МАГІСТЕРСЬКА РОБОТА

МР. ІПм - 41.00.00.000 ПЗ

Група ІПм-24-2

Новальківський Роман

2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Новальківський Роман Русланович

(прізвище, ім'я, по батькові)

УДК 004.9

(індекс)

МАГІСТЕРСЬКА РОБОТА

Застосування "data-driven" методів та засобів при моделюванні процесів в соціальних мережах

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Новальківський Р.Р.

(підпис, ініціали та прізвище здобувача освітнього ступеня)

Науковий керівник Михайлюк Ірина Романівна, к.п.н., доцент

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

Допущено до захисту

Завідувач кафедри

доц. Бандура В.В.

(посада) (підпис) (дата) (ініціали та прізвище)

Нормоконтроль

доц. Вовк Р.Б.

(посада) (підпис) (дата) (ініціали та прізвище)

Робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Івано-Франківськ – 2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Освітній рівень магістр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:
Зав. кафедрою **ІПЗ**

доц. В.В. Бандура

“ 04 ” вересня 2025 р.

ЗАВДАННЯ

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Новальківському Роману Руслановичу

(прізвище, ім’я, по-батькові)

1. Тема магістерської роботи “ Застосування "data-driven" методів та засобів при моделюванні процесів в соціальних мережах ”

керівник проекту (роботи) Михайлюк І.Р., к.п.н., доцент

затверджені наказом закладу вищої освіти від “ 05 ” листопада 2025 р. № 695/7

2. Строк подання студентом проекту (роботи) 15 грудня 2025 р.

3. Вихідні дані до проекту (роботи) Концепції та формальні моделі і методи побудови інформаційних технологій моделювання процесів в соціальних мережах

4. Зміст розрахунково - пояснювальної записки(перелік питань, які потрібно розробити)

1. Аналіз предметної області моделювання процесів приватності в соціальних мережах
2. Моделювання процесів приватності в соціальних мережах на основі оцінки вразливості графів
3. Фреймворк для оцінки вразливості графів до атак деанонімізації
4. Моделювання активності в соціальних мережах з урахуванням екзогенних факторів

5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)

1. Огляд структури системи використання безпечних графових даних SecGraph (рис. 1.1)
2. Фреймворк обміну інформацією (рис. 1.2)
3. Фреймворк для вимірювання приватності та корисності (рис. 2.1)
4. Порівняння сили атаки залежно від різних варіантів вибору перекриття (рис. 2.2)
5. Транзитивність (C) та асортативність (r) на графах моделі Leader-Follower (рис. 2.3)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата
Перевірка на плагіат	доц., к.т.н. Вовк Р.Б.	

7. Дата видачі завдання 04 вересня 2025 р.

Керівник _____
(підпис)

Завдання прийняв до виконання_____
(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назви етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1	Підбір і вивчення літератури по темі магістерської роботи	15.09.2025	виконано
2	Аналіз предметної області моделювання процесів приватності в соціальних мережах	01.10.2025	виконано
3	Моделювання процесів приватності в соціальних мережах на основі оцінки вразливості графів	17.10.2025	виконано
4	Фреймворк для оцінки вразливості графів до атак деанонімізації	02.11.2025	виконано
5	Моделювання активності в соціальних мережах з урахуванням екзогенних факторів	19.11.2025	виконано
6	Представлення узагальненої архітектури фреймворку та потоку даних представлення активності в соціальній мережі	02.12.2025	виконано
7	Затвердження пояснювальної записки роботи завідувачем кафедри	15.12.2025	виконано

Студент – магістр _____
(підпис)

Керівник роботи _____
(підпис)

АНОТАЦІЯ

Магістерська робота: 79 с., 19 рис., 3 табл., 45 джерел.

Тема: Застосування "data-driven" методів та засобів при моделюванні процесів в соціальних мережах

Метою роботи є підвищення ефективності моделювання процесів у соціальних мережах шляхом розробки та застосування data-driven методів для симуляції динаміки активності користувачів під впливом екзогенних факторів.

Об'єкт дослідження: процеси інформаційної взаємодії, функціонування даних та поширення інформаційних потоків у соціальних мережах.

Предмет дослідження: методи та моделі оцінки вразливості структур графів до деанонімізації, а також методи симуляції динаміки активності користувачів із застосуванням data-driven підходів.

Результати дослідження

В роботі проведено моделювання активності в соціальних мережах через розробку модульної архітектури симулятора, яка інтегрує модуль обробки екзогенних факторів (зовнішніх подій) у процес генерації дерев ретвітів, що підвищує точність прогнозування інформаційних каскадів

Висновок

Розроблено модульну архітектуру симулятора процесів у соціальних мережах, яка, на відміну від класичних моделей поширення інформації, враховує вплив зовнішніх (екзогенних) подій.

СОЦІАЛЬНІ МЕРЕЖІ, DATA-DRIVEN МЕТОДИ, МАШИННЕ НАВЧАННЯ, ПРИВАТНІСТЬ, ДЕАНОНІМІЗАЦІЯ, ГРАФОВІ СТРУКТУРИ, ЕКЗОГЕННІ ФАКТОРИ, СИМУЛЯЦІЯ АКТИВНОСТІ, ІНФОРМАЦІЙНІ КАСКАДИ

ABSTRACT

Master Thesis: 79 pp., 19 fig., 3 tab., 45 sources.

Topic: Application of "data-driven" methods and tools in modeling processes in social networks

The aim of the work is to increase the efficiency of modeling processes in social networks by developing and applying data-driven methods for simulating the dynamics of user activity under the influence of exogenous factors.

Object of research: processes of information interaction, data functioning and dissemination of information flows in social networks.

Subject of research: methods and models for assessing the vulnerability of graph structures to deanonymization, as well as methods for simulating the dynamics of user activity using data-driven approaches.

Research results

The work simulates activity in social networks through the development of a modular simulator architecture that integrates a module for processing exogenous factors (external events) into the process of generating retweet trees, which increases the accuracy of predicting information cascades

Conclusion

A modular architecture for a simulator of processes in social networks has been developed, which, unlike classical models of information dissemination, takes into account the influence of external (exogenous) events.

SOCIAL NETWORKS, DATA-DRIVEN METHODS, MACHINE LEARNING, PRIVACY, DE-ANONYMIZATION, GRAPH STRUCTURES, EXOGENEOUS FACTORS, ACTIVITY SIMULATION, INFORMATION CASCADES

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ПСМ - Платформа соціальної мережі

NRMSE - Normalized Root Mean Square Error

SMAPE - Symmetric Mean Absolute Percentage Error

EM - Earth Movers (Distance)

BERT - Bidirectional Encoder Representations from Transformers

GNIP - Data API Tool/Service

GDELT - Global Database of Events, Language, and Tone

MCAS - Modular Cascade Activity Simulator

GT-Seed - Ground Truth Seed

ВСТУП

Актуальність теми.

Сучасний етап розвитку інформаційного суспільства характеризується стрімким зростанням обсягів даних, що генеруються користувачами у соціальних мережах. Соціальні графи стали не лише середовищем для комунікації, але й потужним джерелом інформації для бізнес-аналітики, соціологічних досліджень та систем прийняття рішень. Проте, ця відкритість породжує дві фундаментальні проблеми: загрозу приватності користувачів та складність прогнозування динаміки інформаційних потоків в умовах постійного впливу зовнішніх інформаційних подій.

Актуальність дослідження зумовлена необхідністю розв’язання протиріччя між потребою в аналізі великих даних (Big Data) соціальних мереж та необхідністю захисту конфіденційності осіб, чиї дані формують ці мережі. Традиційні методи анонімізації виявляються недостатньо ефективними проти сучасних атак деанонімізації, що використовують алгоритми машинного навчання. Зловмисники здатні відновлювати особистість користувачів, аналізуючи топологічні структури графів, що вимагає розробки нових, «data-driven» підходів до оцінки вразливості мереж.

Водночас, моделювання активності в соціальних мережах часто обмежується аналізом внутрішніх взаємодій (ендогенних факторів), ігноруючи вплив зовнішніх новинних приводів (екзогенних факторів). Розробка моделей, здатних симулювати реакцію мережі на зовнішні події, є критично важливою для розуміння механізмів поширення інформації, виявлення фейків та прогнозування соціальних трендів.

Таким чином, застосування методів, керованих даними (data-driven), є найбільш перспективним шляхом для комплексного вирішення задач захисту та прогнозування в соціальних мережах.

Мета і завдання дослідження.

Метою роботи є підвищення ефективності моделювання процесів у соціальних мережах шляхом розробки та застосування data-driven методів для симуляції динаміки активності користувачів під впливом екзогенних факторів.

