

Дослідження методів аналізу соціальних мереж як середовища інформаційних війн

Войтович О. П., Буда А. Г., Головенько В. О.
кафедра захисту інформації
Вінницький національний технічний університет, Вінниця, Україна
voytovych.olesya@vntu.edu.ua,
antbuda@vntu.edu.ua
torvald124@gmail.com

Analyses methods research of social networks as a environment of warfare

O. Voitovych, A. Buda, V. Holovenko
Department of information security
Vinnytsia National Technical University
Vinnytsia, Ukraine
voytovych.olesya@vntu.edu.ua,
antbuda@vntu.edu.ua
torvald124@gmail.com

Анотація – Соціальні мережі все частіше використовуються як джерело інформації, включаючи інформацію, що стосується глобальних світових подій. У цьому дослідженні розглядаються соціальні мережі як джерела великих даних для отримання та аналізу даних про різноманітні події. Проводиться аналіз підходів та методів дослідження соціальних мереж, розглядаються основні метрики, які використовуються в якості вхідних даних для дослідження. Проаналізовані відомі методи дослідження соціальних мереж, наведені їх переваги та недоліки. Наведено результати експерименту аналізу реакції користувачів соціальних мереж Facebook та Twitter на основі другого туру дебатів кандидатів у президенти США.

Ключові слова—соціальні мережі; інформаційні війни; методи аналізу соціальних мереж; метрики соціальних мереж; структура соціальних мереж

Abstract – Social networks are increasingly being used as a source of information, including information of global events. This study examines social networks as a source of large data for obtaining and analyzing data on a variety of events. An analysis of social network approaches and methods is under consideration, and the main metrics that are used as inputs for research are considered. The well-known methods of researching social networks are analyzed, their advantages and disadvantages are presented. The experiment results of the analysis of social networks Facebook and Twitter user's reaction on the second round of debates of US presidential candidates are presented.

Keywords—social networks; information warfares; methods of social network analysis; social network metrics; social networks structure

ВСТУП

Соціальні мережі отримали широке розповсюдження за останній час, в першу чергу як інструменту спілкування, обміну думками та отримання інформації.

В теперішній час постійного вдосконалення інформаційних та комунікативних технологій, будь-який конфлікт має своє відображення в мережі Інтернет. Дуже часто таке мережеве відображення впливає на результат протиборства конкуруючих сторін. Активне залучення до віртуальної мережі багатомільйонної аудиторії дозволяє маніпулювати громадською думкою і суттєво впливати на процеси протиборчих сторін [1].

Розуміння значення інтернету та соціальних мереж змушує передові держави вкладати величезні інвестиції в соціальні мережі, які нині стають не тільки засобом комунікації, а й ефективною політичною зброєю.

Сучасні соціальні мережі істотно змінили постановку питання – сьогодні у дослідників є «безкоштовний» ресурс для досліджень, а стрімке розповсюдження соціальних онлайн-сервісів та розвиток технології великих даних ініціювали інтерес до використання інформації з соціальних мереж в різних галузях [2]. Спільне використання структурних і контентних даних потенційно дозволяє застосовувати соціальні мережі для вирішення широкого кола задач, в тому числі для дослідження та маніпулювання громадською думкою, особливо під час ведення інформаційних війн.

Активне використання соціальних мереж дозволяє оперативно впливати на думку і поведінку людей, що перетворює їх на вогнище битв груп різних інтересів.

Відсутність цензури і різного роду перешкод створює сприятливу основу успішним діям у віртуальному просторі. Все це вказує на те, що соціальні мережі перетворюються на поле арени інформаційного протиборства.

АНАЛІЗ СОЦІАЛЬНИХ МЕРЕЖ

Соціальна мережа – це соціальна структура, що складається з множини агентів (суб'єктів – індивідів, спільнот, груп індивідів чи організацій) і визначеного на ній множині відношень (сукупності зв'язків між агентами, таких як знайомства, дружби, спілкування тощо) [3]. Соціальна мережа являє собою граф $G(N, E)$, де $N = \{1, 2, \dots, n\}$ – кінцева множина вершин (агентів) та E – множина ребер, що відображає взаємодію агентів.

Основними завданнями інформаційно-аналітичної підтримки пасивної і активної роботи з соціальними мережами є моніторинг та аналіз соціальних мереж, прогнозування та управління [4].

