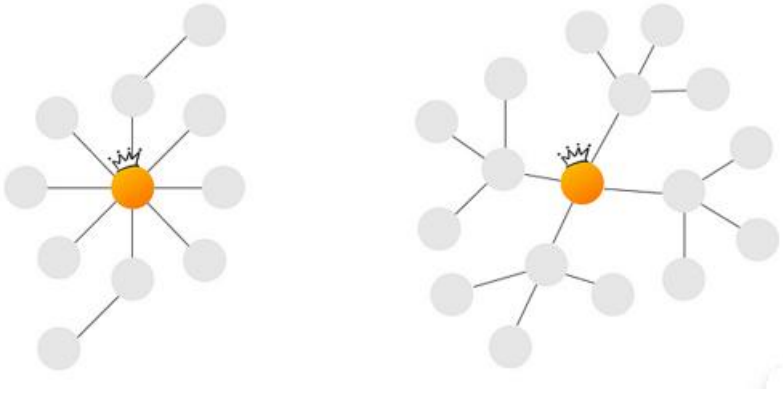


Рисунок 2.7 –вузол із високим ступенем центральності та високим показником власного вектору

Інший спосіб поглянути на важливість вузла — дослідити кількість найкоротших шляхів, частиною яких є вузол[17]. Щоб дізнатися, який вузол поширює інформацію найшвидше, оскільки він є найближчим до багатьох інших вузлів, проаналізуйте графік за допомогою центральності близькості ( і знайдіть

одну людину, яка дружить з усіма іншими і знає все).

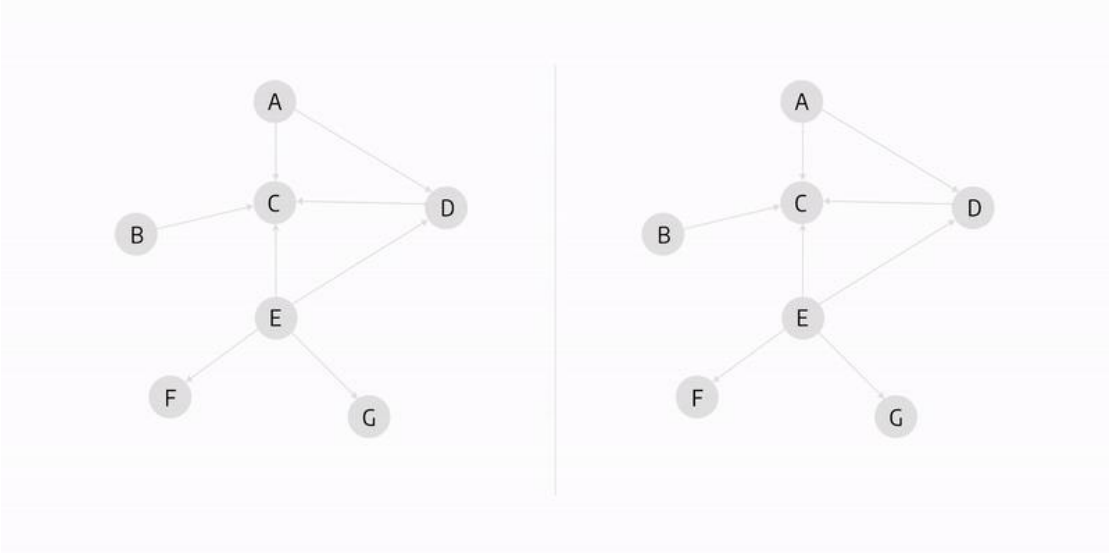



Рисунок 2.8 –вузли із високим ступенем центральності

На графіку вище вузли C і E мають високу центральність, оскільки вони можуть отримати доступ до всіх інших вузлів найшвидше.

Щоб ідентифікувати вузли, які контролюють передачу інформації, вам потрібно з’ясувати центральність між вузлами[18] .

За винятком того, що він мучить вас орфографією, центральність між ними виявляє вузли, які мають значний вплив на мережу, оскільки вони відіграють роль мосту. Центральність між вузлами визначається як кількість найкоротших шляхів, які проходять через вузол, поділена на загальну кількість найкоротших шляхів між усіма парами вузлів.

Це людина, яка об’єднує багатьох різних людей. Але коли цей зв’язок не в місті, деякі з його друзів залишаються ночувати наодинці перед телевізором.

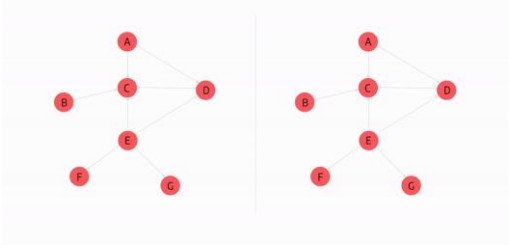


Рисунок 2.9 – вузли із найвищими ступенями центральності

На наведеному вище графіку вузол Е є вузлом із найвищим показником центральності проміжності, оскільки, якщо він стане неробочим, він від’єднає найбільшу кількість вузлів від решти графіка[19].

Іншими словами, у той момент, коли вузол із високим показником центральності між будь-яким чином не виконує те, для чого він був розроблений, настав час виправити проблеми, оскільки деякі вузли більше не приєднані до мережі. Отже, давайте подивимося на центральність проміжності ближче, щоб дізнатися, як уникнути проблем.

Міжцентрованість може допомогти виявити больові точки в мережах і графіках знань, побудованих навколо різних галузей.

Під час аналізу відомої організації шахрайства центральність зв’язків допомагає визначити вузли, які діють як мости між клієнтами, оскільки вони з більшою ймовірністю скоїть шахрайство. Після підтвердження підозри дані можуть бути передані в модель машинного навчання для виявлення підозрілої поведінки та кластерів у великих наборах даних або прогнозування шахрайства.

Якщо говорити про шахрайство, то багато інформації, товарів і діяльності протікає через важливі вузли злочинної мережі, особливо через ті з високими показниками центральності між ними. Ці вузли можуть бути особливо цікавими для зриву. Виявлення та припинення найефективнішого шахрая може сповільнити діяльність усієї злочинної мережі.

Міра центральності між ними може точно визначити, які ресурси контролюються обмеженою кількістю осіб. Доступ, обмежений лише кількома ключовими особами, може сповільнити потік інформації, якщо деякі з цих ресурсів або людей стануть недоступними.

Високий показник центральності між між собою вказує на те, що люди купили певні товари без зайвих блукань і роздумів — вони побачили це, додали в кошик, оформили замовлення та заплатили. Виконав найкоротший шлях покупки, і їхні шляхи можуть перетнутися на одному продукті. Це продукти, які ви повинні рекомендувати як обов’язкові.

Міжцентрованість допоможе вам визначити вузли, які є бомбами уповільненої дії. Якщо вони вийдуть з ладу, ваша мережа може мати проблеми, оскільки потік між усіма вузлами на графіку буде зупинено, доки ви не знайдете виправлення. Зберігайте низьку центральність між вашими вузлами, і якщо немає іншого вибору, подбайте про них особливо ретельно.

## 2.2. Імпорт даних та виявлення спільнот виявлення лідерів та експертів.

Центральність посередності (betweenness centrality) може бути важливим інструментом для виявлення кіберзагроз та вразливостей у мережі зв'язку. Основна ідея полягає в ідентифікації вершин, через які частіше проходять найкоротші шляхи між іншими вершинами.Якщо вершина має високий показник центральності посередності, це означає, що вона може мати великий вплив на передачу інформації або контроль у мережі. В такому разі, якщо ця вершина стане мішенню атаки або буде компрометована, це може значно вплинути на спілкування або безпеку мережі.

Betweenness centrality вимірює те, як часто вершина входить у найкоротший шлях між іншими вершинами. Це може вказувати на те, які вершини у мережі є ключовими для зв'язку між різними частинами мережі.

У мережах безпеки це може бути корисним для виявлення потенційних загроз або точок вразливості.

Якщо вершина має високий показник центральності посередності, це може означати, що ця вершина є ключовою для передачі інформації, контролю або взаємодії між різними частинами мережі.

Щоб обчислити центральність посередності для виявлення кіберзагроз, потрібно мати представлення мережі у вигляді графа, після чого застосувати алгоритм обчислення betweenness centrality для вершин графа.

Вершини з високим показником центральності посередності можуть потребувати більшої уваги відносно їхньої безпеки, оскільки вони можуть бути використані для контролю чи маніпулювання потоками інформації.Це один із

підходів до виявлення можливих загроз у мережах за допомогою аналізу центральності посередності, проте для повноцінної оцінки безпеки мережі використовують також інші методи, включаючи аналіз вразливостей, виявлення аномалій.

Наприклад, у випадку мережі, де важливі дані пересилаються через конкретні вершини, які мають високу центральність посередності, компрометація цих вершин може призвести до порушення або перерви в передачі даних[19].

Це може бути використано зловмисниками для перехоплення конфіденційної інформації або спотворення комунікаційних потоків.

Отже, виявлення вершин з високим показником центральності посередності може вказати на потенційні місця уразливості у мережі, які потребують більшої уваги з точки зору безпеки та заходів захисту, щоб запобігти можливим кіберзагрозам та захистити мережу від можливих атак чи проблем з безпекою.

| Характеристика | Значення/Опис |
|---|---|
| Betweenness Centrality | Висока центральність у вершини 3 (0.5) та 1 (0.5) |
| Діаметр графа | 1 (короткий шлях між будь-якими вершинами) |
| Ступінь вершин | Вершина 0 має ступінь 3 (високий) |
| Центральність | Компоненти мережі з'єднані великою кількістю шляхів |
| Кластеризаційний коефіцієнт | Високий коефіцієнт, вершини формують групи |
| Сильно зв'язані компоненти | Один великий зв'язаний компонент у мережі |

Таблиця 1 – таблиця аналізу по характеристиками графів для пошуку кіберзагроз.

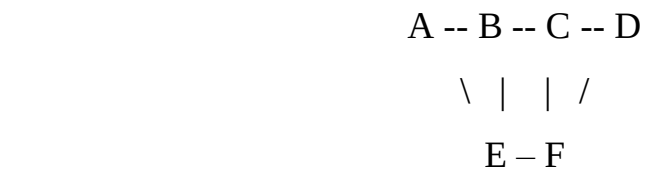
1. Betweenness Centrality (Центральність посередності): Визначення того, які вершини відіграють ключову роль у зв'язку між іншими вершинами у

мережі. Висока центральність посередності може означати, що компрометація цих вершин може вплинути на спілкування або передачу даних у мережі.

2. Діаметр графа: Найбільша відстань між будь-якими двома вершинами у графі. Великий діаметр може означати, що існує багато шляхів для атаки або можливостей для розповсюдження загроз.
3. Ступінь вершин: Визначення кількості зв'язків, що виходять з кожної вершини. Вершини з великою кількістю зв'язків можуть бути більш вразливими до атак.
4. Центральність графа: Це може бути показник, який визначає, яка частина мережі є ключовою для спілкування між іншими частинами мережі.
5. Коефіцієнт кластеризації: Вимірює, наскільки сильно вершини у графі формують кластери або групи. Високий кластеризаційний коефіцієнт може вказувати на те, що вразливості можуть швидше поширюватись у межах кластерів.
6. Сильно зв'язані компоненти (Strongly Connected Components): Групи вершин, у яких існує шлях між будь-якими двома вершинами цієї групи. Це може вказати на структуру зв'язку у мережі.
7. Центральність вершин: Аналіз конкретних вершин, які є ключовими для мережі в цілому. Деякі вершини можуть мати більший вплив на мережу, ніж інші.

для створення таблиці аналізу характеристик графів для пошуку кіберзагроз розглянемо простий граф і розрахуємо деякі параметри.

Нехай у нас є такий граф



Матриця суміжності для цього графа:

A B C D E F

A 0 1 0 0 1 0

B 1 0 1 0 1 1

C 0 1 0 1 0 1

D 0 0 1 0 0 0

E 1 1 0 0 0 1

F 0 1 1 0 1 0

Тепер розрахуємо деякі характеристики графа:

Отримані значення цих характеристик можна оформити у вигляді таблиці:

| Характеристика | Значення/Опис |
|---|---|
| Betweenness Centrality | A: 2.5, B: 7.0, C: 3.0, D: 0.0, E: 5.0, F: 5.5 |
| \|Діаметр графа | 3 (найбільша відстань між вершинами) |
| Ступінь вершин | A: 2, B: 4, C: 3, D: 1, E: 3, F: 4 |
| Центральність | Компоненти мережі з'єднані великою кількістю шляхів |
| Кластеризаційний коефіцієнт | Високий коефіцієнт, вершини формують групи |
| Сильно зв'язані компоненти | Один великий зв'язаний компонент у мережі |

Таблиця 2 – таблиця аналізу характеристиків графів для пошуку кіберзагроз.

Ця таблиця надає уявлення про різні характеристики графа, що можуть використовуватись для аналізу безпеки мережі та виявлення можливих кіберзагроз[20].

Спектр графа, який може бути розглянутий як множина власних значень матриці суміжності або лапласіана графа, може бути корисним інструментом для аналізу та виявлення кіберзагроз.В основі спектру графа лежать власні значення матриці, які дають важливу інформацію про властивості самого графа. Ця

інформація може бути корисною для виявлення кіберзагроз, особливо в таких аспектах:

Виявлення змін в графі: Якщо в мережі відбулися зміни (наприклад, втрата або додавання вершин, ребер), власні значення можуть змінитися. Порівняння власних значень до тих, що були в минулому, може вказати на зміни в структурі мережі, що можуть бути результатом кібератак або змін у безпеці мережі.