Для досягнення поставленої мети вирішено такі завдання:

1. Провести аналіз предметної області, зокрема існуючих методів захисту приватності, моделей атак на соціальні графи та підходів до симуляції інформаційних потоків.

2. Розробити фреймворк для оцінки вразливості соціальних графів до атак деанонімізації із використанням методів машинного навчання.

3. Розробити модульну архітектуру симулятора процесів у соціальних мережах, що враховує вплив зовнішніх (екзогенних) подій на активність користувачів.

4. Виконати емпіричну перевірку запропонованих моделей та оцінити їхню продуктивність на реальних та синтетичних наборах даних.

Об’єктом дослідження є процеси інформаційної взаємодії, функціонування даних та поширення інформаційних потоків у соціальних мережах.

Предметом дослідження є методи та моделі оцінки вразливості структур графів до деанонімізації, а також методи симуляції динаміки активності користувачів із застосуванням data-driven підходів.

Методи дослідження.

У роботі використано комплексний методологічний апарат:

- теорія графів — для аналізу топологічних властивостей соціальних мереж та моделювання зв’язків;

- методи машинного навчання (класифікація, регресія) — для побудови моделей атак деанонімізації та прогнозування активності;

- статистичний аналіз — для виявлення кореляцій між метриками графа та його вразливістю;

- комп’ютерне моделювання та симуляція — для відтворення процесів генерації каскадів ретвітів та перевірки ефективності розроблених фреймворків.

Наукова новизна одержаних результатів

Удосконалено метод оцінки приватності соціальних графів шляхом створення фреймворку на основі машинного навчання, який, на відміну від існуючих підходів, розглядає атаку деанонімізації як задачу класифікації вразливості вузлів, що дозволяє точніше ідентифікувати ризики витоку даних.

Встановлено кількісні залежності (асоціативність та причинність) між специфічними топологічними метриками (центральність, кластеризація) та ймовірністю успішної деанонімізації вузла в умовах структурних атак.

Практичне значення одержаних результатів.

Розроблені моделі та програмні засоби дозволяють власникам та адміністраторам соціальних платформ проводити аудит безпеки даних та виявляти вразливі сегменти мережі перед публікацією анонімізованих датасетів. Аналітикам даних використовувати інструментарій симуляції для прогнозування реакції соціуму на зовнішні інформаційні приводи та маркетингові кампанії.

Структура магістерської роботи. Представлена робота складається зі вступу, трьох розділів та висновків. Загальний обсяг роботи становить 79 сторінок, і містить 19 рисунків, 2 таблиці, перелік використаних джерел із 45 позицій та одного додатку.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ МОДЕЛЮВАННЯ ПРОЦЕСІВ ПРИВАТНОСТІ В СОЦІАЛЬНИХ МЕРЕЖАХ

1.1. Наукові основи захисту приватності та симуляції інформаційних потоків у соціальних мережах

Платформи соціальних мереж (ПСМ), такі як YouTube, X, Reddit та Facebook, набули значної популярності в останні роки, функціонуючи як важливі інструменти для комунікації та соціальної взаємодії. Ці платформи генерують великі мережеві масиви даних (датасети), які часто відображають реальні соціальні взаємодії між суб'єктами, включаючи дружні зв'язки, підписки та професійні відносини. Цінність цих датасетів для дослідницької спільноти є суттєвою, оскільки вони сприяють проведенню різноманітних досліджень, таких як аналіз еволюції спільнот, дослідження поляризації думок, моделювання реагування на катастрофи, оцінка расової/етнічної нерівності, виявлення психологічного стресу та інші.

Ця магістерська робота присвячена двом критично важливим напрямкам досліджень соціальних мереж: забезпеченню приватності окремих осіб у публічно доступних даних ПСМ та симуляції онлайн-активності користувачів на різних платформах соціальних мереж. Перше дослідження обумовлено проблемами доступу та конфіденційності даних ПСМ, що виникають через чутливий характер інформації, яку вони фіксують. Наприклад, виникають серйозні загрози приватності, коли датасети соціальних мереж потенційно розкривають політичні вподобання, сексуальну орієнтацію, корпоративні облікові дані тощо. Мета другої частини дослідження полягає у розробці моделей для симуляції поширення інформації в соціальних мережах у різних контекстах. Здатність прогнозувати майбутню активність у соціальних мережах має прямі практичні застосування. Зокрема, адміністратори платформ можуть використовувати такі прогнози для ідентифікації користувачів, які можуть

публікувати провокаційний контент у дискусії, та для контролю або цензурування їхньої активності. Крім того, ці можливості дозволяють оцінювати ефективність стратегій втручання, спрямованих на обмеження розповсюдження дезінформації.

1.1.1 Проблематика приватності даних соціальних мереж

Незважаючи на значну наукову цінність, публічне оприлюднення датасетів соціальних мереж становить серйозну загрозу для приватності окремих користувачів. Існує безліч векторів для порушення приватності. Наприклад, зловмисники можуть прагнути встановити, чи є певний користувач активним учасником конкретного політичного форуму, чи існує зв'язок між двома користувачами в мережі знайомств, або чи група користувачів у певному географічному районі голосувала за конкретного кандидата.

Для зменшення вторгнення в приватність окремих осіб було запропоновано низку методів захисту даних [2]. Наприклад, ідентичність користувача може бути захищена шляхом наївного знеособлення (санкціонування) через просте видалення ідентифікуючих атрибутів із загальнодоступних даних, або через структурну анонімізацію, при якій вузли та зв'язки в соціальній мережі видаляються/додаються для приховування оригінальної топологічної структури. Проте порушення даних відбуваються регулярно, оскільки зловмисники застосовують складні техніки для подолання існуючих механізмів захисту даних. Атака деанонімізації на датасет дзвінків «Data for Development» (D4D) [3] є яскравим прикладом порушення приватності окремих осіб, спричиненого неналежно знеособленими публічними даними. Датасети D4D містили «анонімізовані» записи дзвінків та SMS-повідомлень, зібрані від користувачів основної комунікаційної мережі. Однак зловмисники успішно розкрили ідентичність користувачів, застосувавши потужну атаку деанонімізації [4]. Вони

використовували інформацію з різних анонімізованих підграфів для декодування анонімізованих облікових записів користувачів.

Ключове наукове питання полягає в тому, як ефективно анонімізувати графи без істотної втрати їхньої корисності [5]. Наприклад, збереження певних мережевих характеристик (наприклад, розподілу ступенів, коефіцієнта кластеризації тощо) в анонімізованому графі є важливим для кінцевих застосувань. Як правило, ті методи, які зберігають більшу корисність, роблять анонімізовані графи вразливими до сучасних атак деанонімізації [6]. Не до кінця з’ясованою залишається взаємодія між гарантіями анонімності, які надаються схемами анонімізації, та потужністю атаки, а також властивостями датасету, що підлягає анонімізації. Або, чи є деякі мережі за своєю природою більш «анонімізованими», тобто стійкими до потужних атак навіть при застосуванні слабких схем структурної анонімізації. Основною науковою проблемою є розробка принципового розуміння того, як кількісно оцінити ефективність схеми анонімізації, і, відповідно, ймовірний успіх атаки деанонімізації. У цій дисертації ми прагнемо зрозуміти, які властивості роблять певні датасети графів більш стійкими до атак деанонімізації.

1.1.2. Моделювання активності в соціальних мережах

Розуміння механізмів поширення інформації в онлайн-соціальних середовищах має значний практичний вплив, починаючи від сфери охорони здоров'я і закінчуючи маркетингом. Значні дослідницькі зусилля були спрямовані на характеристику поширення інформації на різних платформах. Наприклад, в [7] охарактеризували типи інформаційних каскадів у Facebook, продемонструвавши, що типи каскадів залежать від факторів, пов'язаних із зусиллями та соціальною вартістю участі користувачів. У роботі [8] досліджували соціальне зараження шахрайською поведінкою на онлайн-ігрових платформах. В дослідженні [9], на основі колекції твітів, що містять політичні новини, встановили, що неправдива інформація поширюється

швидше, далі, глибше та ширше, ніж правдиві факти. Це явище частково пояснюється людськими факторами, такими як емоційна реакція на здивування, страх та огиду, які частіше викликаються фальсифікованими новинами.

Нашою метою є розробка соціального симулятора, здатного фіксувати поширення інформації як усередині, так і між різними платформами соціальних мереж. Симулятор набуває найбільшої корисності, коли він може прогнозувати реалістичну онлайн-активність користувачів із високою деталізацією (хто, кому, на яку тему відповідає і коли) на майбутньому часовому горизонті без доступу до фактичної активності. Хоча це формулювання виглядає простим, було показано, що такі прогнози є складними. Це частково пояснюється нерегулярними патернами інформаційних потоків, спричиненими впливом як внутрішніх, так і зовнішніх чинників, а також тим, що різні соціальні платформи використовують різні алгоритми для просування контенту.

