

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ІМЕНІ ТАРАСА ШЕВЧЕНКА**

**ФАКУЛЬТЕТ РАДІОФІЗИКИ ЕЛЕКТРОНІКИ ТА КОМП'ЮТЕРНИХ СИСТЕМ**

**Кафедра радіотехніки та радіоелектронних систем**

До захисту допущено:

«На правах рукопису»

Завідувач кафедри \_\_\_\_\_ Ігор АНІСІМОВ

18 травня 2023 р.

**КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА**

на тему:

**«Програмний засіб для дослідження та аналізу соціальних мереж»**

**Виконав:**

студент 2-го курсу магістратури

денної форми навчання

спеціальності 172 Телекомунікації та радіотехніка

ОНП «Інформаційна безпека телекомунікаційних систем і мереж»

Цюрпи́та Михайло Євге́нійович \_\_\_\_\_

**Науковий керівник:**

к.ф.-м. н., ас. Котов Михайло Миколайович \_\_\_\_\_

**Рецензент:**

к.т.н., доц. Четверіков Іван Олександрович \_\_\_\_\_

Засвідчую, що у цій магістерській роботі

немає заповичень з праць інших авторів без

відповідних посилань

Студент \_\_\_\_\_

Робота допущена до захисту в ЕК рішенням кафедри радіотехніки та радіоелектронних систем від 18 травня 2023 р., протокол № 18.

Завідувач кафедри радіотехніки та радіоелектронних систем,

доктор фіз.-мат. наук, професор

Анісімов Ігор Олексійович \_\_\_\_\_

## ЗМІСТ

ВСТУП .....	3
РОЗДІЛ 1. АНАЛІТИКА СОЦІАЛЬНИХ МЕРЕЖ .....	4
1.1. Історія розвитку аналізу та захисту даних соціальних мереж .....	4
1.2. Збір та аналіз даних соціальних мереж.....	6
РОЗДІЛ 2. ПРОГРАМНИЙ ПАКЕТ ДЛЯ АНАЛІЗУ СОЦІАЛЬНОЇ МЕРЕЖІ	10
2.1. Функціональні можливості програмного забезпечення.....	11
2.2. Обґрунтування вибору мови та середовища програмування .....	11
2.3. Опис роботи програми.....	13
2.4. Візуалізація зв'язків у соціальній мережі .....	17
2.5. Подальше вдосконалення ПЗ .....	19
ВИСНОВКИ.....	21
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	22

## ВСТУП

Соціальні мережі стали невід'ємною частиною нашого життя та суттєво впливають на нашу поведінку, взаємодію та спілкування. У зв'язку з цим, обробка та аналіз даних соціальних мереж є важливою задачею в багатьох галузях, включаючи маркетинг, соціологію, психологію та багато інших.

Метою дипломної роботи є розробка програмного забезпечення для аналізу даних соціальної мережі, що дозволяє відстежувати та аналізувати поведінку користувачів, їхні взаємодії та спілкування. Програма має зібрати, обробляти та візуалізувати дані з різних соціальних мереж: Facebook, Twitter, Instagram тощо.

Дана дипломна робота є актуальною та важливою в зв'язку зі зростанням кількості користувачів соціальних мереж та потребою в аналізі даних для розуміння їх поведінки та взаємодії з іншими користувачами. Результати дослідження можуть бути корисними для бізнесу, соціальних наук, політики та інших галузей, які використовують дані соціальних мереж.

## РОЗДІЛ 1. АНАЛІТИКА СОЦІАЛЬНИХ МЕРЕЖ

У теоретичній частині даного дипломної роботи розглянуті основні теоретичні аспекти аналізу даних соціальних мереж. Зокрема, досліджено методи та техніки збору та обробки даних, статистичний аналіз даних, візуалізацію даних, а також методи машинного навчання, які можна використовувати для аналізу соціальних мереж.

### 1.1. Історія розвитку аналізу та захисту даних соціальних мереж

Історія розвитку аналізу та захисту соціальних мереж відносно недавня, оскільки соціальні мережі, які ми знаємо сьогодні, з'явилися зовсім недавно. Перші соціальні мережі, які з'явилися в 1990-х роках, були простими мережами для спілкування між користувачами, що використовувалися головним чином в академічних колах і спільнотах.

З появою Facebook у 2004 році та інших подібних соціальних мереж, таких як Twitter, LinkedIn, Instagram тощо, відбулося значне зростання кількості користувачів і їх активності. Це призвело до створення та розвитку нових методів аналізу соціальних мереж та їх використання у різних галузях, включаючи маркетинг, наукові дослідження, політику, безпеку тощо.

Однак разом з зростанням популярності соціальних мереж з'явилася й загроза безпеці та конфіденційності даних користувачів. Хакерські атаки, витоки даних, шахрайство та інші проблеми стали найбільшою загрозою для безпеки соціальних мереж. Це призвело до розвитку нових методів та технологій захисту соціальних мереж та їх користувачів.

Сьогодні відбувається постійне зростання кількості користувачів соціальних мереж та їх активності. Це призводить до постійного розвитку нових методів аналізу та захисту соціальних мереж, а також до появи нових викликів та загроз, з якими необхідно боротися.

У 2013 році розгорнулася справа про витік даних користувачів Facebook компанією Cambridge Analytica. Ця компанія використовувала

отримані дані для впливу на виборчий процес в США та інших країнах. Цей скандал став поворотним моментом для захисту та регулювання персональних даних в соціальних мережах. В результаті, були прийняті закони та стандарти щодо захисту персональних даних користувачів, такі як загальний регламент про захист персональних даних (GDPR) в Європейському Союзі та Закон про захист персональних даних (CCPA) в Каліфорнії [1].