- Моніторинг включає отримання та структурування первинних даних. Збираються тексти повідомлень, зв'язки між користувачами, посилання на зовнішні ресурси [5].
- Прогноз можливий після ідентифікації математичної моделі інформаційного процесу. Можуть використовуватися статистичні моделі та моделі динамічних процесів на графах [5].
- Управління полягає в наданні цілеспрямованих впливів на соціальну мережу для перекладу інформаційних процесів в бажаний стан [5].
- Завдання аналізу, прогнозування та управління можуть бути різними, в першу чергу, в залежності від того, хто ставить завдання, тобто хто є кінцевим користувачем системи [5].

В аналізі соціальних мереж можна виділити чотири основних напрямки дослідження: структурне, ресурсне, нормативне та динамічне [6]. В кожному з них вирішується доволі велике коло задач і застосовуються методи і різноманітні галузі знань. Вказані підходи дозволяють виконати структурний аналіз та аналіз поведінки зв'язків, статистичний аналіз соціальних мереж з розглядом їх можливого масштабування, аналіз змісту соціальних мереж (аналіз тексту, загальної інформації з довільними типами даних, а також медіаданих), визначення спільнот в рамках соціальних мереж. Рішення останніх двох задач дозволяє визначити регіони мережі, всередині яких відбувається найбільш активна в заданих рамках

взаємодія учасників соціальної мережі. Розгляд змісту соціальних мереж передбачає виділення та аналіз таких видів контенту мережі: загальної інформації з довільними типами даних; текстової інформації; посилань; вкладень [4].

У структурному підході всі учасники мережі розглядаються як вершини графа, які впливають на конфігурацію ребер та інших учасників мережі. Основна увага приділяється геометричній формі мережі та інтенсивності взаємодій (вазі ребер), тому досліджуються такі характеристики як взаємне розміщення вершин, центральність, транзитивність взаємодій.

При структурному аналізі та аналізі поведінки зв'язків використовуються методи статистичного аналізу, визначення спільнот, алгоритми класифікації. Вивчається поведінка вершин в процесі кластеризації і типових тимчасових характеристик соціальних мереж. Наприклад, як змінюються поведінка і розподіл зв'язкових компонентів графа.

Велике значення надається визначенню спільнот в соціальних мережах. Мета - спробувати визначити регіони мережі, всередині яких відбувається активна взаємодія учасників. Алгоритмічно це завдання можна віднести до задачі про розподіл графів. Необхідно розділити мережу на щільні регіони на основі поведінки зв'язків між вершинами. Комп'ютерні соціальні мережі динамічні, що призводить до ускладнень з точки зору виявлення спільнот. У деяких випадках вдається інтегрувати інформаційний вміст мережі в процес визначення спільнот. Тоді контент є допоміжним засобом для виявлення груп учасників зі схожими інтересами.

Найвідомішими серед соціальних мереж є Facebook, Twitter, LinkedIn, Pinterest, Google+, Tumblr, Instagram, Telegram, Flickr, MySpace тощо. Найбільша кількість користувачів знаходиться у соціальній мережі Facebook і дорівнює понад 2 млрд. користувачів. Така кількість користувачів дозволяє проводити дослідження соціальної думки населення на певні події, відслідковувати вподобання користувачів та навіть проводити інформаційні війни.

МЕТРИКИ ДЛЯ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ

Для формування метрик для аналізу даних в соціальних мережах, можна використовувати інформацію про лайки, коментарі, репости до публікацій, ідентифікаційні дані користувача (стать, вік, місто проживання, місце роботи/навчання, адреса тощо), інформація про друзів користувача, його спільноти, статистика перебування користувача в мережі, зв'язки між користувачами тощо [7].

Існують такі типи метрик соціальних мереж:

- на рівні груп – щільність, компоненти, кліки, централізація, фракції, ядро-периферія;
- на рівні вузлів – центральність (залежності, зв'язки, відношення);
- на рівні візуалізації – графі;
- статистичні метрики – аналіз кластерів даних.

Також до числа метрик також належить ріст числа підписників/друзів, лайки або дизлайки до постів/повідомлень, демографічні показники аудиторії, зв'язки між користувачами, часові показники, відповіді та коментарі до постів/повідомлень, чим поділилися користувачі, перехідний трафік із соціальних медіа, частота кліків, негативний фідбек (скільки підписників/друзів було втрачено, до якого контенту було втрачено доступ) тощо [8].