Надійний симулятор повинен реалістично реагувати на внутрішні та зовнішні стимули, забезпечуючи:

- Фіксацію піків активності, спричинених певними темами інтересу.

- Реалістичну реакцію на час зовнішніх подій та внутрішніх інтенсивних дискусій.

- Моделювання активності за темою, навіть якщо теми слабо пов'язані.

- Точне представлення розміру нової залученої аудиторії, яка може значно змінюватися з часом та темами.

Симуляція активності користувачів на онлайн-платформах соціальних мереж має численні переваги. Ці прогнози можуть бути використані для вивчення сценаріїв «що, якщо» в операційному середовищі. Наприклад:

- Яка реакція буде згенерована, якщо певний пост буде опублікований конкретним обліковим записом користувача? Тобто, наскільки великою буде реакція з точки зору повідомлень та залучення користувачів з часом?

- Що, якщо це повідомлення буде опубліковане іншим користувачем (наприклад, урядовою організацією проти облікового запису бота)?

Крім того, дослідники можуть тестувати ефекти втручання в платформу для впливу на активність:

- Чи суттєво вплине блокування певних облікових записів на кампанію дезінформації?

- Наскільки пізно у проведенні інформаційної операції втручання буде ефективним, враховуючи, що може знадобитися час для ідентифікації інформаційної кампанії та її операторів?

Інші застосування для такого симулятора включають генерацію реалістичних датасетів для заповнення прогалин у даних, зібраних для різних наукових досліджень; вивчення міжплатформного поширення інформації; або ідентифікацію користувачів, які прагнуть просувати насильство під час виборчого сезону.

1.2. Загроза приватності даних у соціальних мережах

Соціальні мережі підлягають інтенсивному аналізу з метою виявлення структури та функціональних особливостей представлених ними взаємодій. Незважаючи на значну наукову цінність цих даних для дослідницької спільноти, їхнє публічне оприлюднення супроводжується суттєвими ризиками для приватності окремих суб'єктів.

Класичними ілюстраціями порушення приватності внаслідок публікації неналежно знеособлених (санкціонованих) даних є інциденти з AOL та Netflix. Перший скандал у 2006 році був пов'язаний з оприлюдненням компанією AOL анонімізованих журналів пошукових запитів. Ці записи містили веб-запити понад 500 000 американських користувачів пошукової системи AOL протягом тримісячного періоду. Двоє журналістів видання New York Times здійснили реанонімізацію (деанонімізацію), зіставивши особисту ідентифікаційну інформацію, присутню в анонімізованих записах, із

публічно доступними телефонними довідниками, успішно розкривши ідентичність кількох користувачів. Найбільш відомим повторно ідентифікованим обліковим записом став Користувач № 4417749, 62-річна вдова Тельма Арнольд, чиї пошукові запити включали теми, як-от «оніміння пальців», «60 самотніх чоловіків» тощо [5]. Було встановлено, що багато інших облікових записів, що належали онкохворим пацієнтам, вагітним матерям та студентам, також могли бути повторно ідентифіковані за допомогою подібної методології. Це порушення приватності призвело до подання колективного позову проти AOL. Другий інцидент стосувався публікації оцінок фільмів Netflix у рамках конкурсу Netflix Prize. Два академічні дослідники зіставили ці записи з оцінками з бази даних Internet Movie Database (IMDb). Вони змогли ідентифікувати багатьох користувачів, присутніх в обох датасетах, незважаючи на те, що їхні ідентичності були анонімізовані в наборі даних Netflix.

Для мінімізації ризиків вторгнення в приватність при публічному оприлюдненні графових даних було запропоновано численні методи анонімізації. Сучасний підхід полягає в анонімізації соціальних графів шляхом достатньої модифікації структури графа для роз'єднання конкретної ідентичності вузла з його соціальними зв'язками, водночас зберігаючи загальні характеристики графа. Було розроблено різні рішення: деякі базуються на збуренні оригінальної структури графа, інші — на кластеризації, а ще інші — на генерації графів з графової сигнатури.

Для всіх методів структурної анонімізації графів існує принципова напруженість між забезпеченням приватності у зміненій структурі графа та збереженням точності структурних характеристик оригінального графа у зміненому графі, що є критично важливим для його дослідницької корисності.

На одному кінці спектра знаходиться метод, де анонімізований граф ізоморфний оригіналу, що максимально зберігає структурну корисність даних, але, як наслідок, робить його найбільш вразливим до базових атак

деанонімізації. На протилежному кінці спектра знаходиться генерація випадкових графів, яка може розглядатися як метод анонімізації для створення графа, що повністю неізоморфний оригінальному. Хоча цей метод забезпечує вищий рівень приватності, значна втрата оригінальної структури графа може негативно вплинути на точність аналізу з використанням анонімізованих даних. Як правило, чим більше методи анонімізації зберігають корисність графа, тим більш вразливими вони є до сучасних атак деанонімізації [9]. Отже, ключовим питанням залишається: як ефективно анонімізувати графи без нівелювання їхньої корисності, при цьому забезпечуючи захист приватності користувачів.

Різні дослідження розглядали цю проблему, зазвичай у контексті специфічних методів анонімізації та конкретних метрик корисності [8, 9]. Наприклад, в [10] провели порівняльний аналіз методів анонімізації на основі збурення щодо збереженої корисності та стійкості до конкретних атак деанонімізації. Однак у поточному стані досліджень відсутнє систематичне розуміння обмежень анонімності, що накладаються вимогами корисності.

Як було з'ясовано під час дослідження існуючих методів анонімізації та деанонімізації (ДА), усі вони мають обмеження при оцінці їхньої ефективності. Наприклад, розуміння стійкості/вразливості сучасних схем анонімізації до новітніх атак деанонімізації все ще залишається відкритою науковою проблемою.

Для вирішення цієї проблеми було реалізовано систему публікації/ спільного використання безпечних графових даних (SecGraph). Загальний огляд системи SecGraph представлений на рис. 1.1.

SecGraph складається з трьох основних модулів:

- Модуля анонімізації (AM),
- Модуля оцінки корисності (UM),
- Модуля оцінки деанонімізації (DM).

Основні функції кожного модуля коротко підсумовані нижче.

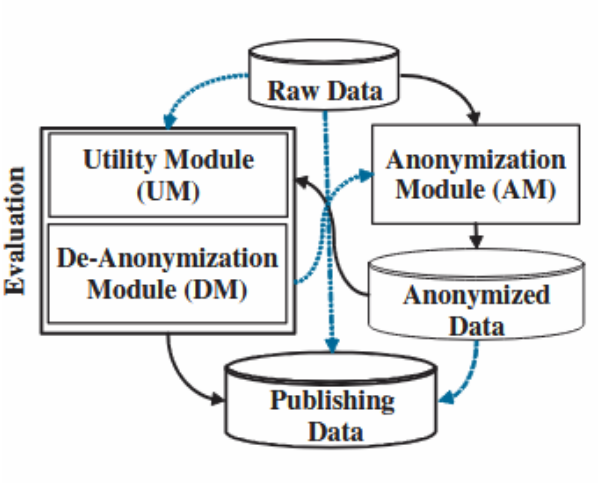


Рис. 1.1. Огляд структури системи використання безпечних графових даних SecGraph

Основна функція модуля модуля анонімізації полягає в анонімізації сирих графових даних та генерації анонімізованих даних. В роботі автори реалізували 11 сучасних схем анонімізації графових даних, які включають алгоритми на основі ентропії ребер, алгоритми, що базуються на k-анонімності та її варіантах, алгоритми на основі агрегації/класу/ кластеризації, алгоритми, засновані на диференційній приватності, а також алгоритм, що базується на випадкових блуканнях.

Конкретно, виникає питання: які властивості графів надають найбільше інформації, дозволяючи ідентифікувати значну частину вузлів? Розуміння відповіді на це питання є корисним у багатьох практичних аспектах.

1. Прийняття рішень щодо анонімізації.

Це може допомогти фахівцям з даних визначити, які властивості графів слід уникати зберігати в анонімізованій версії датасету з метою підвищення анонімності вузлів. Наприклад, якщо встановлено, що спільний розподіл ступенів є надмірно розкриваючим, то метод анонімізації, який зберігає розподіл ступенів оригінального графа, повинен розглядатися як такий, що несе значні ризики приватності, і його слід уникати.

2. Розробка нових алгоритмів.