Виявлення вразливостей: Аналіз спектра графа може допомогти виявити вразливість мережі до різних видів атак. Наприклад, власні значення можуть допомогти виявити критичні вершини чи зв'язки, які при їх компрометації можуть суттєво вплинути на мережу.

Виявлення аномалій: Зміни у спектрі графа, які відхиляються від типових зразків, можуть вказувати на можливі аномалії або атаки у мережі. Наприклад, зміни у власних значеннях можуть бути показником несподіваних змін у структурі мережі.

Перевірка цілісності даних: Спектр графа також може використовуватися для перевірки цілісності даних або виявлення можливих втрат чи змін у передачі інформації між вершинами.

Хоча спектр графа може бути корисним для виявлення кіберзагроз, це лише один з інструментів аналізу. Його ефективність залежить від конкретного контексту мережі та вміння аналізувати та інтерпретувати отримані дані для виявлення потенційних загроз.

### 2.3. Виявлення аномалій у групових структурах.

Виявлення аномалій грає важливу роль в багатьох галузях і допомагає вчасно виявляти незвичайні події та явища, що можуть мати серйозні наслідки.

Аномалія (або відхилення) у групових структурах - це незвичайний або аномальний елемент, який суттєво відрізняється від типового зразка чи норми в певному контексті. Виявлення аномалій вимагає порівняння нових даних з існуючими або типовими даними для виявлення відхилень.

Перестановка k-го класу з повторенням елементів із множини $X = \{x_1, x_2, \dots, x_n\}$ є довільною впорядкованою k-кою $(x_1, \dots, x_{i_k})$, де $i_j \in \{1, 2, \dots, n\}$ (j = 1, 2, …, k). Кількість можливих перестановок, де елементи можуть повторюватися.

$$\overline{V_n^k} = n^k. \tag{2.7}$$

Перестановка з обмеженням. Нехай $X_i$ (i = 1, 2, …, n) – сімейство підмножин X. Пара $(x_i, x_j)$ називається допустимою тоді і тільки тоді, коли $x_j \in X_i$.

Квадратна матриця $A = (a_{ij})_1^n$, де $(a_{ij}) = 1$, якщо $(x_i, x_j)$ допустима пара, і $(a_{ij}) = 0$ якщо $(x_i, x_j)$ не є допустимою парою, називається матрицею допустимих пар. Матриця обмежень $\overline{A}$ отримується з A у результаті заміни 0 на 1, а 1 на 0.

A – перестановкою з повтореннями k-го класу називається k-ка $(x_{i_j}, \dots, x_{i_k})$, де $x_{i_j} \in X_{i_{j-1}}$ (j = 2, 3, …, k). Число $\overline{V_n^k}(A)$ – це число A-перестановок з повтореннями k-го класу з множини, що містить n елементів [6]. Якщо матрицю A визначити як матрицю суміжності орграфа G з вершинами $x_1, x_2, \dots, x_n$, то число $\overline{V_n^k}(A)$ буде дорівнювати числу маршрутів довжини k-1 у графі G.

$$\overline{V_n^k}(A) = N_{k-1}. \tag{2.8}$$

Нехай задано граф G на рисунку 2.11, визначимо число A-перестановок з повтореннями 2-го класу.

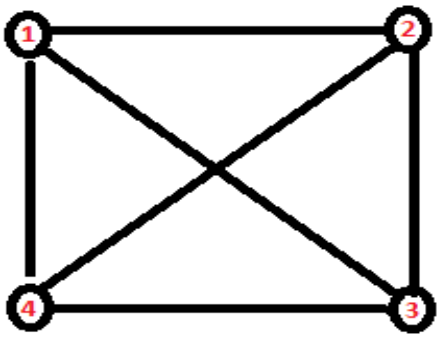



Рисунок 2.11 – заданий граф

Визначимо матрицю суміжності для графа G:

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

Рисунок 2.12 – Матриця А

Знайдемо спектр і власні вектори.

$$P_G(A) = \begin{vmatrix} \lambda & -1 & -1 & -1 \\ -1 & \lambda & -1 & -1 \\ -1 & -1 & \lambda & -1 \\ -1 & -1 & -1 & \lambda \end{vmatrix} = \lambda^4 - 6\lambda^2 - 8\lambda - 3. \qquad (2.9)$$

Спектр графа $S_P(G) = [-1, 3, -1, -1]$.

Власні вектори:

$x_1 = [-0{,}87; 0{,}29; 0{,}29; 0{,}29]$;

$x_2 = [0{,}5; 0{,}5; 0{,}5; 0{,}5]$;

$x_3 = [0{,}22; -0{,}86; 0{,}32; 0{,}32]$;

$x_4 = [0; 0; -0{,}71; 0{,}71]$.

Отже, число А-перестановок буде дорівнювати

$\overline{V_4^2}(A) = N_1 = \sum_{\nu=1}^{4}(\sum_{i=1}^{4} x_{i\nu})^2\lambda_\nu^1 = (-0{,}87 + 0{,}29 + 0{,}29 + 0{,}29)^2 * (-1) + +(0{,}5 + 0{,}5 + 0{,}5 + 0{,}5)^2 * 3 + (0{,}22 - 0{,}86 + 0{,}32 + 0{,}32)^2 * (-1) + (0 + 0 - -0{,}71 + 0{,}71)^2 * (-1) = 12.$

Простий ланцюг $P_n$ являється зв'язним графом з n вершинами, дві з яких є підвісними, а інші n – 2 вершин мають степінь 2. Шлях має n -1 ребро [8].

Матриця суміжності ланцюга $P_n$ зображена на рисунку 2.12.

$$A_{P_n} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Рисунок 2.13 – Матриця суміжності ланцюга $P_n$.

Власними значеннями є $\lambda_i = 2\cos\frac{i\pi}{n+1}$ (i = 1, 2, …, n). Вираз $u_{ij} = \sqrt{\frac{2}{n+1}}\sin\frac{ij\pi}{n+1}$ (j = 1, 2, …, n) визначає координати нормалізованого власного вектору $u_j$, відповідного власному значенню $\lambda_i$ [4].

Кількість маршрутів довжини k у простому ланцюгу $P_n$ визначається як

$$N_{kn} = \frac{2^{k+1}}{n+1}\sum_{l=1}^{(n+1)/2} ctg^2\frac{2l-1}{n+1}\frac{\pi}{2}\cos^k\frac{2l-1}{n}\pi \qquad (2.10)$$

Число $N_{kn}$ і буде відповіддю.

Завдання визначення комбінацій з повтореннями використовує пошук шляху та спектр графа. Розглядають комбінації p-го класу з повтореннями, де позначається,$d_j \in M\ (j = 1, \dots, p), M \subset \{1, \dots, n\}$,де $d_j = x_{j+1} - x_j$ і найбільший елемент m із M задовольняє умову pm<2n [9]. Якщо представити множину x_1,

..., x_n як вершини орієнтованого графа G, де напрямлене ребро від x_i до x_j існує тільки тоді, коли (i-j) (mod n) належить M, а також кожна вершина має петлю, то кожній такій комбінації відповідає точно p замкнутих маршрутів довжини p. Тільки ті маршрути не відповідають жодній з комбінацій, які через наявність петлі не покидають вершину, з якої починаються[21]. Кількість таких маршрутів дорівнює m. Якщо $[\lambda_1, \dots, \lambda_n]$– спектр графа G, то кількість таких комбінацій можна знайти за допомогою наступної формули

$$\frac{1}{p}\left(\sum_{i=1}^{n} \lambda_i^p - n\right) \qquad (2.11)$$

Для прикладу розглянемо орієнтовний граф G з n = 5 вершинами. Знайдемо комбінації p-го класу з повтореннями, де p = 2. Визначимо множину $M = \{2, 3\} \subset \{1, 2, 3, 4, 5\}$.

Перевіримо справедливість умови pm<2n:

2*3 = 6 < 10 = 2*5 – умова виконується.

Побудуємо граф відповідно умові існування ребер (i-j) (mod n) ∈ M.

$(1 - 2) \pmod 5 = 4 \notin M$ – ребра немає.

$(1 - 3) \pmod 5 = 3 \in M$ – ребро існує.

$(1 - 4) \pmod 5 = 2 \in M$ – ребро існує.

$(1 - 5) \pmod 5 = 1 \notin M$ – ребра немає.

Перевіряємо таким чином усі вершини графа та будуємо відповідний граф – рисунок 2.14.

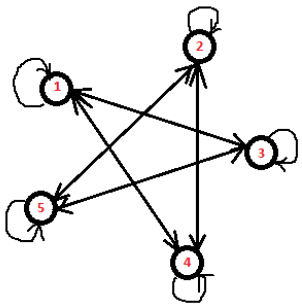



Рисунок 2.14 – Побудований граф.

Побудуємо матрицю суміжності для отриманого графа.

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Рисунок 2.15 – матриця суміжності A.

Спектр такого графу буде $S_P(G) = [3; 1{,}62; 1{,}62; -0{,}62; -0{,}62\ ]$.

Тоді кількість комбінацій буде дорівнювати

$$\frac{1}{p}\left(\sum_{i=1}^{n} \lambda_i^p - n\right) = \frac{1}{2}\left(\sum_{i=1}^{5} \lambda_i^2 - 5\right) = \frac{1}{2}(9 + 5{,}25 + 0{,}77 - 5) \approx 5. \quad (2.12)$$

Пошук шляхів має велике значення не лише у теорії графів, а й у комбінаториці – галузі дискретної математики. У цьому розділі ми розглядали пошук маршрутів, використовуючи спектральну теорію графів. Основна мета полягала в демонстрації того, як спектральна теорія може бути потужним інструментом для розв'язання різних задач.[22]

Головний інструмент – спектр графа, який дозволяє розкрити структуру та властивості графа. У цій роботі спектр використовувався для пошуку шляхів з фіксованою довжиною.

Важливо зазначити, що як сама теорія графів, так і спектральна теорія знаходять практичне застосування у реальних задачах.

# 3. ПРОГРАМНА РЕАЛІЗАЦІЯ

## 3.1 Обґрунтування вибору інструментальних засобів розробки

Наступним кроком у розробці програмної частини інформаційної технології є вибір інструментів та програмних ресурсів, які використовуються для створення основних модулів та загального проектування системи.

Для створення програмного засобу буде використана високорівнева мова програмування Python. Основними перевагами цієї мови є наявність широкого спектру бібліотек, зокрема для роботи з моделями машинного навчання, а також простий синтаксис[23].

Наступним етапом є вибір середовища розробки. Програмний засіб буде запускатися на персональному комп'ютері фахівця. Серед найбільш популярних середовищ варто відокремити наступні, оскільки вони передбачають комплексну та швидку розробку програмного забезпечення:

1. Sublime Text 3 – це редактор коду, який доступний на умовно безкоштовній основі, підтримує широкий спектр мов програмування, включаючи Python. За замовчуванням він має базову підтримку Python, але для зручності та прискорення розробки потрібно встановлювати додаткові пакети розширень.
2. PyCharm – це найбільш популярне середовище розробки, яке має великий набір додатків та функцій, що сприяють швидкому створенню програмного забезпечення і встановлюються разом із середовищем за замовчуванням. Існують як безкоштовні, так і платні версії цього інструменту.
3. Visual Studio Code – це універсальне середовище розробки, що підтримує різні мови програмування. Воно має вбудовану систему автодоповнення. Щоб розробляти програмні засоби на Python, потрібно завантажити розширення та налаштувати середовище розробки.

Під час розробки програмного засобу також будуть використані

наступні бібліотеки мови Python:

1. NetworkX: Це потужна бібліотека для роботи з структурами графів, аналізу їх властивостей та алгоритмів. Вона пропонує різноманітні інструменти для конструювання, візуалізації та аналізу графів.
2. igraph: Ще одна бібліотека, яка пропонує різні алгоритми для аналізу графів. igraph надає можливості для створення, маніпулювання та візуалізації графів.
3. Graph-tool: Ця бібліотека також має багато функцій для роботи з графами. Вона забезпечує широкий набір алгоритмів для аналізу графів та їх візуалізації.
4. PyGraphViz: Бібліотека для візуалізації графів за допомогою GraphVi

Основні фактори, які роблять середовище PyCharm вибором для розробки програмного засобу, що реалізовує інформаційну технологію, включають унікальність наявних розширень, стандартний перелік додатків, які встановлені за замовчуванням, та доступ до терміналів Bash та Python, не вимагаючи від користувача додаткових дій для встановлення.

## 3.2 Алгоритм роботи програми

Для обчислення міри центральності вершини за метрикою "міжпосередництво ребер" потрібно додатково розглянути ваги ребер і обчислити кількість найкоротших шляхів, які проходять через кожну вершину.