Важливим завданням є прогноз формування зв'язків в соціальних мережах. У більшості додатків для аналізу соціальних мереж зв'язку вважаються динамічними і можуть змінюватися з плином часу. В процес прогнозування зв'язків можуть бути залучені як структура мережі, так і інформація про особливості різних вершин. Для вирішення таких завдань пропонується будувати різноманітні структурні і реляційні моделі.

МЕТОДИ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ

Ресурсний підхід [6] розглядає можливості учасників по залученню індивідуальних і мережевих ресурсів для досягнення певних цілей і диференціює учасників, які перебувають в ідентичних структурних позиціях соціальної мережі, за їх ресурсами. У якості індивідуальних ресурсів можуть виступати знання, престиж, багатство, раса, стать. Під мережевими ресурсами розуміються вплив, статус, обсяг і характер інформації. Основним показником, який визначає відмінності в ресурсах учасників мережі, є сила структурної позиції учасника. Важливе завдання ресурсного підходу - аналіз змісту соціальних мереж. Мережевий контент служить джерелом для широкого спектру додатків, орієнтованих на вилучення та аналіз даних. Використання змісту мережі допомагає значно поліпшити якість висновків при аналізі соціальних мереж, наприклад, в задачах кластеризації та класифікації.

Аналіз надання переваги [6] полягає у збиранні інформації та створенні індивідуальних анкет, у яких міститься інформація про порівняння двох членів спільноти третім членом за заданим критерієм. У результаті порівняння у результаті отримується тримісне відношення, якому відповідає булева матриця. Після цього інформація обробляється за алгоритмом перетворення булевих матриць до зважених графів. Структура таких графів відображає структуру колективу та може

аналізуватися шар за шаром, залежно від ваги зв'язків.

Ідентифікація користувачів у різних соціальних мережах [9] полягає у виявленні всіх профілів, які представляють конкретного користувача у соціальних мережах. Вихідними даними для пошуку можуть слугувати паспортні дані, назва компанії, де працює користувач, номер телефону, адреса пошти, місце навчання, список друзів тощо. Найпростішим способом ідентифікації є пошук за точним співпадінням усіх відомих характеристик користувача.

Метод аналізу пропаганди SCAME та контрпропаганди [6]. SCAME – аббревіатура з таких слів: Source (джерело), Content (зміст), Audience (аудиторія), Media (медіа), Effect (ефект).

За допомогою лінгвістичних методів можна визначити наскільки часто у текстовій інформації, зчитаній з облікового запису користувача, зустрічаються певні терміни, і, при досягненні певного рівня їх появи можна зробити відповідні висновки. Для такого методу необхідно знати термінологію, отримати яку можна зі словників чи тезаурусів за конкретною предметною областю. Крім того, необхідно розуміти різноманітні форми одного й того ж слова.

Алгоритми машинного навчання розраховані на структуровані дані, тому перед їх застосуванням тексти замінують на набори слів, що зустрічаються в них, або на набір тематик, що характеризують ці тексти. Для цього використовуються лінгвістичні алгоритми виділення значущих слів, їх нормалізація, створення лексичного профілю тексту, виявлення тематики тощо.

Пошук спільнот користувачів [9] працює на основі соціальних зв'язків між користувачами. Основою алгоритму є процес обміну мітками спільнот між вершинами у відповідності з динамічними правилами взаємодії. Додатковим кроком алгоритму є визначення спільнот з недостатнім внутрішнім зв'язком та розділення їх на більш зв'язані підспільноти.

За допомогою методу візуалізації відповідних графів [10] зв'язки формують граф. Математичний апарат аналізу графів дозволяє розраховувати цілий ряд параметрів та дати кількісні відповіді на запити. Шлях між вершинами – це послідовність вершин та ребер, що з'єднують дві вершини. Відстань між вершинами – це кількість кроків, які потрібно зробити, щоб дістатися від однієї вершини до іншої.

Для методу розрахунку індексів для соціальної мережі вцілому [10] використовують елементарні параметри, такі як число вершин чи ребер, та деякі параметри, такі як щільність, число асиметричних діад, число транзитивних діад, геодезійна відстань, діаметр соціальної мережі, структурні дірки тощо.

Виділення підструктур мережі [10] дозволяє виявити ті чи інші структури, приховані в соціальній мережі. Виділення блоків можна робити за різними атрибутами агентів мережі. У блоки також можна об'єднувати еквівалентних агентів мережі.

Використання скрепінгу [11] полягає у використанні інтерфейсу програмування додатків (API). Використовуючи API, дослідник не зобов'язаний мати справу з HTML, а замість цього він може отримати дані у вигляді посилань.