Можуть бути розроблені нові методи анонімізації з явною метою збурення (приховування) саме тих властивостей оригінального графа, які виявилися надто інформативними. Наприклад, якщо для певної мережі коефіцієнт асортативності ступеня (тенденція вузлів зі схожими ступенями бути з'єднаними) суттєво сприяє повторній ідентифікації вузлів, може бути розроблений алгоритм анонімізації, спрямований на збурення цього коефіцієнта. Це відкриває новий напрямок у просторі алгоритмів анонімізації графів, де традиційні цілі проектування зосереджувалися на збереженні, а не на явному збуренні структурних властивостей.

В даній роботі пропонується фреймворк моделювання для: 1) кількісної оцінки вразливості графа до атаки повторної ідентифікації; та 2) кількісної ідентифікації тих структурних властивостей, які найбільше сприяють цій вразливості графа. Корисність цього фреймворку демонструється на великому наборі синтетично згенерованих графів з контрольованими властивостями, які імітують характеристики реальних соціальних мереж.

1.3. Огляд літератури по темі дослідження

За останнє десятиліття було досягнуто значного прогресу у вирішенні проблем, пов'язаних з анонімізацією графів. Для контекстуалізації наших результатів у широкій літературі з анонімізації графів, ми обговорюємо пов'язані роботи, структуруючи їх відповідно до наших основних наукових внесків.

1.3.1. Приватність та корисність графових даних

Оскільки корисність традиційно визначається через (відстань між) графові метрики, які, своєю чергою, описують властивості мережі, наше дослідницьке питання щодо того, які саме властивості мережі підвищують вразливість графів до атак деанонімізації (ДА), тісно пов'язане з проблемою компромісу між корисністю та анонімністю.

Значні зусилля були спрямовані на розуміння внутрішнього конфлікту між досягнутим рівнем приватності та збереженою корисністю при публікації графових датасетів. Наприклад, незважаючи на те, що будь-яка схема анонімізації, яка зберігає розподіл ступенів, є вразливою до атак ДА, збурення розподілу ступенів у процесі анонімізації призводить до значної втрати корисності, що виражається у зміні важливих властивостей графа в анонімізованих графах. Частка висячих вузлів (вузлів лише з одним сусідом) є важливим фактором у підтримці анонімності, оскільки вони інтуїтивно несуть мінімальну інформацію для розкриття ідентичності свого (єдиного) сусіда. Крім того, емпірично було показано, що корисність зменшується швидше, ніж досягається приріст приватності.

Були запропоновані теоретичні фреймворки для кількісної оцінки компромісу між приватністю та корисністю. В роботі [11] розробили теоретичну модель для кількісної оцінки деанонімізації графових датасетів, що враховує топологічну важливість вузлів. Вони дійшли висновку, що приватність залежить від високого середнього ступеня вузлів. В роботі [12] проаналізували зв'язок між корисністю анонімізованого графа та його вразливістю до атаки повторної ідентифікації вузлів на основі спільних сусідів. Вони сформулювали умови успіху атак ДА, ґрунтуючись на двох метриках корисності, заснованих на відстані між анонімізованим (або допоміжним) та оригінальним графами.

Відмінності між дослідженнями приватності/анонімності та нашим підходом полягають у наступному:

- Наше питання стосується властивостей оригінального графа, а не будь-якої конкретної анонімізованої версії. Таким чином, отримані відповіді є незалежними від методу анонімізації і стосуються внутрішніх властивостей оригінальної мережі.

- Не зосереджуючись виключно на метриках корисності, ми не обмежені вибором підмножини «корисних» властивостей мережі для

конкретного контексту. Це дозволяє провести більш широке дослідження графових властивостей та їхнього впливу на вразливість.

1.3.2. Моделі атак деанонімізації та метрики успіху

Загальноприйнята атака деанонімізації графів використовує інформацію з допоміжного графа для повторної ідентифікації вузлів у анонімізованому графі [13]. Успіх такої атаки визначається рівнем правильної повторної ідентифікації оригінальних вузлів у мережі. Загалом, атаки ДА використовують структурні характеристики вузлів, які є унікально відмінними. Більшість таких атак можна класифікувати на атаки з початковими даними (seed-based) та атаки без початкових даних (seed-free), залежно від наявності у атакуючого попередньої інформації про початкові вузли.

Процес деанонімізації здійснюється для повторної ідентифікації вузлів та зв'язків з використанням підроблених вузлів (sybil nodes) [14] або відомих відображень вузлів у допоміжному графі. Ефективність таких атак корелює з якістю початкових даних.

Щодо атак без початкових даних, то у цьому випадку проблема деанонімізації зазвичай моделюється як задача узгодження графів [15] (також відома як вирівнювання мереж). Мета узгодження мереж — знайти коректне відображення між наборами вузлів двох структурно корельованих графів. Недавні роботи пропонують інформаційно-теоретичні умови, за яких можливе ідеальне відображення [16, 17]. Більшість цих досліджень базуються на моделях Ердеша-Реньї (теоретичні моделі, що не завжди репрезентативні для реальних датасетів) і припускають необмежені обчислювальні ресурси, тоді як інші роблять непрактичні припущення щодо початкових даних, наприклад, наявність вузлів-хабів як початкових даних.

Деякі дослідження запропонували статистичні моделі для повторної ідентифікації вузлів без початкових даних, зокрема байєсівська модель або моделі оптимізації [18]. Для процесу поширення повторної ідентифікації

було застосовано численні евристики, використовуючи такі характеристики графів, як ступінь, k-окіл, коваріація зв'язків, ексцентриситет або властивість належності до спільноти.

Деякі методи анонімізації ґрунтуються на збуренні набору ребер в оригінальному графі в межах заданого бюджету приватності. Наприклад, диференціальна приватність фіксує кількість введеного шуму, що також є популярною теоретичною метрикою для кількісної оцінки приватності анонімізованого графа. Проте диференціальна приватність є високочутливою до бюджету приватності, який вимірює максимальну кількість прийнятних запитів без витоку секретів. Крім того, було показано, що метрики приватності, засновані на диференціальній приватності, можуть переоцінювати приріст приватності.

Припускається, що користувач надає свої дані через систему обміну інформацією з метою отримання певної послуги (корисності). Ми також виходимо з припущення, що користувачі прагнуть захистити свою чутливу інформацію при обміні даними з недовіреними суб'єктами. Наприклад, у випадку надання даних з геолокаційними мітками постачальнику послуг, користувач може мати на меті приховати точні відвідані місця, їхню семантику або діяльність, яку можна виснувати з цих локацій. Ми називаємо цю чутливу інформацію користувача її секретом (s).

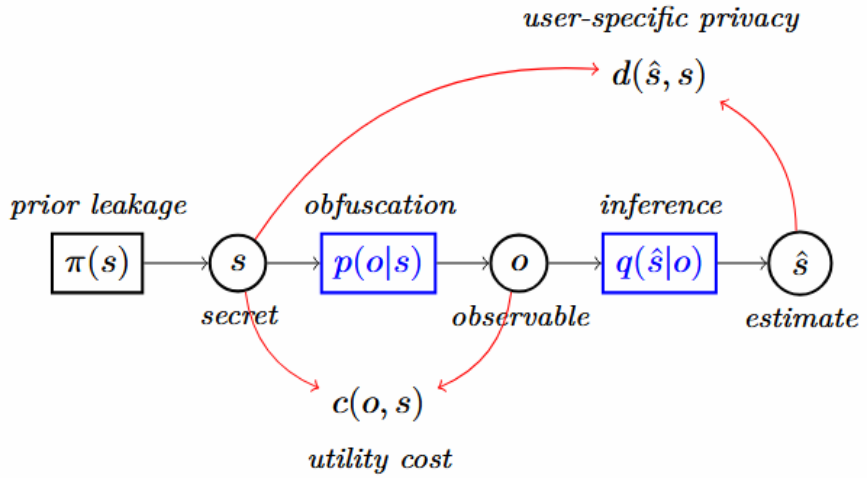


Рис. 1.2. Фреймворк обміну інформацією

Для захисту своєї приватності ми припускаємо, що користувач обфускує свої дані перед їхнім обміном або публікацією. На рис. 1.2 проілюстровано потік інформації.

Розподіл імовірності π кодує апріорну оцінку користувача щодо витоку інформації про секрет s. Секрет обфускується за допомогою механізму захисту p, виходом якого є спостережуване значення o. Адаптивний зловмисник (передбачений користувачем) виконує атаку висновку q на o та отримує розподіл імовірності за оцінками s^. Функція відстані c позначає вартість корисності механізму захисту, спричинену обфускацією. Функція відстані d позначає приватність користувача (для метрики приватності на основі спотворень) або необхідну нерозрізненість між секретами (для метрики диференційної приватності). Користувач визначає функцію відстані d відповідно до своїх чутливостей приватності.