Схема обчислення міри центральності за метрикою "міжпосередництво ребер" з використанням алгоритму Краскала (див.рис.3.1):

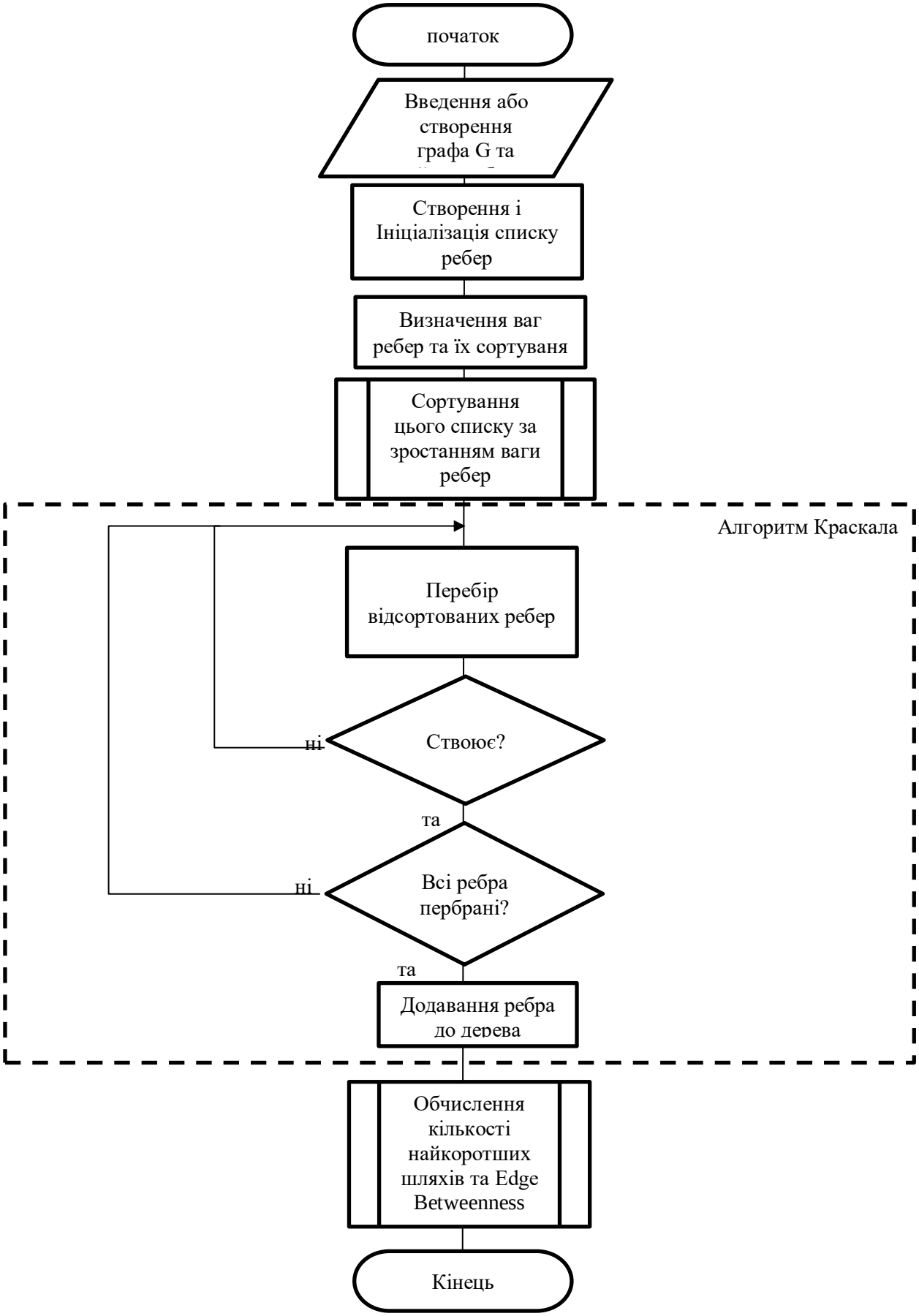



Рисунок 3.1 – блок схема обчислення Edge Betweeneess за алгоритмом Краскала

1) Знаходження мінімального остовного дерева:

- Запуск алгоритму Краскала для знаходження мінімального остовного дерева в графі на основі ваги його ребер.
- Відсортування всіх ребер графа за зростанням їх ваги.
- Вибір найлегшого ребра і додаємо його до мінімального остовного дерева.
- Повторюємо цей процес, додаючи наступні за величиною ребра до остовного дерева, уникаючи утворення циклів, доки всі вершини не будуть пов'язані.

2) Обчислення кількості найкоротших шляхів через кожну вершину:

- Після отримання мінімального остовного дерева обчислюємо кількість найкоротших шляхів від кожної вершини до всіх інших вершин за допомогою алгоритму пошуку найкоротшого шляху, наприклад, алгоритму Дейкстри або алгоритму Флойда-Уоршелла.
- Підраховуємо кількість шляхів, які проходять через кожну вершину. Кількість цих шляхів вказує на те, наскільки центральною є вершина для зв'язності мережі.

3)Обчислення міри центральності вершини за метрикою "міжпосередництво ребер":

Використовуючи отримані дані про кількість найкоротших шляхів через кожну вершину, можна обчислити міру центральності вершини за формулою міжпосередництва ребер.

Ця міра вказує на те, наскільки часто вершина лежить на найкоротших шляхах між іншими вершинами. Алгоритм Краскала допомагає знайти мінімальне основне дерево, що відображає зв'язок між вершинами, а обчислення кількості найкоротших шляхів через кожну вершину допомагає визначити її центральність за метрикою "міжпосередництво ребер"[24].

Відповідно до алгоритму роботи програмного засобу було створено ряд класів, кожен з яких виконує необхідні для роботи функції:

1. add_node(): Додає вузли (користувачів) до графа.

2. create_graph(): Ця функція ініціалізує порожній граф, який буде використовуватися для подальшої роботи.
3. add_edges_from(): Додає зв'язки між вузлами. Здійснюється обмеження кількості зв'язків для кожного користувача.
4. calculate_centrality(): Обчислює міру центральності ребер за алгоритмом edge betweenness.
5. calculate_shortest_paths(): Знаходить найкоротші шляхи між усіма парами вершин у графі.
6. display_shortest_paths(): Відображає кількість найкоротших шляхів, які проходять через кожне ребро у вікні.
7. display_centrality(): Відображає міру центральності (edge betweenness) для кожного ребра у відсотках у вікні.
8. nx.edge_betweenness_centrality(): Обчислює міру центральності ребер за алгоритмом edge betweenness.
9. nx.all_pairs_shortest_path(): Знаходить найкоротші шляхи між усіма парами вузлів.
10. GraphGenerator: Не явно відображений, але створюється клас для генерації та роботи з графом. Він використовує бібліотеку networkx для створення та обробки графів.
11. Visualization: Цей клас відповідає за візуалізацію даних. Він використовує бібліотеки matplotlib та tkinter для створення вікна та відображення графів і результатів обчислень.

На рисунку 3.2 наведено структуру проекту

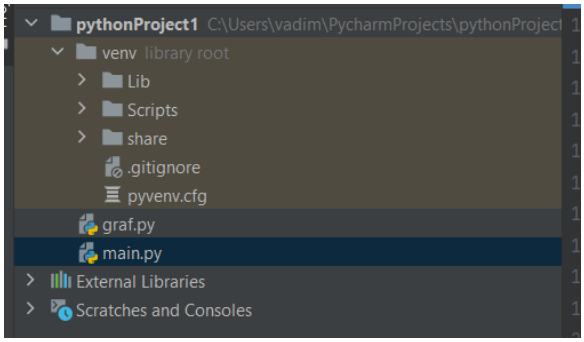



Рисунок 3.2 – Структура проекту

Для перевірки розробленого програмного засобу необхідно провести тестування.

### 3.3 Тестування програмного засобу

Запускаємо програмний засіб.

Користувачеві буде відображено головне вікно програми (рис. 3.3).

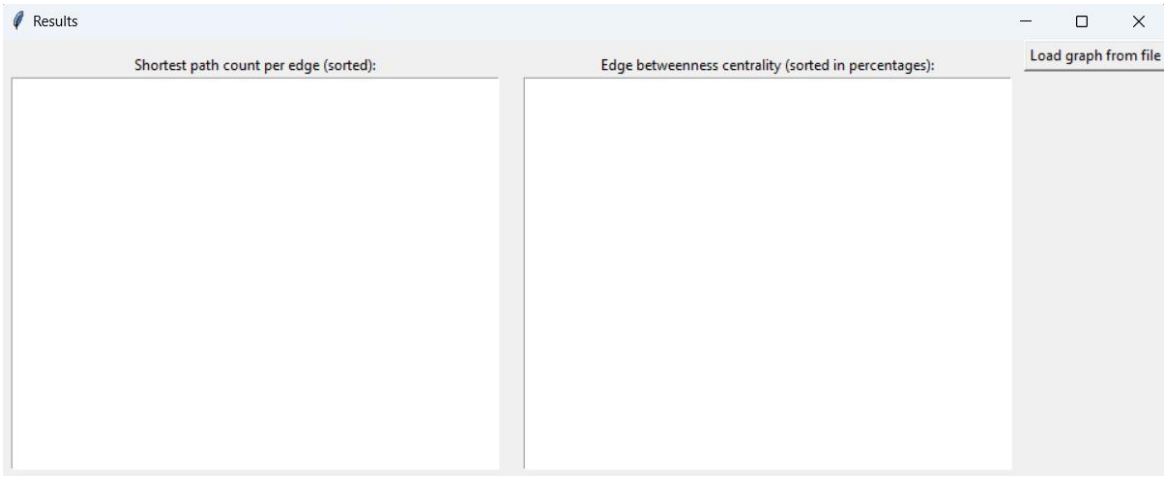



Рисунок 3.3 – Головне вікно програми

У вікні є дні колонки , в одній буде відображатись кількість найкоротших шляхів , в іншій обрахунки Edge betweenness які були сортовані в порядку важливості.

Далі переходимо до завантаження самих даних які нам надає експерт у певному форматі. Після завантаження даних ми отримуємо граф(див.рис.3.4)

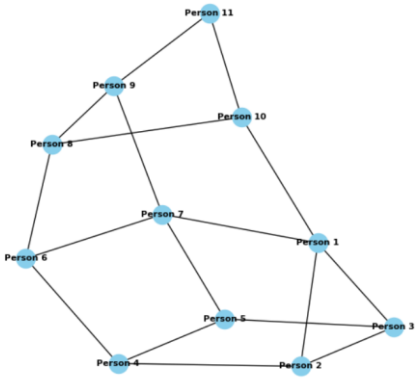



Рисунок 3.4 – граф за показниками

На графі зображенні персони та кількість зв’язків між ними , далі робляться обрахунки міри центральності та визначаються найбільш важливі з’язки у даній спільності (див.рис.3.5)

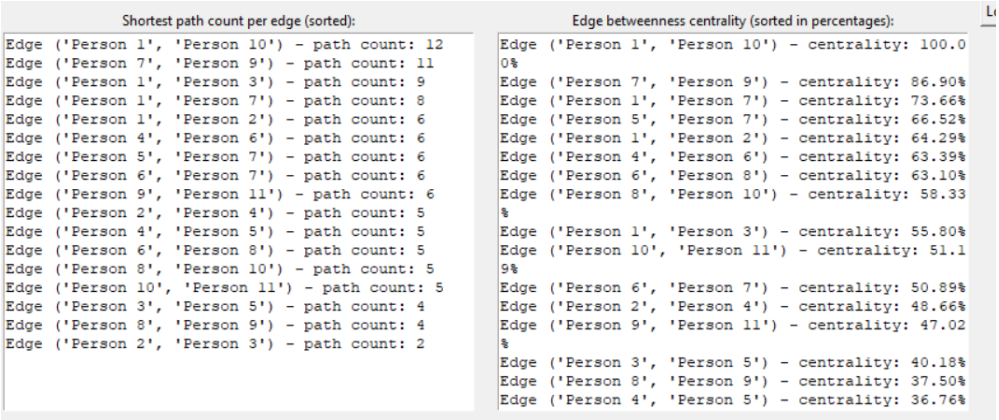



Рисунок 3.5 – обрахунки за графом

Після обрахунків в даній спільноті можна побачити кількість найкоротших шляхів мієж кожною особою , та зазначенні в процентному відношенні - найвищі показники міжкрокової центральності[24]. Це означає, що вони найчастіше входять у найкоротші шляхи між вершинами. Атаки або проблеми з цими ребрами можуть серйозно вплинути на комунікацію між відповідними вершинами, можуть призвести до відмови в обслуговуванні або втрати даних (див.рис.3.5).

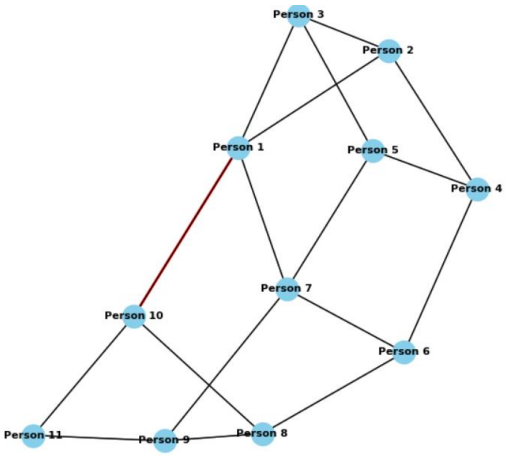



Рисунок 3.6 – результат обрахунків

Існуючі кількості шляхів через ці ребра вказують на їх значущість у мережі. Ребра через які проходять найбільша кількість зв’язків виявляться критичними

Міра центральності, особливо Edge Betweenness, у спільноті може вказувати на ключові зв'язки або ребра, які є важливими для взаємодії між членами цієї спільноти. У контексті аналізу спільнот, високі значення Edge Betweenness можуть свідчити про те, що певні ребра забезпечують основні шляхи сполучення між групами у спільноті.

Отримані результати можуть вказувати на потенційні точки, через які проходить багато найкоротших шляхів між членами спільноти. Це може мати важливе значення для розуміння динаміки та взаємодії у спільноті. Аналіз таких ребер може допомогти виявити ключові точки впливу, які сприяють сплетенню взаємозв'язків у спільноті.