Окрім захисту даних, аналіз соціальних мереж також став важливим для бізнесу та політики. Дані, які надходять з соціальних мереж, дозволяють компаніям та політикам аналізувати поведінку користувачів, їхні інтереси та потреби. Наприклад, це допомагає бізнесу визначити цільову аудиторію та забезпечити їм належну рекламу, а політикам зрозуміти настрої громадськості та розробити ефективну виборчу стратегію.

В даний час аналіз та захист соціальних мереж є одним з найбільш актуальних та широко обговорюваних питань в галузі інформаційної безпеки та приватності. Наслідком цього є поява нових технологій та рішень, що дозволяють ефективно аналізувати дані та захищати користувачів соціальних мереж від шкідливих впливів.

Аналіз соціальних мереж широко використовується в ряді застосунків і дисциплін. Деякі поширені програми мережевого аналізу охоплюють збір і накопичення даних, моделювання поширення мережі, моделювання мережі і вибірок, аналіз характерних ознак і поведінки користувача, ресурсної підтримки, забезпечувальною спільнотою, аналіз взаємодії на основі місця розташування, соціальний обмін і відбір, розвиток систем рекомендацій, а також прогнозування зв'язків і аналіз об'єктів. У приватному секторі фірми використовують аналіз соціальних мереж для підтримання такої діяльності, як взаємодія та аналіз клієнтів, маркетинг і бізнес-аналітика. Використання аналізу соціальних мереж державним сектором включає розвиток стратегій участі керівництва, аналіз індивідуальної та групової участі, використання засобів масової інформації і засноване на спільнотах вирішення проблем.

Аналіз соціальних мереж так само використовується у розвідувальних, контррозвідувальних та правоохоронних заходах. Ця техніка дозволяє аналітикам відобразити на карті нелегальну або приховану організацію, таку як шпигунське коло, організовану злочинну групу або вуличну банду. Агентство національної безпеки (NSA) використовує програми таємних масових систем електронного спостереження для генерування даних, необхідних для подання цього типу аналізу в терористичних осередках і інших мережах, що стосуються національної безпеки. У процесі мережевого аналізу Агентство національної безпеки здійснює пошук у глибину на три вузли. Після того, як завершилося початкове відображення соціальної мережі, виконується аналіз для визначення структури мережі і, наприклад, лідера мережі. Це дозволяє військовим або правоохоронним органам завдати нищівних ударів для захоплення або знищення найзначущих цілей, які займають лідерські позиції, що призводить до порушення функціонування мережі.

Невдовзі після терористичного акту 11 вересня Агентство національної безпеки розпочало використання аналізу соціальних мереж під час детальної реєстрації викликів, які являють собою метадані [2].

## 1.2. Збір та аналіз даних соціальних мереж

### Збір даних

Один з найважливіших етапів аналізу даних соціальних мереж - це збір даних. Збір даних може бути здійснений шляхом використання API соціальної мережі, якщо такий існує, або за допомогою веб-скрапінгу. Веб-скрапінг - це процес автоматичного збору даних з веб-сайтів, що може бути здійснений за допомогою програмних засобів, таких як BeautifulSoup або Scrapy.

### Обробка даних

Після збору даних наступним етапом є їх обробка. Обробка даних може включати в себе очищення даних, обрізання, стандартизацію та інші операції,

які допоможуть підготувати дані для подальшого аналізу. Також може бути проведений аналіз сутностей, що дозволяє виокремити зі зібраних даних корисну інформацію, таку як імена користувачів, дати, локації тощо.

#### Статистичний аналіз даних

Статистичний аналіз даних - це набір методів та технік, які дозволяють виконувати описовий та інференційний аналіз даних. Описовий аналіз даних дозволяє отримати загальну інформацію про дані, таку як середнє значення, медіану, моду тощо.

#### Інференційний аналіз

Інференційний аналіз даних дозволяє здійснювати висновки про популяцію на основі вибірки даних. Для проведення інференційного аналізу можна використовувати різноманітні методи, такі як t-тести, аналіз дисперсії, кореляційний аналіз тощо.

#### Візуалізація даних

Візуалізація даних є важливим етапом аналізу даних, оскільки вона дозволяє зрозуміти дані та знайти закономірності та зв'язки, які можуть бути невидимими при аналізі самої інформації. Візуалізація даних може бути здійснена за допомогою різноманітних графіків та діаграм, таких як стовпчикові діаграми, кругові діаграми, лінійні графіки, теплові карти тощо.

Візуальне подання соціальних мереж важливе для розуміння даних мережі і передавання результатів аналізу. У більшості випадків аналітичне програмне забезпечення має модулі для візуалізації мережі. Дослідження даних проводиться відображенням вузлів і зв'язків у різних шарах, а також призначення вузлам кольорів, розмірів та інших додаткових властивостей. Візуальне подання мереж може виступати потужним методом передавання складної інформації, але слід обережно інтерпретувати вузли і властивості графу, ґрунтуючись виключно на відображенні, оскільки структурні особливості, які найкращим чином охоплює кількісний аналіз, можуть спотворити. Типове подання візуалізації соціальних мереж зображене на рис. 1.1 [3].