В роботі [19] запропоновано загальну модель загрози для вимірювання успіху атаки деанонімізації, яка є незалежною від схеми анонімізації. Він представив фреймворк машинного навчання для порівняння схем анонімізації графів на основі збурень щодо збереженої корисності та стійкості до конкретних атак ДА. Цей фреймворк досліджує приховані інваріанти та подібності для повторної ідентифікації вузлів в анонімізованих графах. Важливо, що цей фреймворк легко адаптується для моделювання різних типів атак. У цьому дослідженні ми базуємося на підході Sharad.

Висновки до розділу

Цей розділ заклав теоретичну базу для дослідження, визначивши критичну проблему приватності даних у соціальних мережах на тлі зростаючої складності інформаційних потоків та активності користувачів. Було встановлено, що моделювання процесів у соціальних мережах вимагає інтеграції двох аспектів: симуляції інформаційних потоків та оцінки загроз приватності. Аналіз літератури виявив ключові компроміси між приватністю

та корисністю графових даних, підкреслюючи, що висока корисність часто корелює з високою вразливістю. Особлива увага приділена моделям атак деанонімізації, які використовують топологічні та атрибутивні ознаки для ідентифікації користувачів. Таким чином, якісне моделювання має не лише симулювати активність, але й кількісно оцінювати ризики, пов'язані з несанкціонованим розкриттям даних. Це стало основою для розробки спеціалізованого фреймворку оцінки вразливості у наступному розділі.

РОЗДІЛ 2. МОДЕЛЮВАННЯ ПРОЦЕСІВ ПРИВАТНОСТІ В СОЦІАЛЬНИХ МЕРЕЖАХ НА ОСНОВІ ОЦІНКИ ВРАЗЛИВОСТІ ГРАФІВ

2.1. Фреймворк для оцінки вразливості графів до атак деанонімізації із застосуванням машинного навчання

Основна мета цього дослідження полягає у кількісній оцінці взаємозв'язку між структурними властивостями графа та рівнем приватності його вузлів. Приватність вузла визначається як його здатність зберігати захищеною ідентичність. Інтуїтивно, у регулярних графах, де ступінь кожного вузла є однаковим, приватність висока, оскільки топологічна інформація не дозволяє диференціювати вузли. На протилежному полюсі, вузол з унікальною топологічною характеристикою, наприклад, центр зіркоподібного графа, легко ідентифікується за наявності допоміжної інформації. Реальні мережеві набори даних демонструють топології, які знаходяться між цими двома крайніми випадками.

Захист ідентичності вузла може бути досягнутий шляхом наївної санкціонізації (видалення ідентифікаційних атрибутів) або структурної анонімізації (модифікація топологічної структури шляхом видалення/додавання вузлів і ребер). У контексті цієї роботи розрізнення між цими сценаріями не є критичним, оскільки дослідження зосереджується на зв'язку між властивостями графа (незалежно від того, чи є він вихідним, чи структурно збуреним) та рівнем приватності його вузлів.

Ключове питання: які структурні властивості графа (задана топологія) містять найбільший обсяг інформації, потенційно придатної для ідентифікації його вузлів? Це питання стосується внутрішньої вразливості вихідної мережі до атаки повторної ідентифікації або, у випадку збуреного графа, вразливості структурно анонімізованого графа до атаки деанонімізації.

У подальшому викладі терміни "атака повторної ідентифікації" та "атака деанонімізації" використовуються як взаємозамінні.

2.1.1 Архітектура фреймворку

Для вирішення поставленого питання розроблено трикомпонентний фреймворк (рис. 2.1).

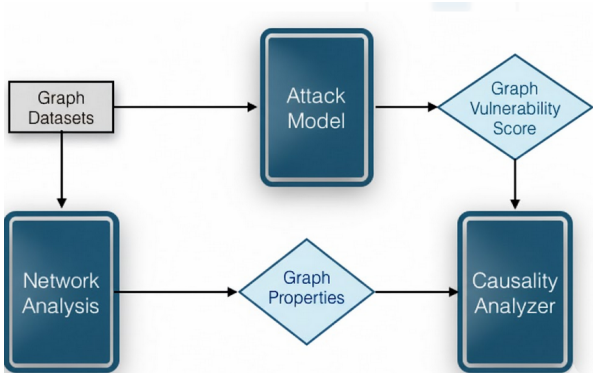


Рис. 2.1. Фреймворк для вимірювання приватності та корисності, де аналізується взаємозв'язок між вразливістю графа та його властивостями

Компонент "Модель Атаки" реалізує атаку повторної ідентифікації на вхідному графі та генерує показник вразливості. Цей компонент є гнучким, дозволяючи інтегрувати будь-який алгоритм атаки. Для експериментів застосовано алгоритм на основі машинного навчання, що базується на визначеній моделі загрози. Визначення показника вразливості залежить від конкретної реалізованої моделі атаки.

Компонент "Аналіз Мережі" виконує обчислення традиційних мережевих метрик (наприклад, показників центральності, щільності, розподілів тощо). Будь-які метрики, що становлять інтерес, можуть бути отримані у вигляді числових значень або розподілів. Детальний опис імплементації не вимагається через широку доступність спеціалізованих бібліотек для мережевого аналізу.

Компонент "Аналізатор Причинності" на вхід отримує метрики мережі та показник вразливості вихідного графа. Виконує ретельний аналіз

взаємозв'язку між вразливістю графа та його структурними властивостями. Вихідні дані цього компонента надають статистичну відповідь на дослідницьке питання.

2.1.2. Модель загрози

Для кількісної оцінки вразливості графа до атак повторної ідентифікації вузлів використовується підхід на основі машинного навчання, що має на меті знайти бієктивне відображення між вузлами двох різних, але перекриваючихся графів (Gsan та Gaux).

Класична модель загрози [2] - зловмисник прагне узгодити вузли з двох мереж, чиї набори ребер корельовані.

Наведемо приклад. Припустимо, що внаслідок порушення приватності облікових записів Unix, зловмисник отримує частковий доступ до двох різних, потенційно перекриваючихся, підмереж (наприклад, Facebook та X) студентів. Деякі особи присутні в обох графах, хоча їхні ідентифікатори видалені. Завдання зловмисника — встановити бієктивне відображення між вузлами, що відповідають спільним особам.

Виконаємо формалізацію. Зловмисник має доступ до санкціонованого графа (Gsan) та допоміжного графа (Gaux). У прикладі Gsan — мережа Facebook, Gaux — мережа X.

Для моделювання використовується реальний набір даних $G = (V,E)$, розділений на два підграфи G1=(V1,E1) та G2=(V2,E2), де $V1 \subset V$, $V2 \subset V$ та $V1 \cap V2 = V\alpha \neq \phi$. Частка перекриття α вимірюється коефіцієнтом Жаккара:

$$\alpha = \frac{|V_1 \cap V_2|}{|V_1 \cup V_2|}$$

Вузли в Vα (спільний підграф) зберігають свої ребра, але можуть мати різні ребра до вузлів, які є унікальними для G1 або G2. Оптимістичний сценарій атаки припускає, що зловмисник має доступ до частини

оригінального графа (G1) як допоміжних даних (Gaux) та незбуреного підграфа (G2) як санкціонованих даних (Gsan). Рекурсивне розділення може бути використане для оцінки можливості деанонімізації у великих мережах. Перекриття є джерелом знань, що використовується зловмисником для деанонімізації.

Встановлено фіксоване значення α=0.2. Для експериментування з різною "силою" атаки, Vα будується чотирма способами:

1) R: Випадкова колекція вузлів.

2) HD: Вибір вузлів з найвищим ступенем.

3) BFS-R: Побудова дерева пошуку вширину (BFS) від випадково вибраного вузла.

4) BFS-HD: Побудова дерева пошуку вширину (BFS) від вузла з найвищим ступенем.

2.1.3. Алгоритм атаки на основі машинного навчання

Реалізовано алгоритм атаки на основі машинного навчання (Random Forest), що забезпечує потужний еталон для оцінки вразливості [14]. Атака використовує структурну інформацію для узгодження вузлів на основі подібних структурних характеристик, навчаючись структурним шаблонам мережі.

Для кожного вузла використовується розподіл ступенів сусідства (NDD). NDD є стійким до шуму та загальним методом представлення [15].

NDD-підпис: Об'єднання бінованої версії NDDu1 та NDDu2. Відстань q=2 є достатньою для соціальних мереж.

Використовується розмір бінів 50. 21 бін відповідає ступеню вузла до 1050; більші значення агрегуються в останньому біні. Ця стратегія розроблена для захоплення агрегатної структури его-мереж.

Завдання класифікації: використовується ансамблевий алгоритм Random Forest для бінарної класифікації: 1 для ідентичних пар вузлів (Gsan ∩Gaux), 0 для неідентичних пар.