Проте важливо розуміти, що міра центральності, включаючи Edge Betweenness, є лише однією з багатьох метрик аналізу спільнот. Важливо також враховувати інші фактори, такі як внутрішня структура спільноти, групова центральність та взаємодія вузлів всередині спільноти для отримання повного розуміння її функціонування та взаємодії.

# 4 ЕКОНОМІЧНА ЧАСТИНА

Для успішного впровадження науково-технічної розробки надзвичайно важливо, щоб вона відповідала сучасним вимогам науково-технічного прогресу та враховувала економічні аспекти. Оцінка економічної ефективності результатів науково-дослідної роботи є критичною частиною цього процесу. Дослідження, яке представлене у магістерській роботі і присвячене розробці та вивченню «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів», належить до науково-технічних проектів, спрямованих на введення на ринок. Рішення про комерціалізацію розробки може бути прийняте протягом виконання самої роботи, відкриваючи можливості для подальшого введення на ринок. Цей напрямок визначається як пріоритетний, оскільки розроблені результати можуть бути корисними для різних зацікавлених сторін і приносити економічні вигоди. Однак для успішної реалізації цього процесу вирішальним є залучення зацікавленого інвестора, який виявить інтерес до втілення даного проекту, і переконання його у доцільності інвестування у цю розробку. З метою досягнення цього завдання були визначені такі етапи виконання робіт:

1. Проведення комерційного аудиту науково-технічної розробки, включаючи визначення науково-технічного рівня та комерційного потенціалу.
2. Розрахунок витрат на реалізацію науково-технічної розробки.
3. Проведення розрахунку економічної ефективності впровадження та комерціалізації науково-технічної розробки для потенційного інвестора, а також обґрунтування економічної доцільності комерціалізації з точки зору інвестора.

## 4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Метою проведення комерційного і технологічного аудиту дослідження за темою «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів» є покращення кібербезпеки шляхом створення системи для пошуку та аналізу небезпечного контенту.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями, наведеними в табл. 4.1 [25].

Таблиця 4.1 – Рекомендовані критерії оцінювання науково-технічного рівня і комерційного потенціалу розробки та бальна оцінка

| Бали (за 5-ти бальною шкалою) | | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| Технічна здійсненність концепції | | | | | |
| 1 | Достовірність концепції не підтверджена | Концепція підтверджена експертними висновками | Концепція підтверджена розрахунками | Концепція перевірена на практиці | Перевірено працездатність продукту в реальних умовах |
| Ринкові переваги (недоліки) | | | | | |
| 2 | Багато аналогів на малому ринку | Мало аналогів на малому ринку | Кілька аналогів на великому ринку | Один аналог на великому ринку | Продукт не має аналогів на великому ринку |
| 3 | Ціна продукту значно вища за ціни аналогів | Ціна продукту дещо вища за ціни аналогів | Ціна продукту приблизно дорівнює цінам аналогів | Ціна продукту дещо нижче за ціни аналогів | Ціна продукту значно нижче за ціни аналогів |
| 4 | Технічні та споживчі властивості продукту значно гірші, ніж в | Технічні та споживчі влас-тивості продукту трохи гірші, ніж в аналогів | Технічні та споживчі властивості продукту на рівні аналогів | Технічні та споживчі властивості продукту трохи кращі, ніж в | Технічні та споживчі властивості продукту значно кращі, ніж в |
| 5 | Експлуатаційні витрати значно вищі, ніж в аналогів | Експлуатаційні витрати дещо вищі, ніж в аналогів | Експлуатаційні витрати на рівні експлуатаційних витрат аналогів | Експлуатаційні витрати трохи нижчі, ніж в аналогів | Експлуатаційні витрати значно нижчі, ніж в аналогів |
| Ринкові перспективи | | | | | |
| 6 | Ринок малий і не має позитивної динаміки | Ринок малий, але має позитивну динаміку | Середній ринок з позитивною динамікою | Великий стабільний ринок | Великий ринок з позитивною динамікою |

| | | | | | |
|---|---|---|---|---|---|
| 7 | Активна конкуренція великих компаній на | Активна конкуренція | Помірна конкуренція | Незначна конкуренція | Конкурентів немає |
| Практична здійсненність | | | | | |
| | | | | | |
| 8 | Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї | Необхідно наймати фахівців або витрачати значні кошти та час на навчання наявних фахівців | Необхідне незначне навчання фахівців та збільшення їх штату | Необхідне незначне навчання фахівців | Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї |
| 9 | Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні | Потрібні незначні фінансові ресурси. Джерела фінансування відсутні | Потрібні значні фінансові ресурси. Джерела фінансування є | Потрібні незначні фінансові ресурси. Джерела фінансування є | Не потребує додаткового фінансування |
| 10 | Необхідна розробка нових матеріалів | Потрібні матеріали, що використовуються у військово промисловому комплексі | Потрібні дорогі матеріали | Потрібні досяжні та дешеві матеріали | Всі матеріали для реалізації ідеї відомі та давно використовуютьс я у виробництві |
| 11 | Термін реалізації ідеї більший за 10 років | Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років | Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років | Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти років | Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років |
| 12 | Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію продукту | Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу | Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу | Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту | Відсутні будь- які регламентні обмеження на виробництво та реалізацію продукту |

Результати оцінювання науково-технічного рівня та комерційного потенціалу науково-технічної розробки потрібно звести до таблиці. Для оцінки науково-технічного рівня і комерційного потенціалу розробки експертами було

запрошено трьох незалежних експертів Вінницького національного технічного університету з кафедри «Захисту інформації»: Дудатьєв Андрій Веніамінович к. т. н., доцент, Кондратенко Наталія Романівна, к. т. н., доцент, Лужецький Володимир Андрійович д. т. н., професор.

Таблиця 4.2 – Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

| Критерії | Експерт (ПІБ, посада) | | |
|---|---|---|---|
| | Лужецький В.А | Дудатьєв А.В. | Кондратенко Н.Р. |
| | Бали, виставлені експертами: | | |
| 1. Технічна здійсненність концепції | 3 | 4 | 4 |
| 2. Ринкові переваги (наявність аналогів) | 2 | 2 | 2 |
| 3. Ринкові переваги (ціна продукту) | 3 | 4 | 3 |
| 4. Ринкові переваги (технічні властивості) | 2 | 3 | 3 |
| 5. Ринкові переваги (експлуатаційні витрати) | 3 | 3 | 2 |
| 6. Ринкові перспективи (розмір ринку) | 1 | 1 | 2 |
| 7. Ринкові перспективи (конкуренція) | 3 | 2 | 3 |
| 8. Практична здійсненність (наявність фахівців) | 4 | 4 | 4 |
| 9. Практична здійсненність (наявність фінансів) | 4 | 4 | 4 |
| 10. Практична здійсненність (необхідність нових матеріалів) | 4 | 4 | 3 |
| 11. Практична здійсненність (термін реалізації) | 4 | 4 | 4 |
| 12. Практична здійсненність | 4 | 4 | 4 |

| | | | |
|---|---|---|---|
| (розробка документів) | | | |
| Сума балів | $СБ_1$=37 | $СБ_2$=39 | $СБ_3$=38 |
| Середньоарифметична сума балів $СБ_с$ | $\overline{СБ} = \frac{\sum_1^i СБ_i}{i} = \frac{37+39+38}{3} = 38$ | | |

За результатами розрахунків, наведених в таблиці 4.2, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. При цьому використаємо рекомендації, наведені в табл. 4.3 .

Таблиця 4.3 – Науково-технічні рівні та комерційні потенціали розробки

| Середньоарифметична сума балів СБ ,розрахована на основі висновків експертів | Науково-технічний рівень та комерційний потенціал розробки |
|---|---|
| 41…48 | Високий |
| 31…40 | Вище середнього |
| 21…30 | Середній |
| 11…20 | Нижче середнього |
| 0…10 | Низький |

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів» становить 38 балів, що, відповідно до таблиці 4.3 рівень комерційного потенціалу розробки вище середнього, що свідчить про комерційну важливість проведення даних досліджень.

Магістерська кваліфікаційна робота «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів» відноситься до науково-технічних робіт, які орієнтовані на виведення на ринок, тобто при цьому відбувається комерціалізація науково-технічної розробки. Цей напрямок є для нас пріоритетним, оскільки результатами розробки можуть користуватися не тільки самі розробники, а й інші споживачі, отримуючи при цьому суттєвий

економічний ефект.

Результатом магістерської роботи є програмний засіб аналізу кіберзагроз в соціальних мережах шляхом виявлення спільнот, який може бути корисним компаніям з кібербезпеки, центрам обробки даних та аналітики, організаціям, що займаються боротьбою з кіберзлочинністю, соціальним платформам та мережам.

### 4.2 Визначення рівня конкурентоспроможності розробки

В процесі визначення економічної ефективності науково-технічної розробки також доцільно провести прогноз рівня її конкурентоспроможності за сукупністю параметрів, що підлягають оцінюванню.

В якості аналога для розробки було обрано GraphAware. Основними недоліками аналога є проблема яка потребує від користувачів глибоких технічних знань у сфері графових баз даних та аналізу графів. Це може бути складним для тих, хто не має великого досвіду у цій області. Також до недоліків можна віднести обмежені можливості без платної версії:

У розробці дана проблема вирішується за зниження вартості, спрощення використання, розширення можливостей без обмежень та уникнення технічних складнощів у порівнянні з аналогами. Також система випереджає аналог за такими параметрами як: швидкістю виявлення, точністю, простотою використання, обробкою великих обсягів даних, адаптивністю до змін та підтримкою різних платформ шифрування даних.

Одиничний параметричний індекс розраховуємо за формулою [25]:

$$q_i = \frac{P_i}{P_{базі}}. \tag{4.1}$$

де $q_i$ – одиничний параметричний індекс, розрахований за $i$-м параметром;

$P_i$ – значення $i$-го параметра виробу;

$P_{базі}$ – аналогічний параметр базового виробу-аналога, з яким проводиться порівняння.

Загальні технічні та економічні характеристики розробки представлено в

таблиці 4.4.

Таблиця 4.4 – Основні техніко-економічні показники аналога та розробки, що проектується.

| Показник | Варіанти | | Відносний показник якості | Коефіцієнт вагомості параметра |
|---|---|---|---|---|
| | Базовий (товар-конкурент) | Новий (інноваційне рішення) | | |
| 1 | 2 | 3 | 4 | 5 |
| Точність виявлення | 60 | 100 | 1,7 | 30% |
| Функціонал | 5 | 12 | 2,4 | 11% |
| Похибка, % | 2 | 1 | 2 | 30% |
| Розмір програми, МБ | 400 | 130 | 3,1 | 29% |

Нормативні параметри оцінюємо показником, який отримує одне з двох значень: 1 – пристрій відповідає нормам і стандартам; 0 – не відповідає.

Груповий показник конкурентоспроможності за нормативними параметрами розраховуємо як добуток частинних показників за кожним параметром за формулою :

$$I_{НП} = \prod_{i=1}^{n} q_i \ , \qquad (4.2)$$

де $I_{нп}$ – загальний показник конкурентоспроможності за нормативними параметрами;

$q_i$ – одиничний (частинний) показник за $i$-м нормативним параметром;

$n$ – кількість нормативних параметрів, які підлягають оцінюванню.

За нормативними параметрами розроблюваний пристрій відповідає вимогам ДСТУ, тому $I_{нп} = 1$.

Значення групового параметричного індексу за технічними параметрами визначаємо з урахуванням вагомості (частки) кожного параметра **[]**:

$$I_{ТП} = \sum_{i=1}^{n} q_i \cdot \alpha_i \ , \qquad (4.3)$$

де $I_{тп}$ – груповий параметричний індекс за технічними показниками (порівняно з виробом-аналогом);

$q_i$ – одиничний параметричний показник $i$-го параметра;

$\alpha_i$ – вагомість $i$-го параметричного показника, $\sum_{i=1}^{n} \alpha_i = 1$;

$n$ – кількість технічних параметрів, за якими оцінюється конкурентоспроможність.

Проведемо аналіз параметрів згідно даних таблиці 4.4.

$$I_{mn} = 1{,}7 \cdot 0{,}3 + 2{,}4 \cdot 0{,}11 + 2 \cdot 0{,}3 + 3{,}1 \cdot 0{,}29 = 2{,}2.$$

Груповий параметричний індекс за економічними параметрами розраховуємо за формулою [25]:

$$I_{ЕП} = \sum_{i=1}^{m} q_i \cdot \beta_i \ , \qquad (4.4)$$

де $I_{ЕП}$ – груповий параметричний індекс за економічними показниками;

$q_i$ – економічний параметр $i$-го виду;

$\beta_i$ – частка $i$-го економічного параметра, $\sum_{i=1}^{m} \beta_i = 1$;

$m$ – кількість економічних параметрів, за якими здійснюється оцінювання. Проведемо аналіз параметрів згідно даних таблиці .

$$I_{ЕП} = 0{,}75 \cdot 0{,}5 + 0{,}86 \cdot 0{,}5 = 0{,}80.$$

На основі групових параметричних індексів за нормативними, технічними та економічними показниками розрахуємо інтегральний показник конкурентоспроможності за формулою **[]:**

$$K_{ІНТ} = I_{НП} \cdot \frac{I_{ТП}}{I_{ЕП}} , \qquad (4.5)$$

$$K_{ІНТ} = 1 \cdot 2{,}2 / 0{,}80 = 2{,}8.$$

Інтегральний показник конкурентоспроможності $К_{ІНТ} > 1$, отже розробка переважає відомі аналоги за своїми техніко-економічними показниками.