Існують такі статистичні методи: підрахунок кількості лайків, репостів, підписників та інших кількісних характеристик агентів соціальних мереж та їх публікацій; підрахунок згадувань ключових слів з подальшим групуванням [12].

Підходи, що застосовуються для аналізу тональності текстової інформації поділяються на дві основні категорії: інженерно-лінгвістичні методи та методи на основі машинного навчання [12]. Інженерно-лінгвістичні методи використовують спеціальні, попередньо підготовлені тональні словники і/або лінгвістичні правила, на основі яких відбувається аналіз текстових фрагментів. Методи машинного навчання включають в себе метод умовних випадкових полів, штучні нейронні мережі, та інше. Ця група методів використовує математичні моделі, що дозволяють автоматично визначати оптимальний набір параметрів для визначення тональності.

Класифікація використовує у якості вхідних даних тексти повідомлень та поля профілю довільного користувача. Виконується алгоритм класифікації для заданої мови та атрибута. Результатом є значення атрибута обраного користувача. Кластеризація полягає у розбитті соціальної мережі на підмножини, що не перетинаються між собою так, щоб кожен кластер складався з подібних об'єктів, а об'єкти різних кластерів істотно відрізнялися [12]. Виділення кластерів можна здійснювати за різними атрибутами агентів мережі, наприклад за статтю. У кластери можна об'єднувати еквівалентних агентів мережі.

Результати порівняння різних методів та підходів показано в таблиці. Основні переваги та недоліки зазначених методів зведені у таблицю.

Існують декілька підходів по виділенню підструктур у соціальній мережі:

- визначає кліки (підгрупи, в яких агенти пов'язані між собою сильніше, ніж з членами інших клік) в соціальній мережі;
- виділення компонентів (частин графа), які пов'язані всередині і не пов'язані між собою;

ТАБЛИЦЯ – РЕЗУЛЬТАТ ПОРІВНЯННЯ МЕТОДІВ ТА ПІДХОДІВ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ

Метод	Підхід	Переваги	Недоліки
Аналіз надання переваги	На основі графів	Відносно невеликі апаратні затрати для дослідження	Напередзаний аналіз даних
Ідентифікація користувача в різних соціальних мережах	Лінгвістичний	Точність ідентифікації користувача за умови наявності всіх необхідних даних	Не завжди можна чітко ідентифікувати користувача через відсутність необхідних даних
Аналіз пропаганди SCAME та контрпропаганда	Лінгвістичний, нейромережі	Якісна аналітична підготовка, детальна розробленість всіх етапів породження повідомлення	Необхідні великі обчислювальні ресурси для точного результату
Лінгвістичний	Лінгвістич.	Зручність в оперуванні даними	Низька точність результату за умови більше ніж однієї оцінки тональності тексту
Машинне навчання	Лінгвістич., нейромережі	Концентрація уваги на обчисленнях	Висока обчислювальна складність, нечітка апріорна величина
Пошук спільнот користувачів	На основі графів	Імітація людського спілкування між парами індивідумів	Значна обчислювальна складність, нездатність знаходити зв'язки спільнот
Візуалізація відповідних графів	На основі графів, статистичний	Істотна практична потужність, простота у використанні	Складність побудови графів
Розрахунок індексів для соціальної мережі в цілому	На основі графів, статистичний	Можливість розрахувати цілий ряд параметрів	Невеликий діапазон того, що можна було б досліджувати
Виділення підструктур мережі	Сегментація мережі	Дозволяє виявити структури, приховані у соціальній мережі	Дозволяє проводити лише кількісний аналіз соціальної мережі
Скрепінг	З використанням API	Метод підходить для будь-якої HTML-сторінки, можливість використання методу без застосування API	Перевантаження каналу, часові затрати

- знаходження блоків і перемичок. Вершина називається перемичкою, якщо після її видалення граф розпадається на блоки.;
- виділення груп – груп еквівалентних агентів мережі, які мають схожі профілі зв'язків.

ЕКСПЕРИМЕНТ

Було проведено дослідження зміни думки користувачів у соціальних мережах Twitter та Facebook [13]. Для дослідження було обрано подію – дебати кандидатів у президенти США Дональда Трампа і Гіллари Клінтон, що відбулися 10 жовтня 2016 року. Для зчитування та аналізу даних з соціальних мереж було розроблено функції мовою статистичної обробки даних R. Для цього було обрано лінгвістичний підхід, що полягає в аналізі коментарів і постів користувачів та виділенні ключових слів, які вказують на реакцію людей. У якості ключових було обрано наступні слова: good, bad, terrible, beautiful, worst, best, normal.