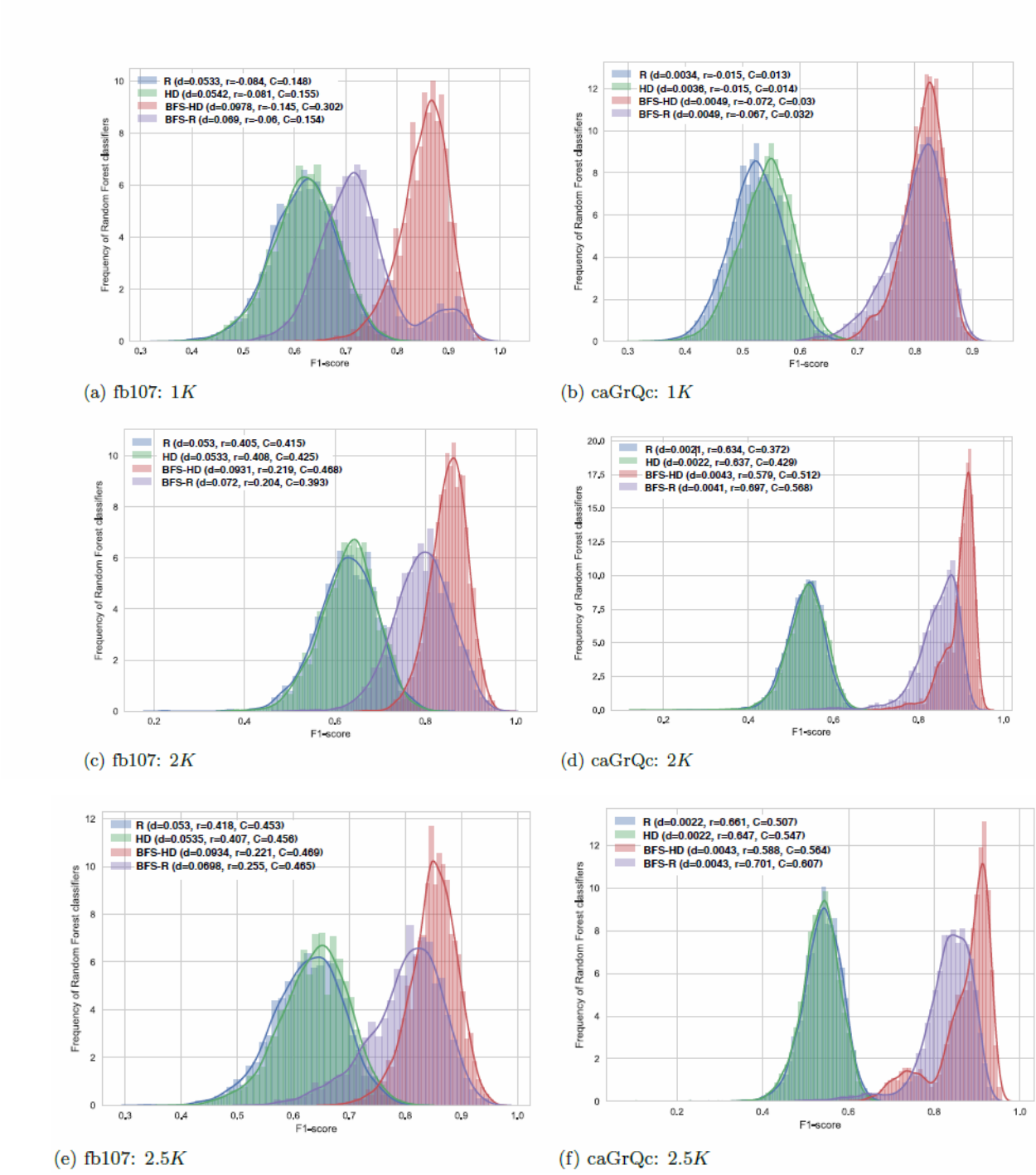


Рис. 2.2. Порівняння сили атаки залежно від різних варіантів вибору перекриття

Варіанти перекриття включають Випадковий (R), Високий Ступінь (HD) та BFS-дерева (з коренем у вузлі з найвищим ступенем BFS-HD і, відповідно, у випадковому вузлі BFS-R). Представлено точність прогнозування ідентичних пар у різних просторах dK. Властивості графа — щільність ($\bar{d}$),

асортативність (r) та транзитивність (C) — усереднені для 8 підграфів на кожен простір dK, що асоційовані з наданим перекриттям.

Навчання та оцінка полягає в наступному:

- Генерація прикладів: випадковий вибір пар вузлів із Gsan та Gaux.

- Небалансованість даних: використовуються техніка резервуарного вибіркового методу [16] для взяття ℓ=1000 збалансованих підвибірок та алгоритм SMOTE [17] (метод надвибірки) для кожної підвибірки.

Кожна вибірка тренується лісом з j=100 випадкових дерев рішень. Загалом: ℓ×j=100,000 сценаріїв атак.

Якість класифікатора вимірюється F1-оцінкою (середнє гармонійне між точністю та повнотою). Використовується 5×2 крос-валідація.

Експериментальні результати (рис. 2.2) підтверджують, що атака є:

- Сильнішою та значущою, коли вузли в перекритті з'єднані (сценарії BFS).

- Сильнішою, коли щільність перекриття вища.

Для решти дослідження використовується BFS-HD для генерації перекриття, оскільки це забезпечує сильнішу атаку завдяки багатшій базі знань. Цей механізм також репрезентативний для методів вирівнювання мереж на основі перколяції [8].

Фреймворк дозволяє систематично досліджувати, які саме структурні характеристики мережі роблять її вузли вразливими до деанонімізації. Застосування потужної, узагальненої атаки на основі машинного навчання (Random Forest з NDD-підписами) служить надійним еталоном для кількісної оцінки цього ризику, незалежно від конкретної схеми анонімізації.

2.2. Аналіз взаємозв'язку між топологічними метриками та вразливістю графа

Мета компонента аналізатора причинності полягає у виявленні впливу топологічних метрик на вразливість графа. У цій реалізації ми досліджуємо

як причинні, так і асоціативні зв'язки. Для обох аналізів використовується однаковий набір вхідних даних: показник вразливості (виміряний як F1-оцінка) та вимірювання графа, отримані за допомогою стандартних методів аналізу соціальних мереж. Застосування різних аналітичних інструментів дозволяє ізолювати силу причинних та асоціативних відносин. Зазначені інструменти є взаємозамінними, що забезпечує гнучкість методології.

2.2.1. Причинність через пояснювальне моделювання

Для кількісної оцінки значущості причинного зв'язку між метриками графів та їхньою вразливістю застосовуються техніки пояснювального моделювання [20].

Ми оцінюємо функцію вразливості графа f за допомогою серії регресійних тестів (як лінійних, так і ентропійних). Кожна модель оцінює індивідуальний вплив структурних метрик на пояснення показника вразливості.

Залежна змінна - показник вразливості (F1-оцінка).

Незалежні змінні - вибрані структурні властивості, що включають макрорівневі метрики: щільність, асортативність, транзитивність, середня довжина шляху та частка вузлів зі ступенем 1. Ці властивості були обрані для дослідження важливості структури спільноти для успіху атак повторної ідентифікації.

Причинність у відносинах вимірюється за допомогою F-тесту [70] та взаємної інформації (MI) [21].

F-тест вимірює значущість кореляції будь-яких незалежних змінних із залежною змінною в межах моделей багатовимірної лінійної регресії.

Взаємна інформація (MI) визначається як нелінійна функція спільного ймовірнісного виміру, що фіксує будь-який тип залежності у просторі змінних. Обидва показники, F-тест та MI, знаходяться в діапазоні [0,1], де вищі значення вказують на більш значущі залежності.

Для виведення причинного зв'язку використовується фреймворк Перла [22], який генерує спрямований ациклічний граф (DAG). Зокрема, застосовується алгоритм IC* (індуктивна каузація), що дозволяє встановити, чи має змінна X прямий причинний вплив на змінну Y. Алгоритм IC* рекурсивно будує DAG, де вузли представляють змінні, а спрямовані ребра — причинні зв'язки, на основі ймовірнісних тестів на умовну незалежність.

Критично важливо, що IC* здатний робити висновки про латентну причинну структуру, припускаючи існування прихованих сплутуючих змінних, що актуально, оскільки аналіз не охоплює весь простір можливих метрик графа.

2.2.2. Асоціативність через прогностичне моделювання

Для виявлення потенційної асоціації між метриками графів та вразливістю кількісно оцінюється рівень передбачуваності показника вразливості на основі структурних властивостей.