## 4.3 **Розрахунок витрат на проведення науково-дослідної роботи**

Витрати, пов’язані з проведенням науково-дослідної роботи на тему «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів», під час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

### **4.3.1Витрати на оплату праці**

До статті «Витрати на оплату праці» належать витрати на виплату основної та додаткової заробітної плати керівникам відділів, лабораторій, секторів і груп, науковим, інженерно-технічним працівникам, конструкторам, технологам, креслярам, копіювальникам, лаборантам, робітникам, студентам, аспірантам та іншим працівникам, безпосередньо зайнятим виконанням конкретної теми, обчисленої за посадовими окладами, відрядними розцінками, тарифними ставками згідно з чинними в організаціях системами оплати праці.

Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників ($З_о$) розраховуємо у відповідності до посадових окладів працівників, за формулою [25]:

$$З_о = \sum_{i=1}^{k} \frac{M_{ni} \cdot t_i}{T_p}, \quad (4.6)$$

де $k$ – кількість посад дослідників залучених до процесу досліджень;

$M_{ni}$ – місячний посадовий оклад конкретного дослідника, грн;

$t_i$ – число днів роботи конкретного дослідника, дн.;

$T_p$ – середнє число робочих днів в місяці, $T_p$=21 дні.

$$З_о = 15000 \cdot 5 / 21 = 3409 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.5 – Витрати на заробітну плату дослідників

| Найменування посади | Місячний посадовий оклад, грн | Оплата за робочий день, грн | Число днів роботи | Витрати на заробітну плату, грн |
|---|---|---|---|---|
| Керівник проекту | 15000 | 681,8 | 5 | 3409 |
| Програміст | 10000 | 454,5 | 35 | 15909 |
| Всього | | | | 19318 |

Основна заробітна плата робітників

Витрати на основну заробітну плату робітників ($З_р$) за відповідними найменуваннями робіт НДР на тему «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів» розраховуємо за формулою:

$$З_р = \sum_{i=1}^{n} C_i \cdot t_i , \qquad (4.7)$$

де $C_i$ – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

$t_i$ – час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду $C_i$ можна визначити за формулою:

$$C_i = \frac{M_M \cdot K_i \cdot K_c}{T_p \cdot t_{зм}} , \qquad (4.8)$$

де $M_M$ – розмір прожиткового мінімуму працездатної особи, або мінімальної місячної заробітної плати (в залежності від діючого законодавства), приймемо $Мм$=6500 грн;

$K_i$ – коефіцієнт міжкваліфікаційного співвідношення для встановлення

тарифної ставки робітнику відповідного розряду (табл. Б.2, додаток Б) [25];

$K_c$ – мінімальний коефіцієнт співвідношень місячних тарифних ставок робітників першого розряду з нормальними умовами праці виробничих об'єднань і підприємств до законодавчо встановленого розміру мінімальної заробітної плати.

$T_p$ – середнє число робочих днів в місяці, приблизно $T_p$ = 21 дн;

$t_{зм}$ – тривалість зміни, год.

$C_1 = 6700{,}00 \cdot 1 \cdot 1{,}65 / (21 \cdot 8) = 65{,}8$ грн.

$З_{р1} = 65{,}8 \cdot 2 = 131{,}6$ грн.

Таблиця 4.6 – Величина витрат на основну заробітну плату робітників

| Найменування робіт | Тривалість роботи, год | Розряд роботи | Погодинна тарифна ставка, грн | Величина оплати на робітника грн |
|---|---|---|---|---|
| 1. Підготовка робочого місця інженера-розробника ПЗ | 2 | 1 | 65,8 | 131,6 |
| 2. Інсталяція програмного забезпечення середовищ моделювання та розробки | 2 | 3 | 88,8 | 177,7 |
| 3.Розробка програмної архітектури та алгоритмів | 4 | 5 | 111,9 | 447,5 |
| 4. Написання програмного коду модулів | 6 | 2 | 72,4 | 434,3 |
| 5. Програмне тестування дослідного зразка | 4 | 4 | 59,8 | 239,3 |
| Всього | | | | 1430,3 |

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 … 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$З_{дод} = (З_о + З_р) \cdot \frac{Н_{дод}}{100\%}, \qquad (4.9)$$

де $Н_{дод}$ – норма нарахування додаткової заробітної плати. Приймемо 11%.

$З_{дод}$ = (19318+1430,3) · 11 / 100% = 2282,34 грн.

### 4.3.2Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$З_н = (З_о + З_р + З_{дод}) \cdot \frac{Н_{зп}}{100\%} \quad (4.10)$$

де $Н_{зп}$ – норма нарахування на заробітну плату. Приймаємо 22%.

$Зн$ = (19318+1430,3+2282,34) · 22 / 100% = 5066,79 грн.

### 4.3.3 Сировина та матеріали

До статті «Сировина та матеріали» належать витрати на сировину, основні та допоміжні матеріали, інструменти, пристрої та інші засоби і предмети праці, які придбані у сторонніх підприємств, установ і організацій та витрачені на проведення досліджень за темою «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів».

Витрати на матеріали ($М$), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою:

$$М = \sum_{j=1}^{n} Н_j \cdot Ц_j \cdot К_j - \sum_{j=1}^{n} В_j \cdot Ц_{вj}, \quad (4.11)$$

де $Н_j$ – норма витрат матеріалу $j$-го найменування, кг;

$n$ – кількість видів матеріалів;

$Ц_j$ – вартість матеріалу $j$-го найменування, грн/кг;

$К_j$ – коефіцієнт транспортних витрат, ($К_j$ = *1,1 … 1,15*);

$B_j$ – маса відходів $j$-го найменування, кг;

$Ц_{вj}$ – вартість відходів $j$-го найменування, грн/кг.

Проведені розрахунки зведемо до таблиці.

Таблиця 4.7 – Витрати на матеріали

| Найменування матеріалу, марка, тип, сорт | Ціна за 1 кг, грн | Норма витрат, кг | Вартість витраченого матеріалу, грн |
|---|---|---|---|
| Папір канцелярський офісний (А4) | 170 | 1 | 170 |
| ручка | 50 | 1 | 50 |
| Флешка | 250 | 1 | 250 |
| Всього | | | 470 |
| З врахуванням коефіцієнта транспортування | | | 517 |

### 4.3.4 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{обл} = \frac{Ц_б}{T_в} \cdot \frac{t_{вик}}{12}, \quad (4.12)$$

де $Ц_б$ – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{вик}$ – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_в$ – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$A_{обл} = (2600 \cdot 2) / (2 \cdot 12) = 216{,}67$ *грн.*

Проведені розрахунки зведемо до таблиці.

Таблиця 4.8 – Амортизаційні відрахування по кожному виду обладнання

| Найменування обладнання | Балансова вартість, грн | Строк корисного використання, років | Термін використання обладнання, місяців | Амортизаційні відрахування, грн |
|---|---|---|---|---|
| 1. IDE PyCharm | 2600 | 2 | 2 | 216,67 |
| 2.Компютерна техніка | 14000 | 2 | 2 | 1166,67 |
| 3. Сервіс шарингу | 3000 | 2 | 1 | 125,00 |
| Всього | | | | 1508,33 |

### 4.3.5 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію ($B_e$) розраховуємо за формулою:

$$B_e = \sum_{i=1}^{n} \frac{W_{yi} \cdot t_i \cdot Ц_e \cdot K_{вni}}{\eta_i}, \qquad (4.13)$$

де $W_{yi}$ – встановлена потужність обладнання на визначеному етапі розробки, кВт;

$t_i$ – тривалість роботи обладнання на етапі дослідження, год;

$Ц_e$ – вартість 1 кВт-години електроенергії, грн; (вартість електроенергії визначається за даними енергопостачальної компанії), приймемо $Ц_е = 7,5$ грн;

$K_{вni}$ – коефіцієнт, що враховує використання потужності, $K_{вni} < 1$;

$\eta_i$ – коефіцієнт корисної дії обладнання, $\eta_i < 1$.

$В_е = 0,25 \cdot 320,0 \cdot 7,5 \cdot 0,5 / 0,8 = 375$ грн.

### 4.3.6 Службові відрядження

До статті «Службові відрядження» дослідної роботи на тему «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів» належать витрати на відрядження штатних працівників, працівників організацій, які працюють за договорами цивільно-правового характеру, аспірантів, зайнятих розробленням досліджень, відрядження, пов'язані з проведенням випробувань машин та приладів, а також витрати на відрядження на наукові з'їзди, конференції, наради, пов'язані з виконанням конкретних досліджень.

Витрати за статтею «Службові відрядження» розраховуємо як 20…25% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{св} = (З_о + З_р) \cdot \frac{Н_{св}}{100\%}, \qquad (4.14)$$

де $Н_{св}$ – норма нарахування за статтею «Службові відрядження», приймемо $Н_{св}$ = 20%.

$В_{св}$ = (19318+1430,3) · 20 / 100% = 4149,70 грн.

### 4.3.8 Інші витрати

До статті «Інші витрати» належать витрати, які не знайшли відображення у зазначених статтях витрат і можуть бути віднесені безпосередньо на собівартість досліджень за прямими ознаками.

Витрати за статтею «Інші витрати» розраховуємо як 50…100% від суми основної заробітної плати дослідників та робітників за формулою:

$$I_в = (З_о + З_р) \cdot \frac{Н_{ів}}{100\%}, \qquad (4.15)$$

де $Н_{ів}$ – норма нарахування за статтею «Інші витрати», приймемо $Н_{ів}$ = 50%.

$І_в$ = (19318+1430,39) · 50 / 100% = 10374,26 грн.

### 4.3.9 Накладні (загальновиробничі) витрати

До статті «Накладні (загальновиробничі) витрати» належать: витрати, пов'язані з управлінням організацією; витрати на винахідництво та раціоналізацію; витрати на підготовку (перепідготовку) та навчання кадрів; витрати, пов'язані з набором робочої сили; витрати на оплату послуг банків; витрати, пов'язані з освоєнням виробництва продукції; витрати на науково-технічну інформацію та рекламу та ін.

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100…150% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{нзв} = (З_о + З_р) \cdot \frac{H_{нзв}}{100\%}\ , \qquad (4.16)$$

де $H_{нзв}$ – норма нарахування за статтею «Накладні (загальновиробничі) витрати», приймемо $Н_{нзв}$ = 100%.

$В_{нзв}$ = (19318+1430,3) · 100 / 100% = 20748,51 грн.

Витрати на проведення науково-дослідної роботи на тему «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$B_{заг} = З_о + З_р + З_{дод} + З_н + М + К_в + B_{спец} + B_{прг} + A_{обл} + B_е + B_{св} + B_{сп} + I_в + B_{нзв}. \qquad (4.17)$$

$В_{заг}$=19318+1430,3+2282,34+5066,79+517+1508,33+375+4149,70+10374,26 +20748,51= 65770,44 грн.

Загальні витрати *ЗВ* на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою:

$$ЗВ = \frac{B_{заг}}{\eta}, \qquad (4.18)$$

де $\eta$ - коефіцієнт, який характеризує етап (стадію) виконання науково-

дослідної роботи, приймемо $\eta$=0,7.

ЗВ = 65770,44 / 0,7= 93957,77 грн.

### 4.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором

В ринкових умовах узагальнюючим позитивним результатом, що його може отримати потенційний інвестор від можливого впровадження результатів тієї чи іншої науково-технічної розробки, є збільшення у потенційного інвестора величини чистого прибутку.

Результати дослідження проведені за темою «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів» передбачають комерціалізацію протягом 3-х років реалізації на ринку.

В цьому випадку основу майбутнього економічного ефекту будуть формувати:

$\Delta N$ – збільшення кількості споживачів яким надається відповідна інформаційна послуга у періоди часу, що аналізуються;

$N$ – кількість споживачів яким надавалась відповідна інформаційна послуга у році до впровадження результатів нової науково-технічної розробки, приймемо 1 особа

$Ц_б$ – вартість послуги у році до впровадження інформаційної системи, приймемо 2000,00 грн;

$\pm\Delta Ц_о$ – зміна вартості послуги від впровадження результатів, приймемо зростання на 500,00 грн.

Можливе збільшення чистого прибутку у потенційного інвестора $\Delta П_i$ для кожного із 3-х років, протягом яких очікується отримання позитивних результатів від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою **[]**:

$$\Delta П_i = (\pm \Delta Ц_o \cdot N + Ц_o \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot (1 - \frac{\vartheta}{100}), \quad (4.19)$$

де $\lambda$ – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2023 році ставка податку на додану вартість складає 20%, а коефіцієнт $\lambda = 0,8333$;

$\rho$ – коефіцієнт, який враховує рентабельність інноваційного продукту). Приймемо $\rho = 40\%$;

$\vartheta$ – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2023 році $\vartheta = 18\%$;

Збільшення чистого прибутку 1-го року:

$\Delta П_1 = (1 \cdot 500 + 20000 \cdot 70) \cdot 0,83 \cdot 0,4 \cdot (1 - 0,18/100\%) = 245221,44$ грн.

Збільшення чистого прибутку 2-го року:

$\Delta П_2 = (1 \cdot 500 + 20000 \cdot (70+60)) \cdot 0,83 \cdot 0,4 \cdot (1 - 0,18/100\%) = 455752,62$ грн.