Отримані результати вказали на різку зміну ставлення користувачів до кандидатів після проведених дебатів та збігаються з результатами аналітики провідного журналу The Telegraph [14].

Результати показали, що думка населення у соціальних мережах Twitter та Facebook була досить суперечливою та неоднозначною у наданні переваги тому чи іншому кандидату у президенти, що свідчить про інформаційне протиборство у соціальних мережах. Як відомо, пізніше було виявлено, що у соціальних мережах Росія змогла вплинути на думку населення США шляхом розповсюдження своїх постів, коментарів та повідомлень [15]. Таким чином, соціальні мережі чітко відображають реакцію користувачів на певну подію, а також є чудовим середовищем для проведення інформаційних війн та пропаганди.

ВИСНОВКИ

У даній статті представлено загальний огляд соціальних мереж, їх структуру та компоненти. Виділено основні завдання для використання соціальних мереж, а саме спілкування користувачів, дослідження та інформаційні війни. Розглянуто найвідоміші та найбільші соціальні мережі світу.

Було розглянуто та проаналізовано основні підходи та їх методи для аналізу соціальних мереж, визначено їх загальну суть, виділено їх основні переваги та недоліки. Розглянуто основні метрики, які використовуються у якості вхідних даних для подальшого аналізу.

Було виділено основні метрики соціальних мереж та їх типи, за допомогою яких проводять аналіз соціальних мереж.

Було проведено експеримент, за допомогою якого досліджувалась реакція людей в соціальних мережах Facebook і Twitter на другий тур дебатів кандидатів у президенти США Гіллари Клінтон та Дональда Трампа, що відбулися 10 жовтня 2016 року. Результат дослідження показав зміну ставлення користувачів до кандидатів після проведення дебатів. Результат експерименту збігається з аналітикою видання The Telegram. Для експерименту було використано лінгвістичний підхід. Дослідження зайняло багато часу та ресурсів через різні фактори, такі як різний доступ до інформації, різна потужність серверів соціальних мереж, різні методи зберігання великих даних.

ЛІТЕРАТУРА REFERENCES

- [1] А. В. Дудатьев, О. П. Войтович, Ю. В. Барішев, “Метод оцінювання безпеки інформаційних ресурсів підприємства на основі аналізу вразливостей,” Вісник Хмельницького національного університету, №4, С. 78-83, 2008. 2
- [2] А.В. Дудатьев, В.А. Лужецкий, Д.А. Коротаев, “Метод оценки информационной устойчивости социотехнических систем в условиях информационной войны,” Восточно-Европейский журнал передовых технологий, № 2(2), С. 4-11, 2016. 3
- [3] Д.А. Губанов, Д.А. Новиков, А.Г. Чхартишвили, Социальные сети: модели информационного влияния, управления и противоборства. Москва: Физматлит, 2010. 8
- [4] С.Ф. Чалый, А.А. Чередниченко, “Исследование методов анализа социальных сетей для определения групп пользователей программного продукта,” 2013. 4
- [5] Н.И. Базенков, Д.А. Губанов, “Обзор информационных систем анализа социальных сетей”. 5
- [6] Т. В. Батура, Н. С. Копылова, Ф. А. Мурзин, А. В. Проскураков, “Методы анализа данных из социальных сетей,” 2013. 6
- [7] О.Горчинская, “Анализ данных социальных сетей”, 2015. 12
- [8] Aaron Aguis, “10 Metrics to Track for Social Media Success”, 2016. 15
- [9] А.Коршунов, И.Белобородов, Н.Бузун, “Анализ социальных сетей: методы и приложения”, 2014. 11
- [10] А. Прохоров, “Социальные сети и Интернет,” 2006. 9
- [11] Берни Хоган, “Анализ социальных сетей в интернете”, 2013. 10
- [12] Д.В. Будыльский, “Автоматизация мониторинга общественного мнения на основе интеллектуального анализа сообщений в социальных сетях,” 2015. 7
- [13] O. Voitovych, V. Holovenko, “Research of social networks as a source of information in warfare”, 2016. 1
- [14] The Telegraph, “Who won the second presidential TV debates”, The Telegraph, 2016. 13
- [15] G. Korte, ‘Election to obstruction: The many tentacles of the Trump-Russia investigation’, USA Today, 2017.