Для моделювання прогнозу будується функція вразливості графа f шляхом навчання на прикладах отриманих показників вразливості та відповідних структурних властивостей. Використовується той самий набір структурних властивостей, що й у пояснювальному моделюванні. Моделі оцінюють здатність прогнозувати показник вразливості для небачених графів, маючи лише їхні структурні метрики.

Для уникнення завищених оцінок прогностичних результатів застосовуються техніки крос-валідації (наприклад, тестові вибірки). Крос-валідація дозволяє імітувати, як модель працюватиме в реальних умовах, використовуючи лише наявний набір даних.

Метод 5×2 крос-валідації є модифікацією, яка часто використовується для порівняння двох різних алгоритмів. Він працює наступним чином:

- Набір даних ділиться на 2 згини (50/50). Згином є рівні підмножини набору даних вибірки.

- Обидва алгоритми тренуються та тестуються на цих двох згинах (спочатку згин 1 — тренувальний, згин 2 — тестовий; потім навпаки).

- Цей процес (дві ітерації) повторюється 5 разів із новим випадковим розділенням на 2 згини.

Це дає загалом 5×2=10 оцінок продуктивності, які потім усереднюються, забезпечуючи більш надійну статистичну оцінку.

Продуктивність регресії оцінюється трьома метриками:

- Середньоквадратична помилка (RMSE) - відповідає очікуваному значенню квадратичної втрати в одиницях залежної змінної. Нижчі значення (ближче до діапазону F1-оцінки: $0 \leq F1 \leq 1$) вказують на точніші прогнози.

- Пояснена дисперсія (EVAR) вимірює значущість дисперсії помилки відносно дисперсії залежної змінної.

- Коефіцієнт детермінації (R2) вимірює ймовірність коректного прогнозування майбутніх прикладів.

EVAR та R2 знаходяться в діапазоні $(-\infty, 1]$, де вищі значення свідчать про більш точні моделі.

2.3. Набори даних та генерація графів

Метою цього розділу є опис наборів даних, використаних для емпіричної оцінки запропонованого фреймворку, спрямованого на кількісну оцінку впливу структурних властивостей на вразливість графів до атак деанонімізації. Дослідження включає аналіз реальних мережевих наборів даних та сімей синтетичних графів, згенерованих для контролю специфічних мережевих метрик.

2.3.1. Емпіричні соціальні мережі

Для забезпечення репрезентативності та різноманітності типів мереж, було обрано чотири публічно доступні набори даних, що представляють реальні соціальні мережі.

- fb107 - мережа его-соціальних кіл із платформи Facebook.

- ca-GrQc - мережа співпраці (коавторства) вчених у галузі загальної теорії відносності та квантової космології.

- soc-anybeat - мережа взаємодій в онлайн-спільноті Anybeat.

- soc-gplus - мережа підписників із соціальної мережі Google+.

Основні топологічні властивості цих наборів даних узагальнені в таблиці 2.1.

Таблиця 2.1.

Властивості графів реальних та синтетичних мережевих наборів даних

Network	space	$\|N\|$	$\|E\|$	$\bar{d}$	r	C	κ	degree-1 (%)
fb107	original	1034	26749	0.0500	0.4316	0.5045	2.9517	1.45
	0K	1034	26749	0.0501	-0.0029	0.0501	2.0210	0.0
	1K	1034	26749	0.0501	-0.0961	0.1466	2.1965	1.45
	2K	1034	26749	0.0501	0.4316	0.3161	2.4020	1.45
	ERGM-apl	1034	26749	0.0501	0.0017	0.0504	2.0193	0.0
	ERGM-cc	1034	26749	0.0501	0.4293	0.5038	2.8796	0.57
	ERGM-dc	1034	26749	0.0501	0.3747	0.1627	2.1197	0.0
	LF (m=2)	1034	2066	0.0039	0.1425	0.2173	10.2155	0.0
	LF (m=5)	1034	5165	0.0097	0.2308	0.2463	5.5336	0.0
	LF (m=10)	1034	10330	0.0193	0.2733	0.2164	3.6806	0.0
caGrQc	original	5242	14496	0.0011	0.6592	0.6298	3.8047	22.83
	0K	5242	14496	0.0011	-0.0011	0.0010	5.2155	2.22
	1K	5241	14484	0.0011	-0.0355	0.0077	4.0002	22.83
	2K	5241	14484	0.0011	0.6593	0.2710	1.0410	22.83
	ERGM-apl	5241	14484	0.0011	0.0390	0.0064	5.4390	0.02
	ERGM-cc	4507	14484	0.0014	0.6804	0.6278	5.6361	10.43
	ERGM-dc	5237	14484	0.0011	0.4547	0.0790	5.5294	0.98
	LF (m=2)	5242	10482	0.0008	0.1536	0.2132	13.0612	0.0
	LF (m=5)	5242	26205	0.0019	0.24	0.2348	7.1527	0.0
	LF (m=10)	5242	52410	0.0038	0.2771	0.1895	4.7513	0.0
soc-anybeat	original	12645	49132	0.0006	-0.1234	0.0217	3.1715	49.51
	0K	12645	49132	0.0006	-0.0001	0.0006	4.8365	0.33
	1K	12645	49132	0.0006	-0.1232	0.0149	2.8779	49.50
	2K	12645	49132	0.0006	-0.1234	0.0176	2.4943	49.50
	ERGM-apl	12635	49132	0.0006	-0.0572	0.0018	3.2206	0.61
	ERGM-cc	12582	49132	0.0006	0.2285	0.1877	4.9853	2.57
	ERGM-dc	12459	49132	0.0006	-0.0831	0.0158	3.3204	8.93
soc-gplus	original	23628	39194	0.0001	-0.3885	0.0037	2.2082	69.16
	0K	23628	39194	0.0001	0.0009	0.0001	7.7045	12.46
	1K	23628	39194	0.0001	-0.3514	0.0137	3.1760	69.16
	2K	23628	39194	0.0001	-0.3885	0.0018	3.8620	69.16
	ERGM-apl	22544	39194	0.0001	-0.0729	0.0004	4.5236	15.32
	ERGM-cc	17784	39194	0.0002	-0.0651	0.0337	5.8122	39.76
	ERGM-dc	22042	39194	0.0001	-0.2407	0.0024	4.0795	30.52

2.3.2 Генерація синтетичних мереж

З метою забезпечення контролю над специфічними структурними характеристиками та дослідження впливу як незалежних, так і взаємозалежних структурних сил, було згенеровано сімейства синтетичних графів. Використовувалися три різні методики генерації, кожна з яких дозволяє контролювати певний простір метрик графа.

Серія dK-випадкових графів представляє набір дескриптивних статистичних метрик, які систематично моделюють топологічні обмеження, пов'язані зі ступенем вузла. Відомо, що зі зростанням порядку d (тобто dK) графи стають менш випадковими та більш структурованими. Хоча розподіли ступенів [28] є важливими для захоплення основних властивостей графа, вони, як правило, не відтворюють інші критичні метрики, такі як коефіцієнт кластеризації.

Для аналізу графів із контрольованим коефіцієнтом кластеризації, що відповідає характеристикам реальних мереж, використовується модель ERGM (Exponential Random Graph Models) [21]. ERGM — це зрілий підхід у соціальному мережевому аналізі, призначений для визначення ймовірності розподілу для мереж у межах заданого набору структурних параметрів. Їхня основна мета — описати структурні та локальні сили (наприклад, гомофілію, кореляцію ступеня, кластеризацію), що формують загальну топологію мережі.

Наш інтерес до ERGM полягає у симуляції графів, які зберігають набір структурної інформації з оригінального графа, забезпечуючи генерацію різноманітних графових структур. Для імплементації використовувалися мова R та пакет statnet. Ми зосередилися на трьох структурних аспектах: коефіцієнт кластеризації, середня довжина шляху та кореляція ступеня (асортативність).

ERGM-cc (кластеризація) - використовувалися параметри ребер (вимірює ймовірність зв'язку) та трикутників (враховує кількість трійок/кластеризацію).

ERGM-apl (середня довжина шляху) - використовувалися параметри ребер та двопутів (кількість двопутів).

ERGM-dc (кореляція ступеня) - використовувалися терміни ребер та degcor (кореляція ступеня всіх пар зв'язаних вузлів).

Незважаючи на широке застосування, наш досвід показав, що ERGM можуть мати обмеження щодо генерації графів із бажаним діапазоном коефіцієнтів асортативності ступеня.

Для забезпечення контролю над асортативністю ступеня та охоплення відповідного простору графів використовується модель Лідер-Наслідник (Leader-Follower, LF).

Ця модель контролює дві популяції вузлів:

- Наслідники (Followers) - випадково вибирають ребра, що призводить до спонтанної поведінки переваги приєднання.