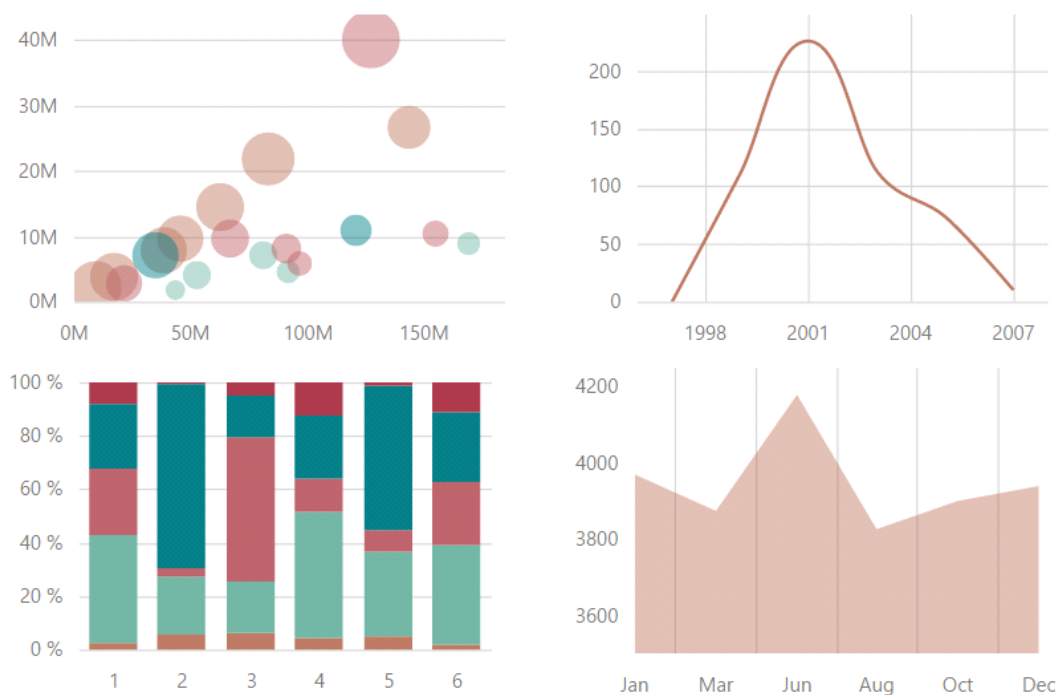


Рис. 1.1. Візуалізація соціальної мережі [3].

Зважені графи можуть використовуватися для ілюстрації хороших і поганих відносин між людьми. Дуги додатної ваги між двома вузлами вказують на позитивні взаємини (дружба, родинні зв'язки, відносини), а дуги від'ємної ваги між двома вузлами вказують на негативні взаємини (ненависть, гнів). Розмічені графи соціальних мереж можна використати для передбачення розвитку графу в майбутньому. У розмічених соціальних мережах існують поняття «збалансованих» і «незбалансованих» циклів. Під збалансованим циклом мають на увазі такий цикл, у якому результат всіх міток додатний. Збалансовані графи подають групу людей, членам якої не хотілося б змінювати свою думку про інших членів групи. Незбалансовані графи подають групу людей, члени якої легко змінюють свою думку про інших членів групи. Наприклад, група з трьох осіб (A, B і C), де A і B мають позитивні взаємини, а C і A негативні, є незбалансованим циклом. Цю групу легко перетворити на збалансований цикл, такий що B буде мати позитивні взаємини з A, і обидва A і B будуть мати негативні взаємини з C. Завдяки використанню збалансованих і незбалансованих циклів розміченого графу

соціальної мережі можна передбачити її розвиток. Приклад зваженого графу продемонстровано на рис. 1.2 [4].

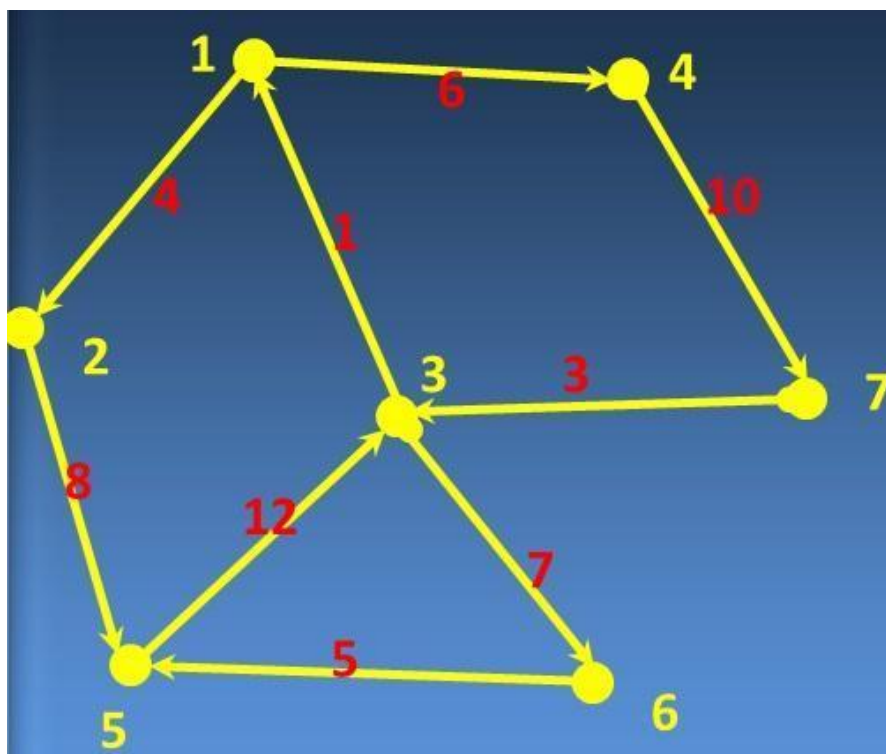


Рис. 1.2. Зважений граф [4].