Збільшення чистого прибутку 3-го року:

$\Delta П_3 = (1 \cdot 500 + 20000 \cdot (70+60+50)) \cdot 0,83 \cdot 0,4 \cdot (1 - 0,18/100\%) = 630849,79$ грн.

Приведена вартість збільшення всіх чистих прибутків *ПП*, що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ПП = \sum_{i=1}^{T} \frac{\Delta П_i}{(1+\tau)^t}, \quad (4.20)$$

де $\Delta П_i$ – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

$T$ – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

$\tau$ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні, $\tau = 18\%$;

$t$ – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$ПП$ =245221,44 /(1+0,18)$^1$+455752,62 /(1+0,18)$^2$+630849,79 /(1+0,18)$^3$=

= 887619,21 грн.

Величина початкових інвестицій $PV$, які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{інв} \cdot ЗВ, \tag{4.21}$$

де $k_{інв}$ – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію, приймаємо $k_{інв}$=2;

$ЗВ$ – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 93957,77 грн.

$PV = k_{інв} \cdot ЗВ$ = 2 · 93957,77 = 187915,54 грн.

Абсолютний економічний ефект $E_{абс}$ для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = ПП - PV \tag{4.22}$$

де $ПП$ – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 887619,21 грн;

$PV$ – теперішня вартість початкових інвестицій, 187915,54 грн.

$E_{абс} = ПП - PV$ = 887619,21 - 187915,54 =699703,67 грн.

Внутрішня економічна дохідність інвестицій $E_в$, які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$E_в = \sqrt[Tж]{1 + \frac{E_{абс}}{PV}} - 1, \quad (4.23)$$

де $E_{абс}$ – абсолютний економічний ефект вкладених інвестицій, грн;

$PV$ – теперішня вартість початкових інвестицій, грн;

$T_ж$ – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримування позитивних результатів від її впровадження, 3 роки.

$$E_в = \sqrt[Tж]{1 + \frac{E_{абс}}{PV}} - 1 = (1+699703{,}67\ /187915{,}54)^{1/3} - 1 = 1{,}04.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій $\tau_{мін}$:

$$\tau_{мін} = d + f, \quad (4.24)$$

де $d$ – середньозважена ставка за депозитними операціями в комерційних банках; в 2023 році в Україні $d$ =0,1;

$f$ – показник, що характеризує ризикованість вкладення інвестицій, приймемо 0,25.

$\tau_{мін}$ = 0,1+0,25 = 0,35 < 1,04 свідчить про те, що внутрішня економічна дохідність інвестицій $E_в$, які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки вища мінімальної внутрішньої дохідності. Тобто інвестувати в науково-дослідну роботу за темою «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів» доцільно.

Період окупності інвестицій $T_{ок}$ які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_в}, \quad (4.25)$$

де $E_в$ – внутрішня економічна дохідність вкладених інвестицій.

$T_{ок}$ = 1 / 1,04 = 1 р.

$T_{ок}$ < 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

### 4.5 Висновки до розділу

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів» становить 38 бали, що, свідчить про комерційну важливість проведення даних досліджень оскільки рівень комерційного потенціалу розробки вище середнього.

При оцінюванні рівня конкурентоспроможності, згідно узагальненого коефіцієнту конкурентоспроможності розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 2,8 рази.

Також термін окупності становить 1 рік, що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже можна зробити висновок про доцільність проведення науково-дослідної роботи за темою «Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів».

# ВИСНОВКИ

У магістерській роботі про "Систему виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів" виявлено, що застосування теорії графів у спільнотах соціальних мереж може бути цінним інструментом для аналізу кіберзагроз.Під час обчислення міри центральності, зокрема Edge Betweenness, виявлено ребра, які є критичними для спільноти. Високі значення цієї метрики можуть вказувати на ребра, через які проходить багато найкоротших шляхів між учасниками спільноти. Це може свідчити про потенційні точки вразливості, через які можуть бути здійснені кібератаки або які можна використовувати для впливу на спільноту.Отримані результати показують, що аналіз теорії графів дозволяє виявляти ключові ребра у спільнотах соціальних мереж, які можуть бути важливими для стійкості та безпеки цих спільнот. Проте, важливо розуміти, що міра центральності, включаючи Edge Betweenness, є лише однією з багатьох метрик, і для повного розуміння кіберзагроз у соціальних мережах необхідно застосовувати комплексний підхід, враховуючи інші аспекти безпеки, структуру спільноти та їхню динаміку.Загалом, модуль моніторингу та аналізу спільнот за допомогою теорії графів може бути корисним інструментом для виявлення потенційних загроз у соціальних мережах, але його ефективність повинна бути підтверджена та доповнена іншими методами аналізу та захисту мережі.

# ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ


1. Спекторський І. Я., Стусь О. В., Статкевич В. М. Дискретна математика. Збірник задач: навч. посіб. Київ: НТУУ «КПІ», 2015. 105 с.
2. Efficient Algorithms for Path Problems in Weighted Graphs. URL: https://people.csail.mit.edu/virgi/thesis.pdf (дата звернення 16.04.2019).
3. Cvetkovic D. The Generating Function for Variations with Restrictions and Paths of the Graph and Self-Complementary Graphs. URL: http://pefmath2.etf.rs/files/86/322.pdf (дата звернення 16.04.2019).
4. Cvetkovic D. Die Zahl der Wege eines Grafen. Beograd: Univ. Beograd, 1970. 210 p.
5. Spectra of Simple Graphs. Owen Jones, 2013. URL: https://www.whitman.edu/Documents/Academics/Mathematics/Jones.pdf (дата звернення 20.04.2019).
6. Карпов Д.В. Теория графов. URL: https://logic.pdmi.ras.ru/~dvk/graphs_dk.pdf (дата звернення 01.05.2019).
7. Chung R. K. Lectures on Spectral Graph Theory. URL: http://www.math.ucsd.edu/~fan/cbms.pdf (дата звернення 05.05.2019).
8. The Shortest Path algorithm. URL: https://neo4j.com/docs/graph-algorithms/current/algorithms/shortest-path/ (дата звернення 27.04.2019).
9. Берцун В. Н. Математическое моделирование на графах. Томск: Издательство Том. ун-та, 2013. 88 с.
10. Домнін Л. Н. Елементи теорії графів: навчальний посібник. Пенза: Видавництво Пенз. ун-ту, 2007. 144 с.
11. Andoni A. Spectral Graph Algorithms. URL: http://www.mit.edu/~andoni/s17_advanced/algorithms/mainSpace/files/scribe13.pdf (дата звернення 25.04.2019).
12. Norman B. Algebraic Graph Theory. Cambridge: Cambridge University Press, 1993. 247 p.
13. Chung R. K. Review of Spectral Graph Theory. URL: https://dl.acm.org/citation.cfm?id=568553 (дата звернення 28.05.2019).

14. Mathematical Python. Eigenvalues and Eigenvectors. URL: https://www.math.ubc.ca/~pwalls/math-python/linear-algebra/eigenvalues-eigenvectors/ (дата звернення 10.05.2019).
15. Graph Plotting in Python. URL: https://www.geeksforgeeks.org/graph-plotting-in-python-set-1/ (дата звернення 12.05.2019).
16. J. D. Fehribach, Diffusion-reaction-conduction processes in porous electrodes: the electrolyte wedge problem, European J. Appl. Math. 12 (2001), 77–96.
17. I. Fishtik and R. Datta, A reaction route network for hydrogen combustion, Physica A 373 (2007), 777–784.
18. J.Newman and K. Thomas-Alyea, Electrochemical Systems, Prentice Hall, Englewood Cliffs, NJ, 3rd edition, 2004
19. Brett W. Bader and Tamara G. Kolda. MATLAB Tensor Toolbox Version 2.2. http://csmr.ca.sandia.gov/ tgkolda/TensorToolbox/, 2007. 30
20. Deng Cai, Zheng Shao, Xiaofei He, Xifeng Yan, and Jiawei Han. Mining hidden community in heterogeneous social networks. In Workshop on Link Discovery: Issues, Approaches and Applications (LinkKDD-2005), 2005. 24
21. Ben Taskar, Eran Segal, and Daphne Koller. Probabilistic classification and clustering in relational data. In Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI-01), 2001. 18
22. A. Francisco, C. Vaz, P. Monteiro, J. Melo-Cristino, M. Ramirez, and J. Carri¸co. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics, 13(1):87, 2012.
23. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. Introduction To Algorithms. MIT Press, 2001.
24. M. Girvan and M. E. J. Newman. Community structure in social and biological networks. PNAS, 99(12):7821–7826, April 2002
25. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. – Вінниця : ВНТУ, 2021. – 42 с.

# ДОДАТКИ

85

**ДОДАТОК А**
**ПРОТОКОЛ ПЕРЕВІРКИ**
**МАГІСТЕРСЬКОЇ КВАЛІФІКАЦІЙНОЇ РОБОТИ**
**НА НАЯВНІСТЬ ТЕКСТОВИХ ЗАПОЗИЧЕНЬ**

Назва роботи: Система виявлення кіберзагроз, що створюються спільнотами в соціальних мережах. Частина 2. Модуль моніторингу та аналізу спільнот за допомогою теорії графів.

Автор роботи: Стаднік Вадим Леонідович

Тип роботи: магістерська кваліфікаційна робота

Підрозділ кафедра захисту інформації ФІТКІ
(кафедра, факультет)

**Показники звіту подібності Unicheck**

Оригінальність – 93,06 %. Схожість – 6,94 %.

Аналіз звіту подібності (відмітити потрібне):

√ 1. Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.

☐ 2. Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і відсутності самостійності її виконання автором. Роботу направити на розгляд експертної комісії кафедри.

☐ 3. Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Особа, відповідальна за перевірку _____ (підпис) Валентина КАПЛУН

Ознайомлені з повним звітом подібності, який був згенерований системою Unicheck щодо роботи.

Автор роботи _____ (підпис) Вадим СТАДНІК

Керівник роботи _____ (підпис) Наталія КОНДРАТЕНКО

# ДОДАТОК Б
## Текст програми частини

```
import networkx as nx
import matplotlib.pyplot as plt
import matplotlib
import tkinter as tk
from tkinter import filedialog
import random

matplotlib.use('TkAgg')  # Зміна бекенду

def highlight_critical_edges(edge_path_count):
    critical_edges = [edge for edge, count in edge_path_count.items() if count ==
max(edge_path_count.values())]

    plt.figure(figsize=(6, 6))
    pos = nx.spring_layout(G)

    non_critical_edges = [edge for edge in G.edges() if edge not in critical_edges]
    nx.draw_networkx_edges(G, pos, edgelist=non_critical_edges, edge_color='gray', ax=plt.gca())

    nx.draw_networkx_edges(G, pos, edgelist=critical_edges, edge_color='red', width=2.0, ax=plt.gca())

    nx.draw(G, pos, with_labels=True, node_size=300, node_color='skyblue', font_weight='bold',
font_size=8, ax=plt.gca())
    plt.title("Community Graph with Limited Connections")
    plt.tight_layout()
    plt.show()

def update_graph_visualization():
    plt.figure(figsize=(6, 6))
    pos = nx.spring_layout(G)
    nx.draw(G, pos, with_labels=True, node_size=300, node_color='skyblue', font_weight='bold',
font_size=8, ax=plt.gca())
    plt.title("Community Graph with Limited Connections")
    plt.tight_layout()
    plt.show()

def calculate_metrics():
    edge_betweenness = nx.edge_betweenness_centrality(G)
    max_centrality = max(edge_betweenness.values())
    edge_centrality_percentage = {edge: (centrality / max_centrality) * 100 for edge, centrality in
edge_betweenness.items()}

    shortest_paths = dict(nx.all_pairs_shortest_path(G))

    edge_path_count = {edge: 0 for edge in G.edges()}
    for paths in shortest_paths.values():
        for target_node, path in paths.items():
            if len(path) > 1:
                for i in range(len(path) - 1):
                    edge = (path[i], path[i + 1])
                    if edge in edge_path_count:
                        edge_path_count[edge] += 1
```

```
    update_graph_visualization()
    update_results_window(edge_path_count, edge_centrality_percentage)
    highlight_critical_edges(edge_path_count)

def update_results_window(edge_path_count, edge_centrality_percentage):
    text_box1.delete(1.0, tk.END)
    text_box2.delete(1.0, tk.END)

    sorted_edge_paths = sorted(edge_path_count.items(), key=lambda x: x[1], reverse=True)
    for edge, count in sorted_edge_paths:
        text_box1.insert(tk.END, f"Edge {edge} - path count: {count}\n")

    sorted_edge_centrality = sorted(edge_centrality_percentage.items(), key=lambda x: x[1], reverse=True)
    for edge, centrality in sorted_edge_centrality:
        text_box2.insert(tk.END, f"Edge {edge} - centrality: {centrality:.2f}%\n")

def load_graph_from_file():
    filename = filedialog.askopenfilename(filetypes=[("Text files", "*.txt")])
    if filename:
        G.clear()
        with open(filename, 'r') as file:
            lines = file.readlines()
            for line in lines:
                nodes = line.strip().split()
                if len(nodes) == 2:
                    node1, node2 = nodes[0], nodes[1]
                    G.add_edge(f"Person {node1}", f"Person {node2}")
        calculate_metrics()
root = tk.Tk()
root.title("Results")

frame1 = tk.Frame(root)
frame1.pack(side=tk.LEFT, padx=10, pady=10)
frame2 = tk.Frame(root)
frame2.pack(side=tk.LEFT, padx=10, pady=10)
frame3 = tk.Frame(root)
frame3.pack()

label1 = tk.Label(frame1, text="Shortest path count per edge (sorted):")
label1.pack()

text_box1 = tk.Text(frame1, width=50, height=20)
text_box1.pack()

label2 = tk.Label(frame2, text="Edge betweenness centrality (sorted in percentages):")
label2.pack()

text_box2 = tk.Text(frame2, width=50, height=20)
text_box2.pack()

load_button = tk.Button(frame3, text="Load graph from file", command=load_graph_from_file)
load_button.pack()

G = nx.Graph()

root.mainloop()
```

