

МЕТОДИ І ПРОГРАМНІ ЗАСОБИ АНАЛІЗУ ПОВЕДІНКИ ГРУП У СОЦІАЛЬНИХ МЕРЕЖАХ

Захарчук Н.Г., науковий керівник Ткаченко О.М.

Соціальні мережі — це структури з вузлів (людей чи об'єктів) та зв'язків між ними. Вони допомагають досліджувати соціальні взаємодії, від дружби до впливу на поведінку[1]. Аналіз поведінки груп вивчає формування, взаємодію та вплив між групами, включаючи поширення інформації та формування норм. Для цього використовуються графи, машинне навчання та штучний інтелект. Практичне застосування охоплює маркетинг, політику, управління брендом, боротьбу з дезінформацією.

Аналіз соціальних мереж дозволяє вивчати структуру взаємодій у суспільстві та корисний для формування теорій соціальної динаміки, таргетування, аналізу споживачів, та ефективного управління політичними ініціативами. Також, це інструмент для розв'язання соціальних проблем та вивчення впливу нових ідей у суспільстві.

В даному дослідженні описуються ключові елементи: вузли, ребра, ступені, типи центральності, щільність мережі, компоненти зв'язності та кластеризація. Ці метрики показують наскільки вузол важливий та як формуються груп.

Розглянемо графові алгоритми, які можуть бути використані:

1. Пошук найкоротших шляхів (алгоритм Дейкстри) — допомагає моделювати розповсюдження інформації.
2. Виявлення спільнот — виявляє групи з тісними зв'язками
3. Алгоритми ранжування (PageRank, HITS) — визначають найвпливовіші вузли.
4. Алгоритми поширення — моделюють вірусний маркетинг та інфо-хвилі.
5. Структурний аналіз — виявляє критичні точки мережі, які можуть порушити її цілісність.

Для дослідження методів аналізу соціальних мереж було обрано тематичний датасет, присвячений повномасштабному вторгненню росії в Україну. Джерелом даних стала Reddit-спільнота “Ukrainian Conflict”[2], яка є активною платформою для обговорення подій, пов'язаних із війною, надання новин, аналітики та коментарів користувачів.

Датасет зібраний автоматизовано за допомогою Python-бібліотеки, а самі дані доступні у відкритому доступі на платформі Kaggle. Такий вибір зумовлений тематичною релевантністю та обсягом даних, відкритістю та потенціалом для аналізу соціальних, емоційних та інформаційних патернів.

Перед проведенням аналізу дані були очищені та структуровані. Зокрема, було видалено дублікати записів за унікальним ідентифікатором id, оскільки під час автоматизованого збору могли повторно фіксуватися ті самі пости або коментарі. Пропущені значення у колонках title і body було заповнено порожніми рядками для уникнення помилок при об'єднанні.

Для подальшого аналізу було створено нову колонку text, яка об'єднує заголовок (title) і основний текст повідомлення (body). Це дозволяє розглядати публікацію як єдину текстову одиницю.

Колонка timestamp була перетворена у формат дати й часу для проведення часових аналізів. Також було витягнуто доменні імена з URL-посилань, що дає змогу аналізувати джерела новин та потенційні канали поширення дезінформації.

Для виявлення основних тем обговорень було проведено тематичний аналіз. Для цього було використано метод тематичного моделювання — Latent Dirichlet Allocation (LDA). Перед його застосуванням тексти були векторизовані з використанням CountVectorizer, а також вилучені англійські стоп-слова.

Додатково були враховані біграми і триграми, щоб зафіксувати стійкі фрази на кшталт “military aid” або “civilian casualties”.

Було задано параметр $n_components = 5$, щоб модель виділила лише п’ять основних тем. Після навчання моделі було проаналізовано топ-10 ключових слів для кожної теми, що дозволило інтерпретувати їх зміст. У результаті були ідентифіковані такі напрями, як військові дії, міжнародна підтримка, гуманітарна криза, політичні заяви та емоційно насичені новини.

Наступним етапом став аналіз емоційного тону повідомлень. Для цього було використано бібліотеку TextBlob, яка обчислює полярність тексту в межах від -1 (негативне) до +1 (позитивне). Було відібрано підмножину з 200 000 постів для прикладного аналізу.

Результати показали, що більшість повідомлень мають нейтральний тон — приблизно 56,6%. Позитивними класифіковано близько 29,5%, а негативними — 13,9%. Було побудовано часову діаграму для виявлення динаміки емоцій у часі, яка показала наявність піків емоційної активності, що часто збігалися з ключовими подіями, такими як військові наступи, гуманітарна допомога або резонансні новини.

Наступним етапом був аналіз числової динаміки публікацій, цей етап був присвячений дослідженню змін у кількості публікацій у часі. Спочатку було виконано агрегацію кількості постів за кожен день, після чого результати були візуалізовані у вигляді стовпчикowego графіка. Це дало змогу ідентифікувати дні з найвищою інформаційною активністю.

Додатково було виділено 10 днів із найбільшою кількістю публікацій. Для кожного з них проведено змістовий аналіз — визначено найпопулярніші заголовки, які найчастіше повторювались у цей період. Наприклад, у дні 23–24 березня 2022 року активно обговорювались теми втрат російської армії, гуманітарної допомоги та знищення ворожих об’єктів.

На наступному етапі було проведено поглиблений аналіз лексики позитивних та негативних повідомлень. Зокрема, створено “хмари слів” для кожної з груп. У позитивних постах домінували слова на кшталт support, freedom, victory, що свідчить про акцент на підтримку та волонтерську діяльність. У негативних — war, killed, missile, що вказує на теми втрат, насильства та тривоги.

Також було здійснено трендовий аналіз частоти згадуваних слів у часовому розрізі. Обрано ключові слова. Аналіз показав, що імена лідерів згадуються стабільно, а слова по типу Bucha мало чіткий пік після оприлюднення злочинів у цьому місті. Таким чином, виявлено зв’язок між інформаційною хвилею та подіями, що її спричинили.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аналіз соціальних мереж [Електронний ресурс] // Вікіпедія. – Режим доступу: https://uk.wikipedia.org/wiki/Аналіз_соціальних_мереж.
2. Огляд моделей аналізу соціальних мереж [Електронний ресурс] // ResearchGate. – Режим доступу: https://www.researchgate.net/publication/279535422_OGLAD_MODELEJ_ANALIZU_SOCIALNIH_MEREZ.
3. Russian Invasion of Ukraine [Електронний ресурс] // Kaggle. — Режим доступу: <https://www.kaggle.com/datasets/gpreda/russian-invasion-of-ukraine>.

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**



ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РОЗРОБКИ КОМП'ЮТЕРНИХ СИСТЕМ

**ЗБІРНИК НАУКОВИХ ПРАЦЬ ЗА МАТЕРІАЛАМИ
VII ВСЕУКРАЇНСЬКОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ
СТУДЕНТІВ І АСПІРАНТІВ
*24 квітня 2025 року***

Київ 2025

УДК 004

Відповідальна за випуск: М.І. Лендел

Збірник наукових праць за матеріалами VII Всеукраїнської науково-практичної конференції студентів і аспірантів «ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РОЗРОБКИ КОМП'ЮТЕРНИХ СИСТЕМ '2025», 24 квітня 2025 року, НУБіП України, Київ. – 452 с. (електронне видання)

Відповідальність за зміст публікацій несуть автори.

Передрук матеріалів, а також використання їх будь-якій формі допускається лише з дозволу авторів