Особливо коли мова йде про використання аналізу соціальної мережі як інструменту сприяння змінам, визнано корисними різні підходи для відображення мережі. Наприклад, учасники/інтерв'юери надають мережеву інформацію, відображаючи мережу (за допомогою ручки та паперу або цифрових засобів) у процесі збирання даних. Одна з переваг цього підходу та, що він дозволяє досліднику у ході збирання мережевої інформації збирати якісні дані й ставити уточнювальні питання.

#### Машинне навчання

Останнім часом машинне навчання стало дуже популярним для аналізу соціальних мереж. Машинне навчання - це галузь штучного інтелекту, яка дозволяє комп'ютерам навчатися на основі даних, використовуючи різноманітні алгоритми. Для аналізу соціальних мереж можна

використовувати методи машинного навчання, такі як класифікація, кластеризація, прогнозування тощо [5].

У процесі виконання дипломної роботи був створений пакет програмного забезпечення, який може знайти застосування в галузях аналізу та захисту даних соціальних мереж. Він може бути корисним і призначеним для таких груп людей:

1. Дослідники соціальних мереж: проект може бути цікавим для дослідників, які вивчають взаємозв'язки та патерни у соціальних мережах;
2. Компанії з маркетингу: проект може бути корисним для компаній, які займаються маркетингом і рекламою в соціальних мережах. Аналіз зв'язків користувачів може допомогти їм вивчати цільову аудиторію, розуміти їх інтереси та поведінку, і розробляти кращі стратегії реклами та маркетингу;
3. Компанії з кібербезпеки: у світі, де конфіденційні дані є цінними активами, мій проект може бути корисним для компаній з кібербезпеки. Це може допомогти виявляти підозрілу активність, аналізувати зв'язки між користувачами та попереджувати потенційні загрози або ризики безпеки в соціальній мережі.

У порівнянні з наявними рішеннями для аналізу зв'язків у соціальній мережі, розроблений пакет ПЗ має такі переваги:

1. Висока продуктивність: розроблений пакет ПЗ має більш швидшу продуктивність та оптимізацію, порівняно з іншими пакетами ПЗ;
2. Гнучкість та налаштування: користувачі можуть легко налаштувати параметри аналізу, вибрати відображення даних за їхніми вподобаннями.

## РОЗДІЛ 2. ПРОГРАМНИЙ ПАКЕТ ДЛЯ АНАЛІЗУ СОЦІАЛЬНОЇ МЕРЕЖІ

### 2.1. Функціональні можливості програмного забезпечення

Розроблений пакет програмного забезпечення надає користувачам засоби збору, обробки та аналізу даних, які можна використати для різних соціальних мереж.

Збір даних - програма дозволяє користувачам зібрати дані з різних соціальних мереж, таких як Facebook, Twitter, Instagram та інших. Для цього користувачу потрібно ввести свої дані для входу до облікового запису та обрати, які дані він хоче зібрати.

Обробка даних - програма дозволяє користувачам обробляти зібрані дані та виконувати операції, такі як фільтрація, сортування, групування тощо. Крім того, програма має можливість автоматичного визначення та виправлення помилок у даних.

Статистичний аналіз даних - програма дозволяє користувачам виконувати статистичний аналіз зібраних даних, такий як розподіл даних, кореляція, регресійний аналіз, дослідження залежності між змінними тощо.

Візуалізація даних - програма дозволяє користувачам візуалізувати зібрані та оброблені дані за допомогою різних графічних засобів, таких як стовпчикові діаграми, кругові діаграми, лінійні графіки, теплові карти тощо.

### 2.2. Обґрунтування вибору мови та середовища програмування

**Мова програмування Python** має кілька переваг, які роблять її привабливим вибором для розробки програми. Нижче перераховані деякі з них:

1. Легка читабельність: Python має простий і легко читабельний синтаксис, що дозволяє швидко створювати програми і аналізувати дані.

2. Велика спільнота розробників: Python має широку спільноту розробників, що постійно вносять нові пакети і бібліотеки для розв'язання різних завдань.

3. Багатофункціональність: Python має велику кількість бібліотек, які дозволяють працювати зі звичайними форматами даних, такими як CSV та JSON, а також з великими базами даних, такими як MySQL та PostgreSQL.

4. Зручність роботи зі структурованими даними: Python має багато пакетів для роботи зі структурованими даними, такими як Pandas, який дозволяє швидко та ефективно обробляти, очищати та аналізувати дані. (**Pandas** - програмна бібліотека, написана для мови програмування Python для маніпулювання даними та їхнього аналізу. Вона, зокрема, пропонує структури даних та операції для маніпулювання чисельними таблицями та часовими рядами.)

5. Багатоплатформність: Python працює на багатьох операційних системах, таких як Windows, MacOS та Linux, що дозволяє легко переносити програми між різними системами.

**Вибір середовища розробки** для роботи з програмним кодом є важливим етапом у розробці будь-якого проекту, включаючи дипломну роботу. Від середовища розробки залежить ефективність роботи розробника, а також зручність редагування, тестування та налагодження коду.