**Додаток В**
**ІЛЮСТРАТИВНА ЧАСТИНА**

# АКТУАЛЬНІСТЬ , МЕТА ТА ЗАДАЧІ МКР

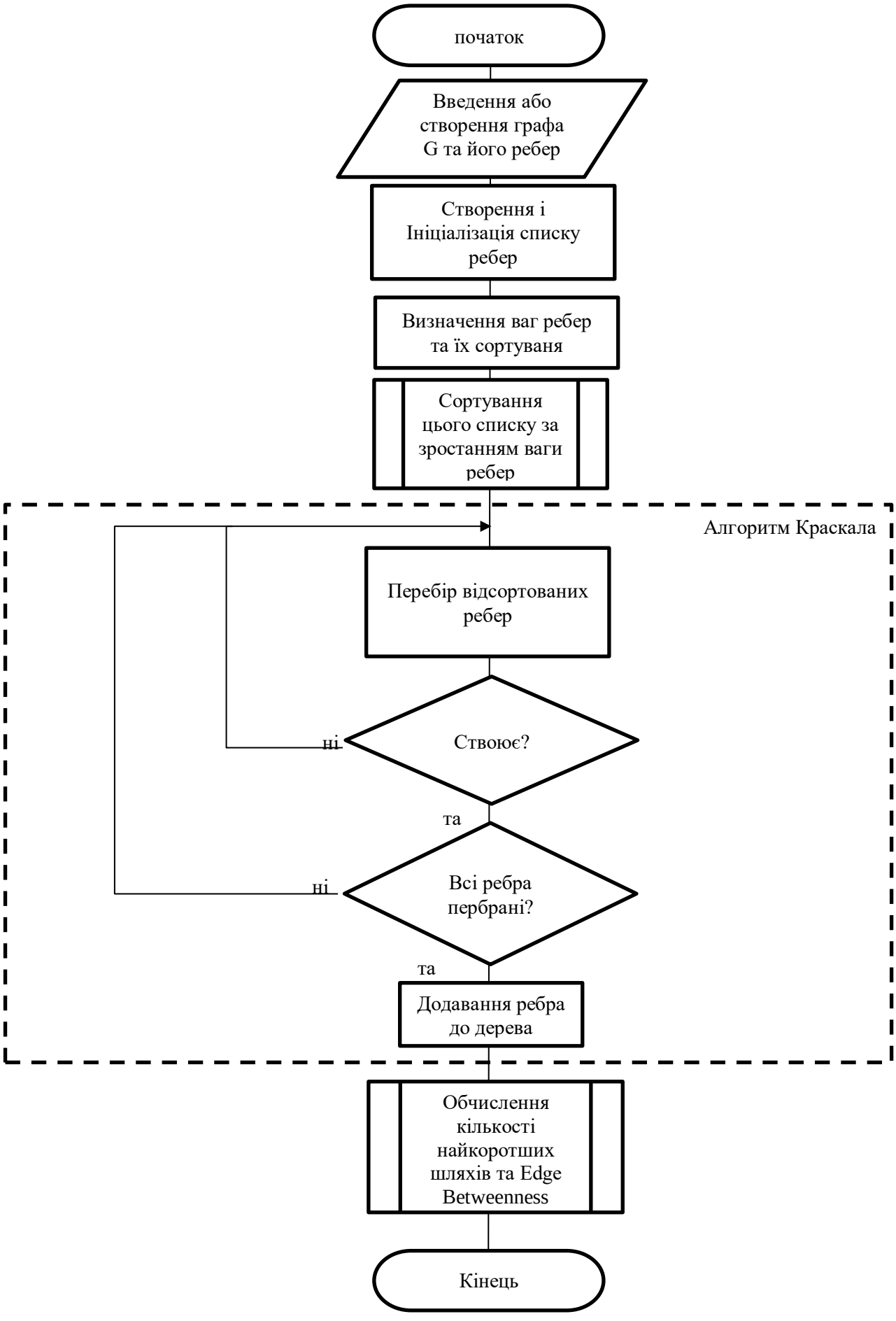
Актуальність дослідження: теорія графів дає уявлення про зв'язки між акаунтами та групами, виявляє впливові точки та аномалії, допомагаючи у боротьбі з дезінформацією та кібератаками. Це потужний інструмент, але вимагає доступу до великого обсягу даних і уважного ставлення до етичних питань щодо приватності користувачів.

Об’єкт дослідження: процес моніторингу та аналізу спільнот за допомогою теорії графів

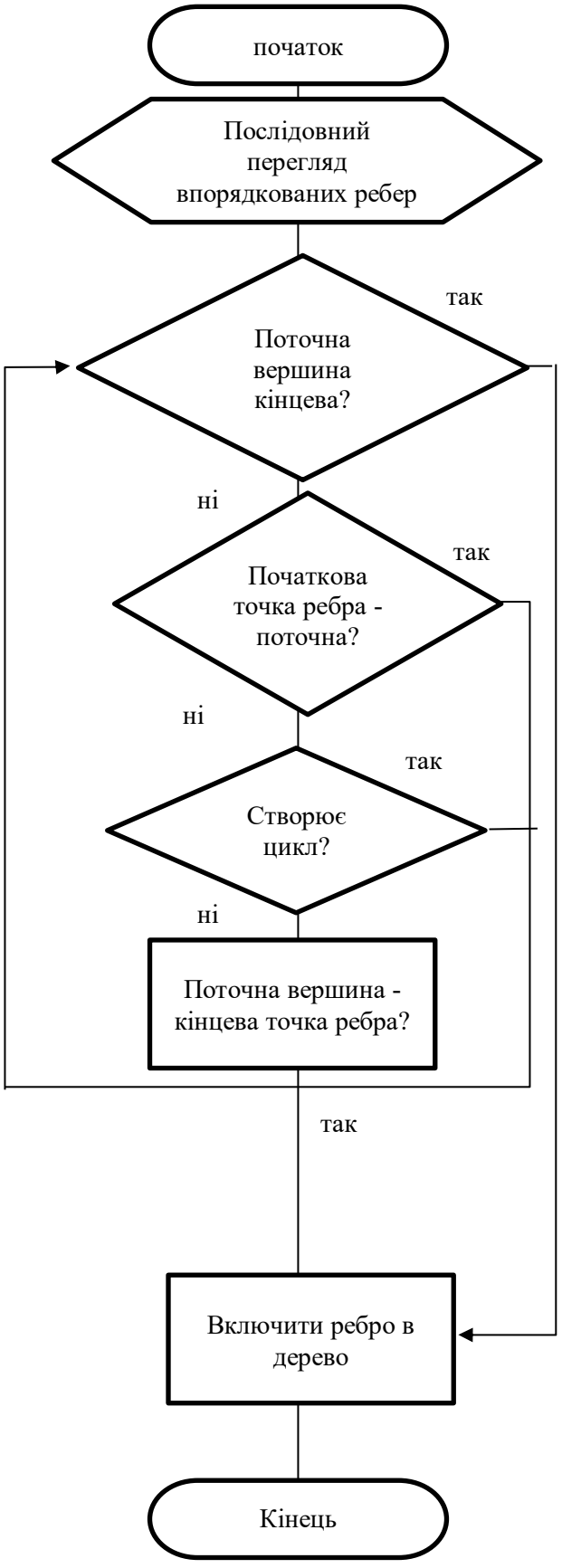
Предмет дослідження: моделі та методи аналізу спільнот за допомогою теорії графів.

Мета дослідження: розширення можливостей для аналізу та моніторингу кіберзагроз, що створюються спільнотами в соціальних мережах, на основі теорії графів.

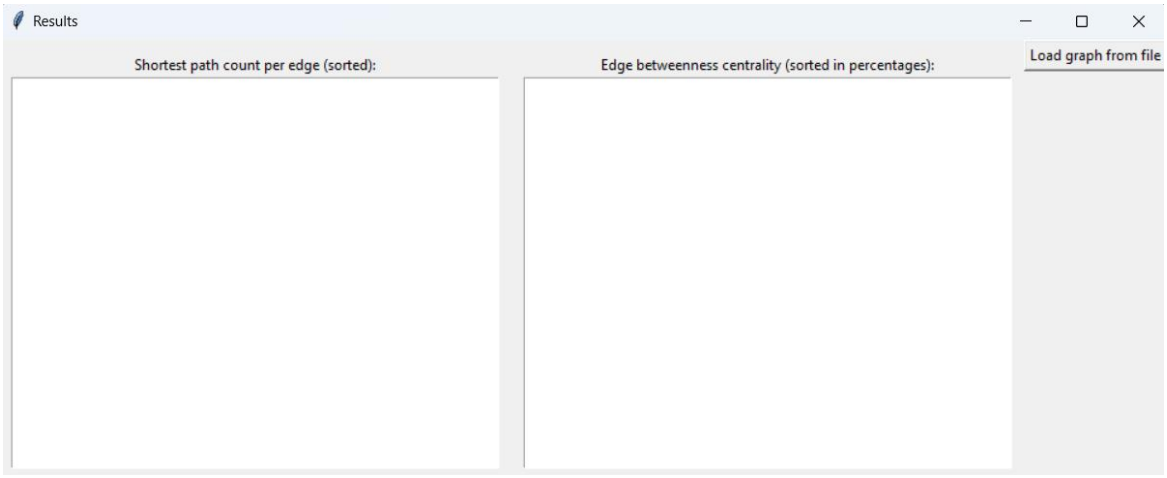
# СХЕМА АЛГОРИТМУ РОБОТИ ПРОГРАМНОГО ЗАСОБУ