Один з варіантів вибору середовища розробки для роботи з мовою програмування Python - це IntelliJ IDEA. IntelliJ IDEA є інтегрованою середовищем розробки (IDE), яка надає розробникам зручні інструменти для роботи з кодом. Використання IntelliJ IDEA для розробки проекту на мові Python надає такі переваги:

- Підтримка Python: IntelliJ IDEA має вбудовану підтримку мови програмування Python, що дозволяє вам працювати з Python-кодом з використанням зручного інтерфейсу.

- Зручні інструменти: IntelliJ IDEA надає широкий вибір зручних інструментів, таких як автодоповнення коду, налагоджувач та інші, що допоможуть зробити процес розробки швидшим та ефективнішим.
  - Інтеграція з іншими інструментами: IntelliJ IDEA інтегрується з різноманітними іншими інструментами розробки, такими як системи контролю версій, засоби тестування та інші.
  - Наявна безкоштовна версія для некомерційного використання.
- Недоліками IntelliJ IDEA є:
- Великі вимоги до ресурсів: IntelliJ IDEA вимагає значних обчислювальних ресурсів, що може призвести до повільної роботи комп'ютера.
  - Висока вартість: IntelliJ IDEA може бути вартісним, особливо якщо використовується комерційна ліцензія [6].

### 2.3. Опис роботи програми

Фрагмент коду на рис. 2.1. відповідає за тривимірну візуалізацію даних з соціальної мережі Facebook. Він використовує бібліотеку `igraph` для створення графа мережі і `Plotly` для відображення його в 3D. Потім завантажує список зв'язків між користувачами Facebook. Далі, він будує неорієнтований граф, використовуючи `igraph`, додає до нього вершини та ребра, та присвоює кожній вершині позначення з її іменем. Після цього він обчислює координати кожної вершини та кожного ребра та виконує 3D-відображення в `Plotly`.

```

5 import json
6
7 import igraph as ig
8 import urllib2
9
10 data = []
11 req = urllib2.Request("https://raw.githubusercontent.com/plotly/datasets/master/miserables.json")
12 opener = urllib2.build_opener()
13 f = opener.open(req)
14 data = json.loads(f.read())
15
16 # In[16]:
17
18 L = len(data['links'])
19 Edges = [(data['links'][k]['source'], data['links'][k]['target']) for k in range(L)]
20
21 Gp = ig.Graph(Edges, directed=False)
22
23 # In[19]:
24
25 print((Edges[0]))
26
27 # In[3]:
28
29 labels = []
30 group = []
31
32 for node in data['nodes']:
33     labels.append(node['name'])
34     group.append(node['group'])
35
36 # In[25]:
37
38 from igraph import *
39
40 G = Graph()

```

Рис. 2.1. Приклад коду з файлу 3d\_network\_vis.py

Код на рис. 2.2. містить функції для зчитування та візуалізації графа з даних, які зберігаються у вигляді списку ребер. Граф зберігається в форматі, що використовується бібліотекою `igraph`. Крім цього, з використанням бібліотеки `plotly`, створюється 3D-візуалізація графа.

```

10 from igraph import *
11
12 g = Graph()
13
14
15 # In[3]:
16
17
18 def addVertex(g, name_str):
19     try:
20         if (name_str not in g.vs['name']):
21             print('Inserted node ', name_str)
22             g.add_vertex(name=name_str)
23         else:
24             print('Node ', name_str, ' already present')
25             print(g.vs.find(name_str).index)
26     except KeyError:
27         g.add_vertex(name=name_str)
28     return g
29
30
31 def write_tuple_to_file(f, t):
32     string = str(t[0]) + ' ' + str(t[1]) + '\n'
33     f.write(string)
34
35
36 def retrieve_edge_name_tuple(g, t):
37     a = (g.vs[t[0]]['name'], g.vs[t[1]]['name'])
38     return a

```

Рис. 2.2. Приклад коду з файлу Central.py

Код на рис. 2.3. визначає функції, а потім використовує їх для створення наборів даних для графіка. Графік представлено за допомогою бібліотеки `igraph`.

`load_dataset`: Ця функція приймає графік як вхідні дані та завантажує дані з файлів, що представляють ребра між вузлами.

`addVertex`: ця функція приймає граф і рядок, що представляє назву вузла як вхідні дані. Він додає до графа нову вершину з заданим ім'ям, якщо вершина з такою ж назвою ще не існує.

```

7  from igraph import *
8  def load_dataset(g):
9      fileNums=[0]
10     for i,eachNum in enumerate(fileNums):
11         print(eachNum)
12         fileName="Datasets/facebook/edges/"+str(eachNum)+".edges"
13         print('fileName=', fileName)
14         f=open(fileName, 'a+')
15         nodeID=eachNum
16         line=f.readline()
17         while(line!=''):
18             c=(line.split())
19             g=addVertex(g,c[0])
20             g=addVertex(g,c[1])
21             print('Adding ',c[0], '-->', c[1])
22             g.add_edge(c[0],c[1])
23             line=f.readline()
24     g.simplify()
25     return
26
27 def addVertex(g, name_str):
28     try:
29         if(name_str not in g.vs['name']):
30             print('Inserted node ', name_str)
31             g.add_vertex(name=name_str)
32         else:
33             print('Node ', name_str, ' already present')
34             print(g.vs.find(name_str).index)
35     except KeyError:
36         g.add_vertex(name=name_str)
37     return g