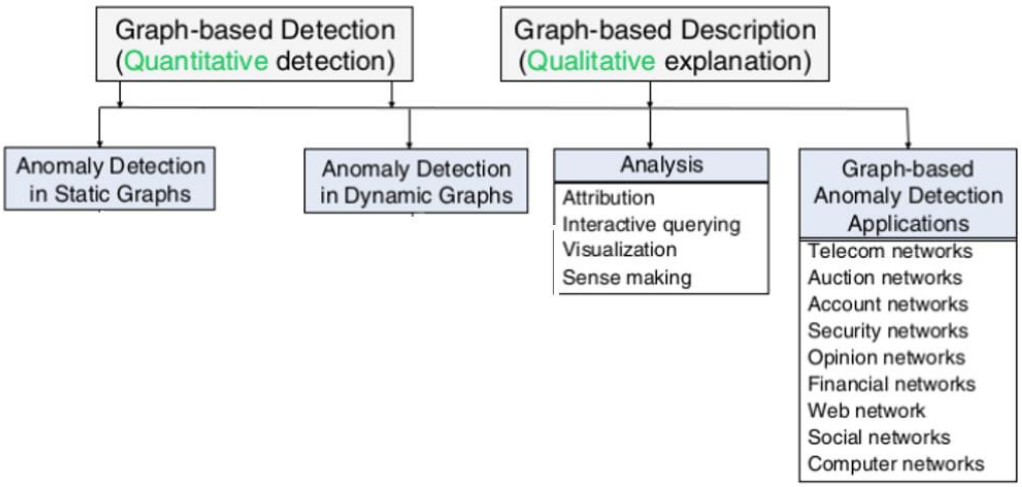
# СХЕМА АЛГОРИТМУ КРАСКАЛА



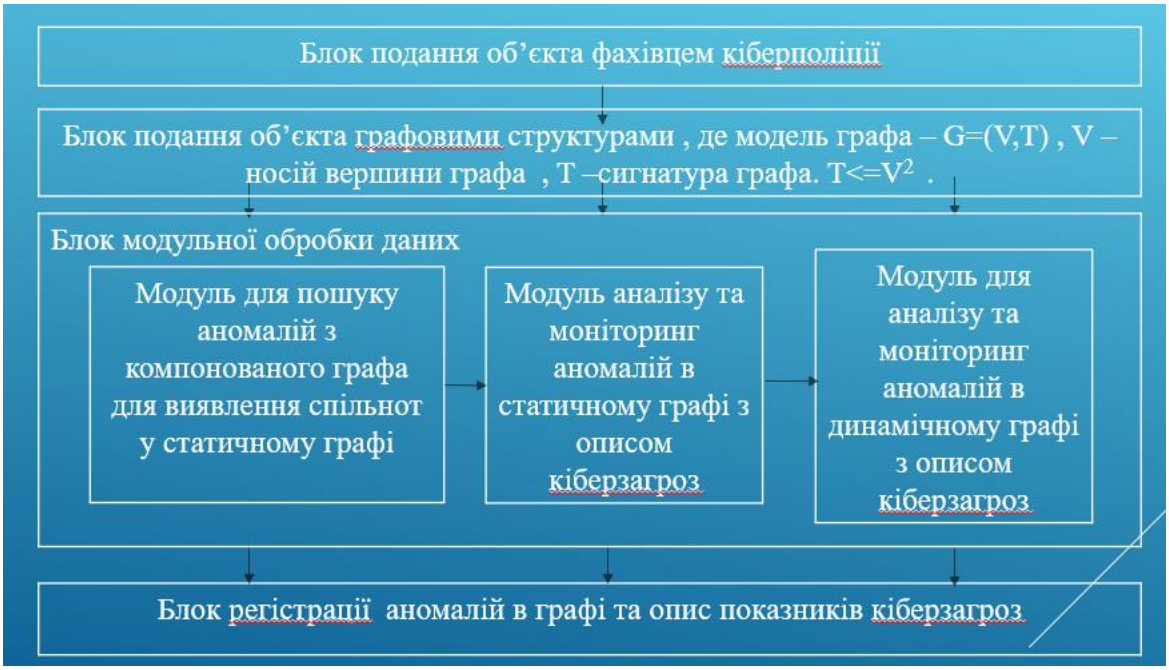
# ІНТЕРФЕЙС ПРОГРАМНОГО ЗАСОБУ



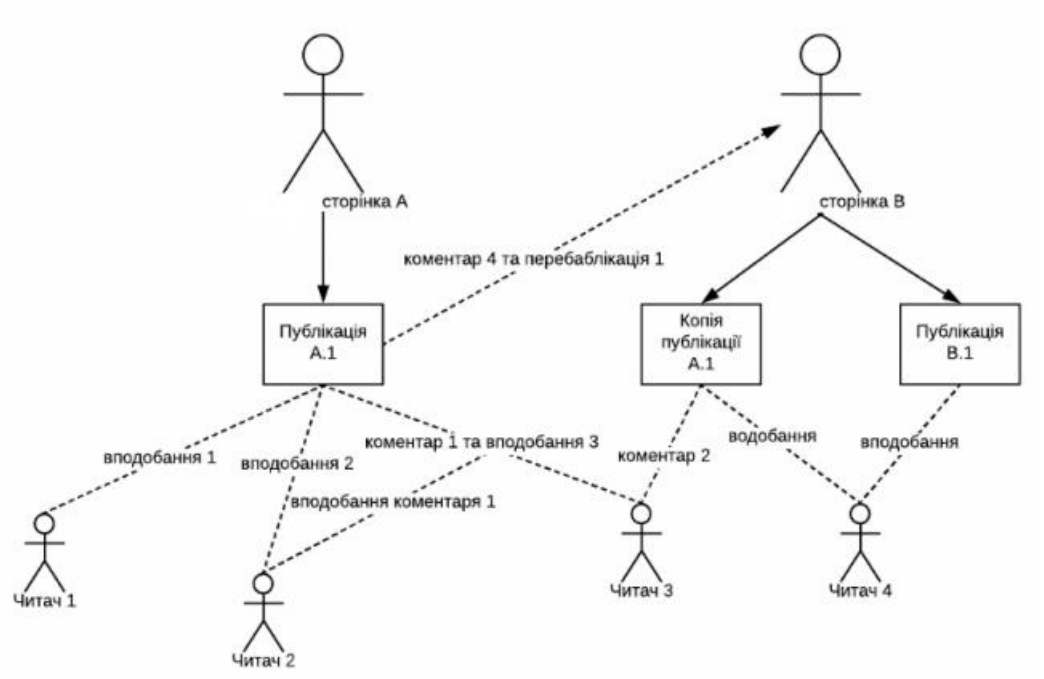
# СУЧАСНІ НАПРЯМКИ МОНІТОРИНГУ МЕРЕЖ НА ОСНОВІ ГРАФІВ



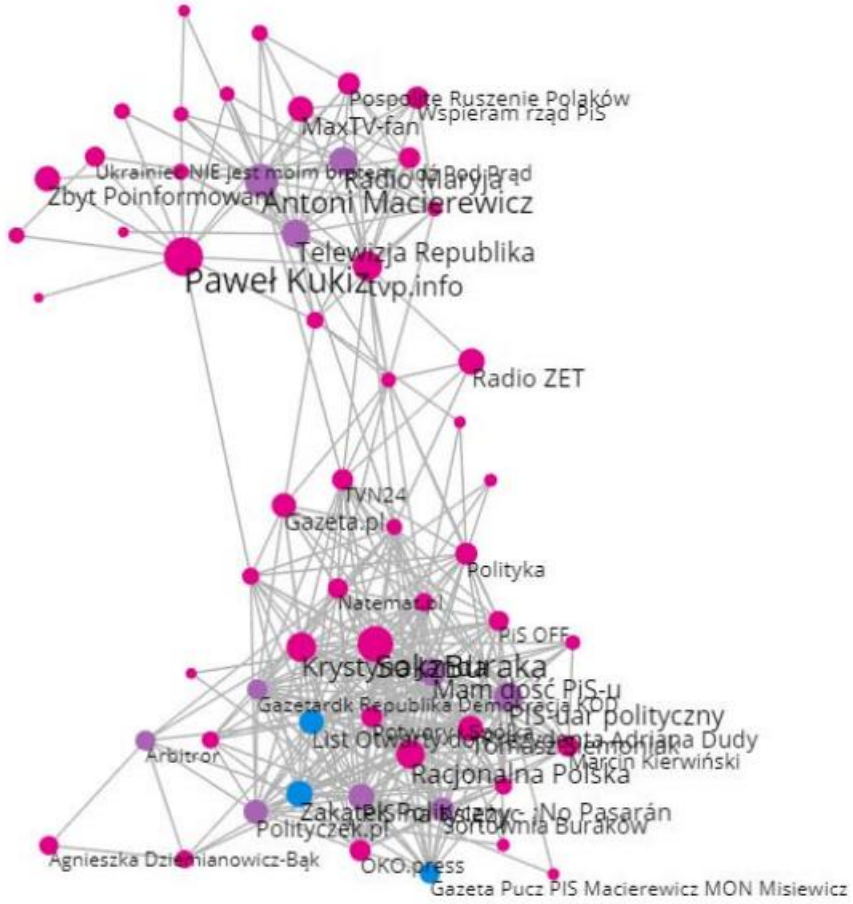
# СИСТЕМНИЙ ПІДХІД ДО ВИЯВЛЕННЯ КІБЕРЗАГРОЗ, ЩО УТВОРЮЮТЬСЯ СПІЛЬНОТАМИ СОЦІАЛЬНИХ МЕРЕЖ НА ОСНОВІ ТЕОРІЇ ГРАФІВ



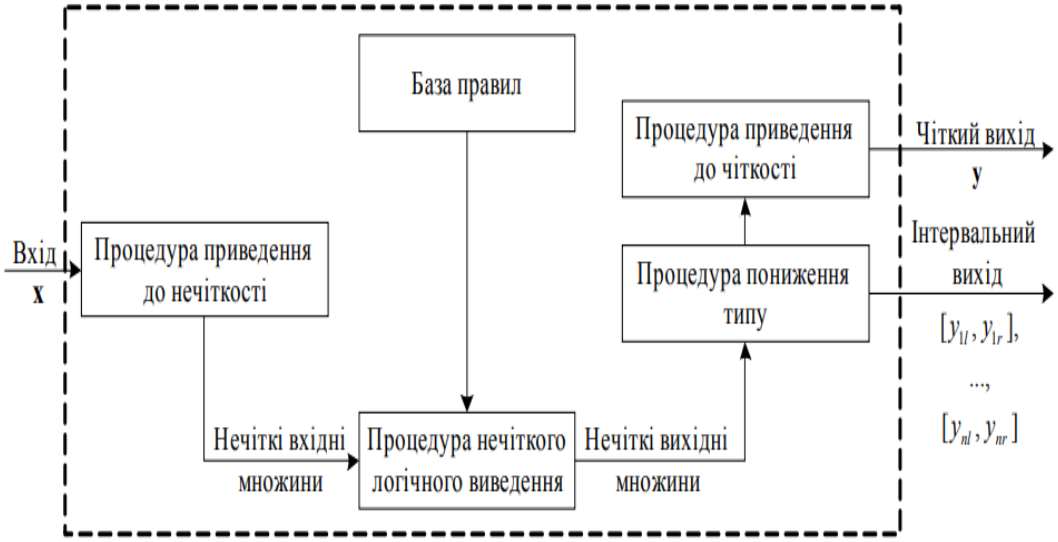
# СТРУКТУРА СПІЛКУВАННЯ МЕРЕЖІ



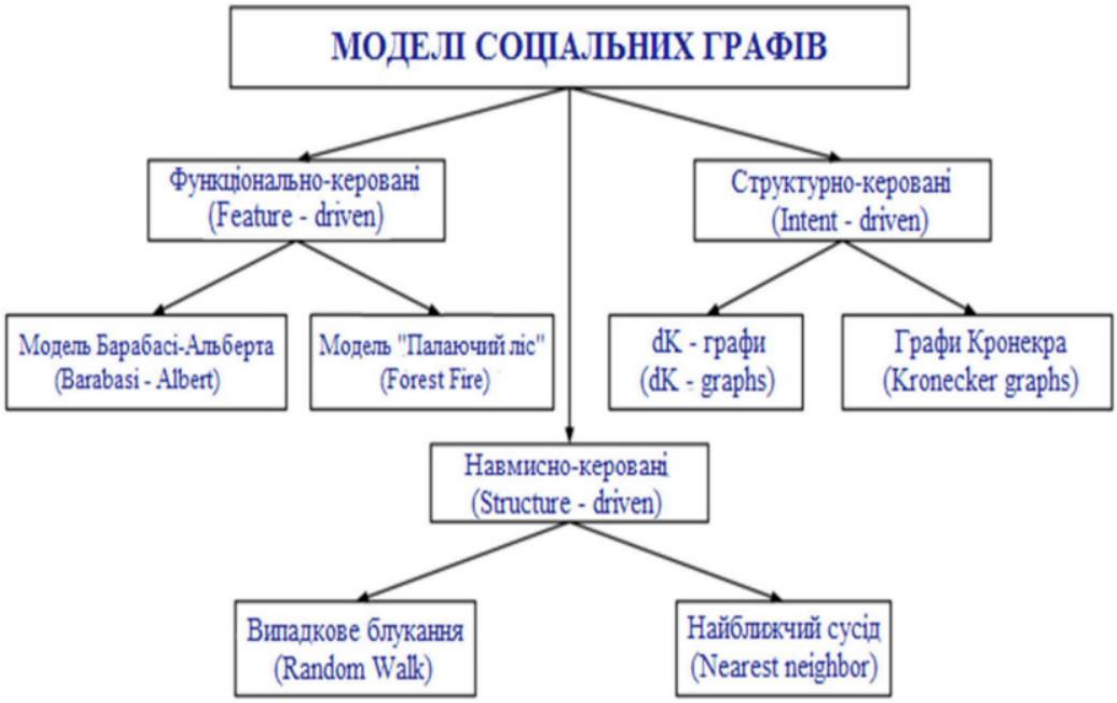
# ПРИКЛАД ПОБУДОВАНОЇ МЕРЕЖІ



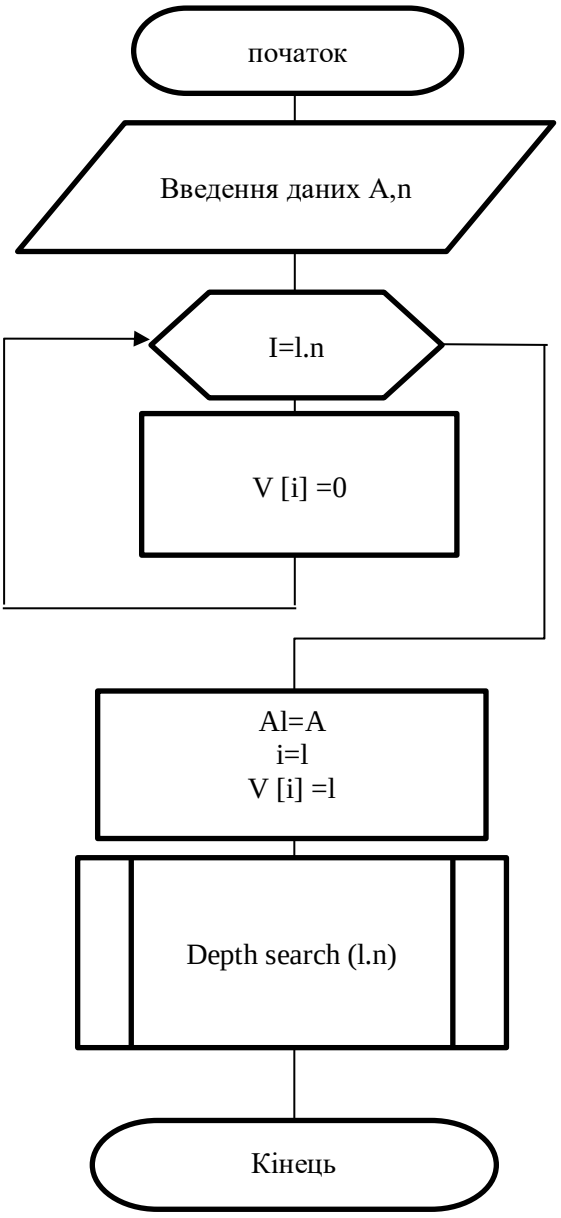
# СТРУКТУРА ІНТЕРВАЛЬНОЇ НЕЧІТКОЇ МОДЕЛІ



# МОДЕЛІ СОЦІАЛЬНИХ ГРАФІВ



# БЛОК СХЕМА АЛГОРИТМУ ПОШУКУ ГРАФА В ГЛИБИНУ



# БЛОК СХЕМА АЛГОРИТМУ ДЕЙКСТРИ