```

Рис. 2.3. Приклад коду з файлу Central.py

Код на рис. 2.4. є реалізацією функцій для зчитування і підготовки даних для задачі класифікації на графах. Код зчитує датасет за допомогою функцій `load_dataset` та `load_neg_dataset` та створює графи `g` та `not_g`, які відповідають позитивним і негативним зв'язком у даних відповідно.

```

7 import numpy as np
8 from igraph import *
9 global num_of_feat
10 num_of_feat=347
11
12
13 # In[3]:
14
15 def load_dataset(fileName,g):
16     fileNums=[0]
17     for i,eachNum in enumerate(fileNums):
18         print(eachNum)
19         print('fileName=', fileName)
20         f=open(fileName)
21         line=f.readline()
22         while(line!=''):
23             c=(line.split())
24             g=addVertex(g,c[0])
25             g=addVertex(g,c[1])
26             print('Adding ',c[0],'-->',c[1])
27             g.add_edge(c[0],c[1])
28             line=f.readline()
29     g.simplify()
30     return
31
32 def load_neg_dataset(fileName,g):
33     fileNums=[0]
34     for i,eachNum in enumerate(fileNums):
35         print(eachNum)
36         print('fileName=', fileName)
37         f=open(fileName)
38         nodeID=eachNum
39         line=f.readline()
40         while(line!=''):

```

Рис. 2.4. Приклад коду з файлу GenOtherDatasets.py

Функція `load_dataset` зчитує інформацію з файлу про зв'язки між вузлами графа, додає нові вершини та ребра до графу `g` за допомогою функції `addVertex`, яка перевіряє, чи існує вузол з іменем `name_str` у графі, та додає його до графу, якщо він ще не існує. Функція `load_neg_dataset` виконує ту ж операцію для графу `not_g`.

## 2.4. Візуалізація зв'язків у соціальній мережі

Програма дозволяє створити 3D візуалізацію вузлів та зв'язків графа на основі даних, отриманих з мережі. Конкретні результати відображення залежать від даних, які завантажуються з мережі та використовуються для побудови графа.

Після цього вона завантажує дані з URL-адреси "<https://raw.githubusercontent.com/plotly/datasets/master/miserables.json>". Тоді вона створює граф `Gp` на основі зв'язків у цих даних, що зображене на рис.

2.5.

```

{"nodes":[{"name":"Myriel","group":1}, {"name":"Napoleon","group":1}, {"name":"Mlle.Baptistine","group":1}, {"name":"Mme.Magloire","group":1}, {"name":"Cravatte","group":1}, {"name":"Count","group":1}, {"name":"OldMan","group":1}, {"name":"Labarre","group":2}, {"name":"Valjean","group":2}, {"name":"Tholomyes","group":3}, {"name":"Listolier","group":3}, {"name":"Fameuil","group":3}, {"name":"Blacheville","group":3}, {"name":"Mme.Thenardier","group":4}, {"name":"Thenardier","group":4}, {"name":"Cosette","group":5}, {"name":"Javert","group":4}, {"name":"Mlle.Baptistine","group":2}, {"name":"Scaufflaire","group":2}, {"name":"Woman1","group":2}, {"name":"Judge","group":2}, {"name":"Champrol","group":2}, {"name":"Boulatruelle","group":6}, {"name":"Eponine","group":4}, {"name":"Anzelma","group":4}, {"name":"Woman2","group":5}, {"name":"Mme.Thenardier","group":8}, {"name":"Gillenormand","group":5}, {"name":"Magnon","group":5}, {"name":"Mlle.Gillenormand","group":5}, {"name":"BaronessT","group":5}, {"name":"Mabeuf","group":8}, {"name":"Enjolras","group":8}, {"name":"Combeferre","group":8}, {"name":"Bossuet","group":8}, {"name":"Joly","group":8}, {"name":"Grantaire","group":8}, {"name":"MotherPlutarch","group":9}, {"name":"Toussaint","group":5}, {"name":"Child1","group":10}, {"name":"Child2","group":10}, {"name":"Brujon","group":4}, {"name":"Mme.Thenardier","group":4}, {"source":3,"target":2,"value":6}, {"source":4,"target":0,"value":1}, {"source":5,"target":0,"value":1}, {"source":6,"target":0,"value":1}, {"source":11,"target":10,"value":1}, {"source":11,"target":3,"value":3}, {"source":11,"target":2,"value":3}, {"source":11,"target":0,"value":1}, {"source":15,"target":11,"value":1}, {"source":17,"target":16,"value":4}, {"source":18,"target":16,"value":4}, {"source":18,"target":0,"value":1}, {"source":20,"target":16,"value":3}, {"source":20,"target":17,"value":3}, {"source":20,"target":18,"value":3}, {"source":20,"target":0,"value":1}, {"source":21,"target":19,"value":3}, {"source":21,"target":20,"value":5}, {"source":22,"target":16,"value":3}, {"source":22,"target":0,"value":1}, {"source":22,"target":21,"value":4}, {"source":23,"target":16,"value":3}, {"source":23,"target":17,"value":3}, {"source":23,"target":0,"value":1}, {"source":23,"target":22,"value":4}, {"source":23,"target":12,"value":2}, {"source":23,"target":11,"value":9}, {"source":24,"target":0,"value":1}, {"source":25,"target":11,"value":12}, {"source":26,"target":24,"value":4}, {"source":26,"target":11,"value":31}, {"source":26,"target":0,"value":1}, {"source":27,"target":25,"value":5}, {"source":27,"target":24,"value":1}, {"source":27,"target":26,"value":1}, {"source":28,"target":0,"value":1}, {"source":29,"target":11,"value":2}, {"source":30,"target":23,"value":1}, {"source":31,"target":30,"value":2}, {"source":31,"target":0,"value":1}, {"source":33,"target":11,"value":2}, {"source":33,"target":27,"value":1}, {"source":34,"target":11,"value":3}, {"source":34,"target":0,"value":1}, {"source":36,"target":34,"value":2}, {"source":36,"target":35,"value":2}, {"source":36,"target":11,"value":2}, {"source":36,"target":0,"value":1}, {"source":37,"target":11,"value":2}, {"source":37,"target":29,"value":1}, {"source":38,"target":34,"value":2}, {"source":38,"target":0,"value":1}, {"source":38,"target":29,"value":1}, {"source":39,"target":25,"value":1}, {"source":40,"target":25,"value":1}, {"source":41,"target":0,"value":1}, {"source":42,"target":24,"value":1}, {"source":43,"target":11,"value":3}, {"source":43,"target":26,"value":1}, {"source":43,"target":0,"value":1}, {"source":47,"target":46,"value":1}, {"source":48,"target":47,"value":2}, {"source":48,"target":25,"value":1}, {"source":48,"target":0,"value":1}, {"source":50,"target":49,"value":1}, {"source":50,"target":24,"value":1}, {"source":51,"target":49,"value":9}, {"source":51,"target":0,"value":1}, {"source":53,"target":51,"value":1}, {"source":54,"target":51,"value":2}, {"source":54,"target":49,"value":1}, {"source":54,"target":0,"value":1}, {"source":55,"target":54,"value":1}, {"source":55,"target":26,"value":21}, {"source":55,"target":11,"value":19}, {"source":55,"target":0,"value":1}, {"source":56,"target":49,"value":1}, {"source":56,"target":55,"value":1}, {"source":57,"target":55,"value":1}, {"source":57,"target":0,"value":1}, {"source":58,"target":27,"value":6}, {"source":58,"target":57,"value":1}, {"source":58,"target":11,"value":4}, {"source":59,"target":0,"value":1}, {"source":60,"target":48,"value":1}, {"source":60,"target":58,"value":4}, {"source":60,"target":59,"value":2}, {"source":61,"target":0,"value":1}, {"source":61,"target":57,"value":1}, {"source":62,"target":55,"value":9}, {"source":62,"target":0,"value":1}, {"source":62,"target":41,"value":1}, {"source":62,"target":61,"value":6}, {"source":62,"target":60,"value":3}, {"source":63,"target":0,"value":1}, {"source":63,"target":58,"value":4}, {"source":63,"target":61,"value":3}, {"source":63,"target":60,"value":2}, {"source":63,"target":0,"value":1}, {"source":64,"target":63,"value":4}, {"source":64,"target":58,"value":10}, {"source":64,"target":61,"value":6}, {"source":64,"target":0,"value":1}, {"source":65,"target":63,"value":5}, {"source":65,"target":64,"value":7}, {"source":65,"target":48,"value":3}, {"source":65,"target":0,"value":1}, {"source":65,"target":59,"value":5}, {"source":65,"target":57,"value":1}, {"source":65,"target":55,"value":2}, {"source":66,"target":0,"value":1}, {"source":66,"target":65,"value":2}, {"source":66,"target":48,"value":1}, {"source":66,"target":63,"value":1}, {"source":66,"target":0,"value":1}, {"source":68,"target":11,"value":1}, {"source":68,"target":24,"value":1}, {"source":68,"target":27,"value":1}, {"source":68,"target":0,"value":1}, {"source":69,"target":11,"value":1}, {"source":69,"target":24,"value":1}, {"source":69,"target":27,"value":2}, {"source":69,"target":0,"value":1}, {"source":70,"target":68,"value":4}, {"source":70,"target":11,"value":1}, {"source":70,"target":24,"value":1}, {"source":70,"target":0,"value":1}, {"source":71,"target":69,"value":2}, {"source":71,"target":68,"value":2}, {"source":71,"target":70,"value":2}, {"source":71,"target":0,"value":1}, {"source":72,"target":26,"value":2}, {"source":72,"target":27,"value":1}, {"source":72,"target":11,"value":1}, {"source":73,"target":0,"value":1}, {"source":75,"target":68,"value":3}, {"source":75,"target":25,"value":3}, {"source":75,"target":48,"value":1}, {"source":75,"target":0,"value":1}, {"source":76,"target":65,"value":1}, {"source":76,"target":66,"value":1}, {"source":76,"target":63,"value":1}, {"source":76,"target":0,"value":1}

```

Рис. 2.5. Зв'язки графа у текстовому вигляді

Далі програма розміщує вузли та зв'язки в 3D просторі за допомогою алгоритму "kk" (layout('kk')). Вузли відображаються як маркери, а зв'язки - як лінії.

Після того програма використовує бібліотеку Plotly, щоб створити 3D візуалізацію. Вона використовує об'єкти Scatter3d для відображення зв'язків та вузлів у тривимірному просторі. Результат відображається у вікні, що відкривається після виклику fig.show().

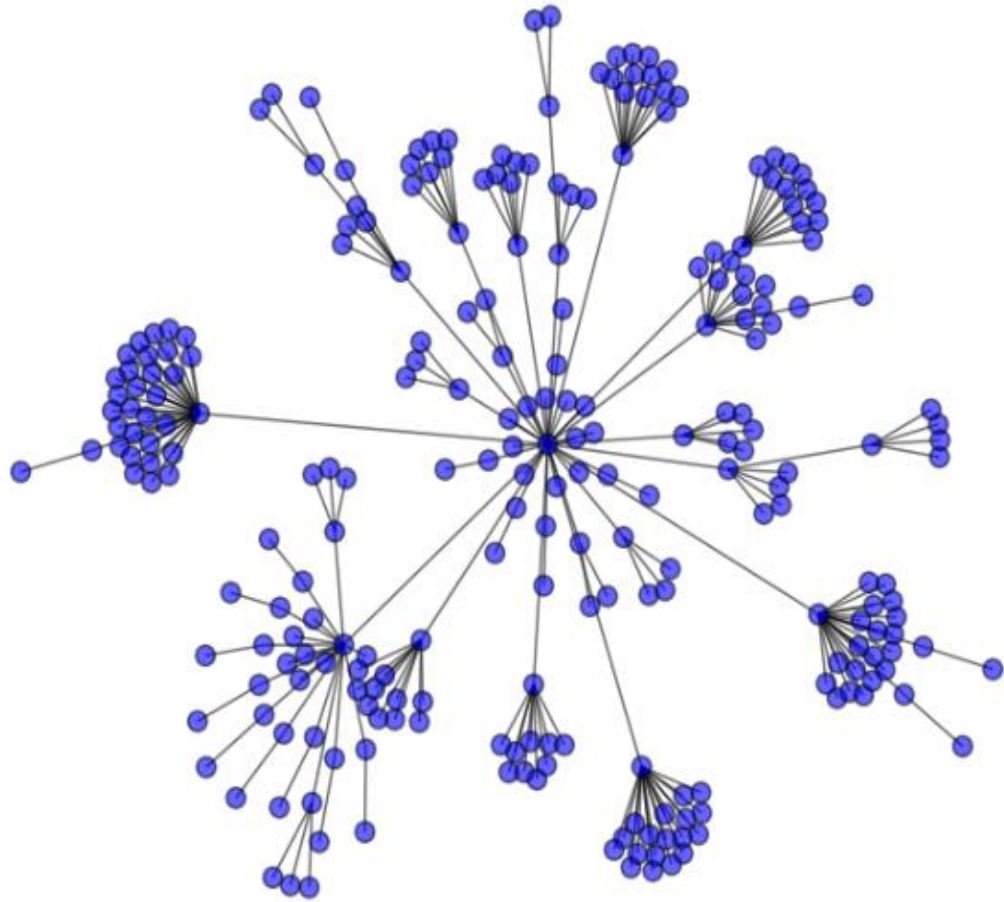


Рис. 2.6. Зв'язки графа у графічному вигляді

Отримана візуалізація (що зображена на рис. 2.6.) є складною та відображає групи вузлів або інші характеристики графа залежно від даних, що використовуються. Візуалізація дозволяє легко виявляти патерни, структуру та організацію соціальної мережі. Це допомагає зрозуміти, як люди або об'єкти взаємодіючи між собою утворюють спільноти та загальну динаміку соціальної мережі.

## 2.5. Подальше вдосконалення ПЗ

Наразі інтерфейс програми знаходиться у стані розробки, також планується розширити функціонал, додавши такі можливості:

1. машинне навчання - програма буде мати можливість використовувати методи машинного навчання для класифікації, кластеризації

та прогнозування результатів. Для цього програма буде мати вбудовані алгоритми машинного навчання, такі як нейронні мережі, рішення дерева та інші.

2. збереження та експорт даних - програма дозволить користувачам зберігати оброблені дані у різних форматах, таких як CSV, Excel, JSON тощо. Крім того, програма дозволить експортувати дані у вигляді графіків та звітів у форматах PDF або HTML. Загалом, програма забезпечуватиме користувачів засобами для повного циклу обробки даних соціальної мережі, починаючи від збору даних та закінчуючи створенням високоякісних графічних відображень та звітів.

## ВИСНОВКИ

Під час виконання дипломної роботи було розроблено пакет програмного забезпечення для аналізу даних соціальної мережі, який забезпечує зручний та ефективний інструмент аналітики, має широкі можливості для обробки та візуалізації даних.

Робота програми перевірялась на основі тестового масиву даних, який імітує набір користувацьких даних у мережі Facebook. Для підтвердження працездатності представлено візуалізацію у тривимірному просторі графу, побудованого за цими даними.

Представлене програмне забезпечення може бути використано у різних галузях, зокрема воно може бути корисне для маркетингових аналітиків для створення персоналізованих рекламних кампаній, оскільки дозволяє вивчати поведінку користувачів на основі їх інтересів, попередніх взаємодій та демографічних характеристик.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction: Stanford, California, USA, 2008. 764 с.
2. Stanley Wasserman. Social Network Analysis: Methods and Applications: University of Illinois, USA, 1994. 356 с.
3. Stephen P. Borgatti, Martin G. Everett, Jeffrey C. Johnson. Analyzing Social Networks: SAGE, Great Britain, 2013. 320 с.
4. Matthew A. Russell. Mining the Social Web, Second Edition: O'Reilly Media, Inc., USA, 2014. 448 с.
5. Reza Zafarani. Social Media Mining an Introduction: University of Cambridge, Great Britain, 2014. 382 с.
6. А. В. Яковенко. Основи програмування. Python. Частина 1 [Електронний ресурс]: підручник для студ. спеціальності 122 "Комп'ютерні науки", спеціалізації "Інформаційні технології в біології та медицині": КПІ ім. Ігоря Сікорського, Київ, 2018. 195 с.