- Лідери (Leaders) - демонструють поведінку анти-переваги, встановлюючи зв'язки з вузлами нижчого ступеня.

Алгоритм генерації вимагає трьох параметрів:

p: Частка лідерів.

m: Максимальна кількість можливих з'єднань для вузла.

l: Ступінь інформації про сусідство, доступної для вузла при початковому з'єднанні (встановлено l=1 для спрощення).

Ефект параметра p:

- При p=0 (відсутність лідерів) згенеровані мережі демонструють сильну перевагу приєднання, що призводить до негативної асортативності ступеня.

- При p=1 отримані графи мають позитивну асортативність ступеня.

Експериментально підтверджено, що p пропорційний асортативності ступеня (як показано на рис. 2.3), при цьому також спостерігається лінійний зв'язок між транзитивністю та асортативністю ступеня. На рис. 2.3 представлено множинні регресійні моделі як функцію від m, де пунктирні,

крапкові та штрих-пунктирні лінії відповідно позначають моделі для m=2, m=5 та m=10.

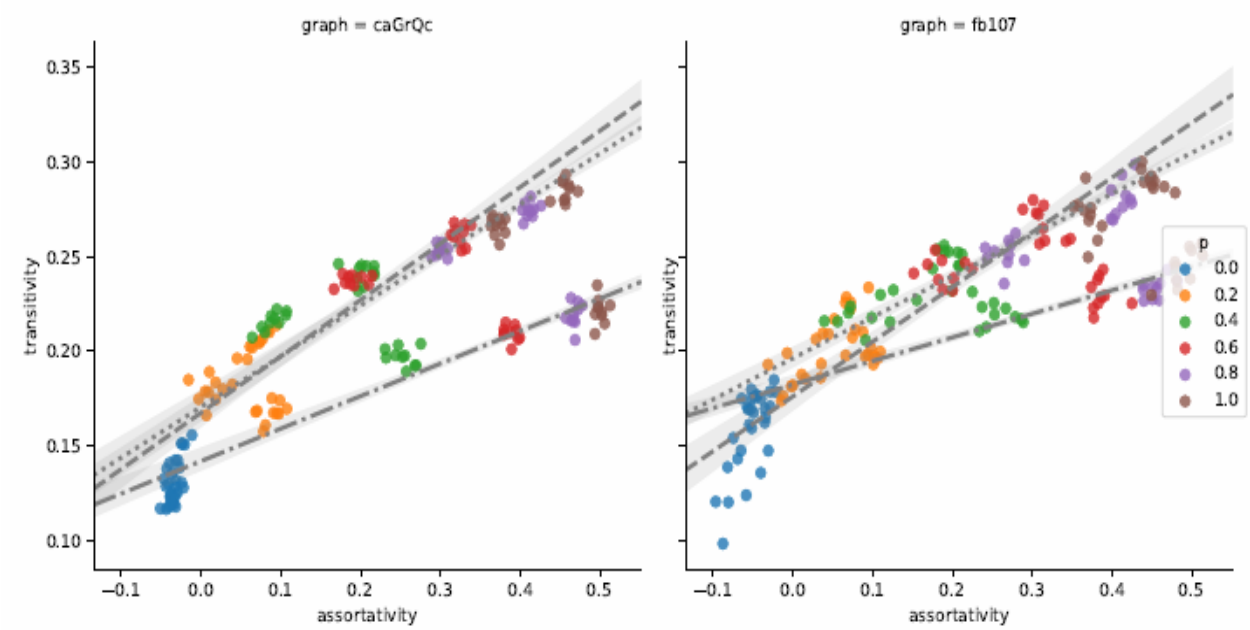


Рис. 2.3. Транзитивність (C) та асортативність (r) на графах моделі Leader-Follower

2.4. Емпірична оцінка фреймворку та аналіз вразливості графів

Емпірична оцінка, представлена тут, має подвійну мету. По-перше, вона спрямована на валідацію корисності запропонованого фреймворку. По-друге, вона використовує цей фреймворк для відповіді на ключове дослідницьке питання: Які саме структурні властивості роблять певні набори даних графів більш вразливими до атак деанонімізації, ніж інші?

Дослідження включає оцінку вразливості колекції реальних та синтетичних графів. Вразливість графа кількісно визначається як функція успішної повторної ідентифікації вузлів. Далі проводиться детальний аналіз взаємозв'язку між вразливістю графа та різними структурними силами, використовуючи як інформаційно-теоретичні, так і прогностичні вимірювання.

2.4.1 Аналіз вразливості графів

Показник вразливості графа визначається як F1-оцінка, що відображає точність прогнозування структурної еквівалентності пар вузлів. На рис. 2.4 представлено порівняння показників вразливості для різних синтетичних просторів графів, що дозволяє зробити три ключові спостереження.

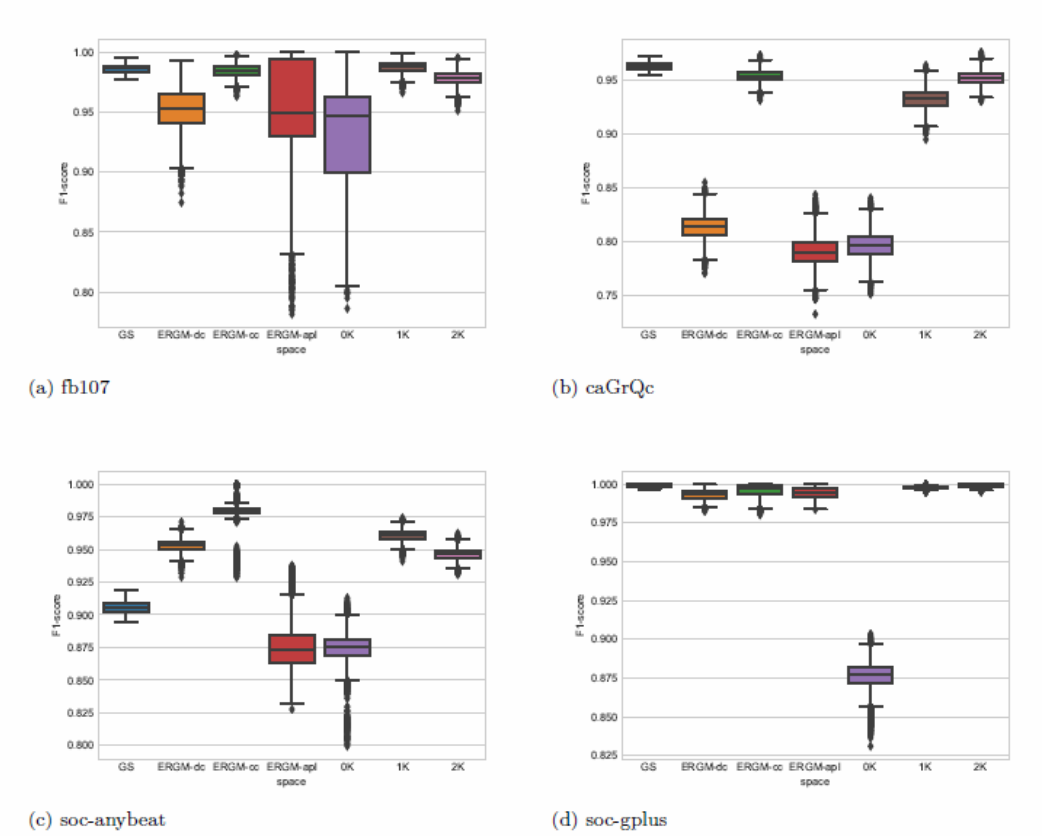


Рис. 2.4. Порівняння F1-оцінки в різних просторах вхідних графів. Кожна оцінка представляє результати прогнозування 5×2 зразків крос-валідації, усереднені для 100 синтетичних графів на простір.

1. Порівняння просторів ERGM. Середнє значення показника вразливості зростає у послідовності: ERGM-apl < ERGM-dc < ERGM-cc. При цьому ERGM-apl демонструє найширший діапазон варіативності. Це спостереження дозволяє припустити, що збереження асортативності та транзитивності як метрик корисності в процесах анонімізації може

потенційно компрометувати анонімність вузлів у графі. Це явище є новим, наскільки нам відомо.

Додатковий аналіз графів LF (де асортативність контролюється, див. рис. 2.5) показує, що показник вразливості досягає локального максимуму при малих значеннях параметра p та спадає до локальних мінімумів у діапазоні $p \in [0.4, 0.6]$. Оскільки p корелює з асортативністю, яка, своєю чергою, корелює з транзитивністю (рис. 2.3), існує висока ймовірність того, що асортативність і транзитивність є значущими факторами вразливості графів.

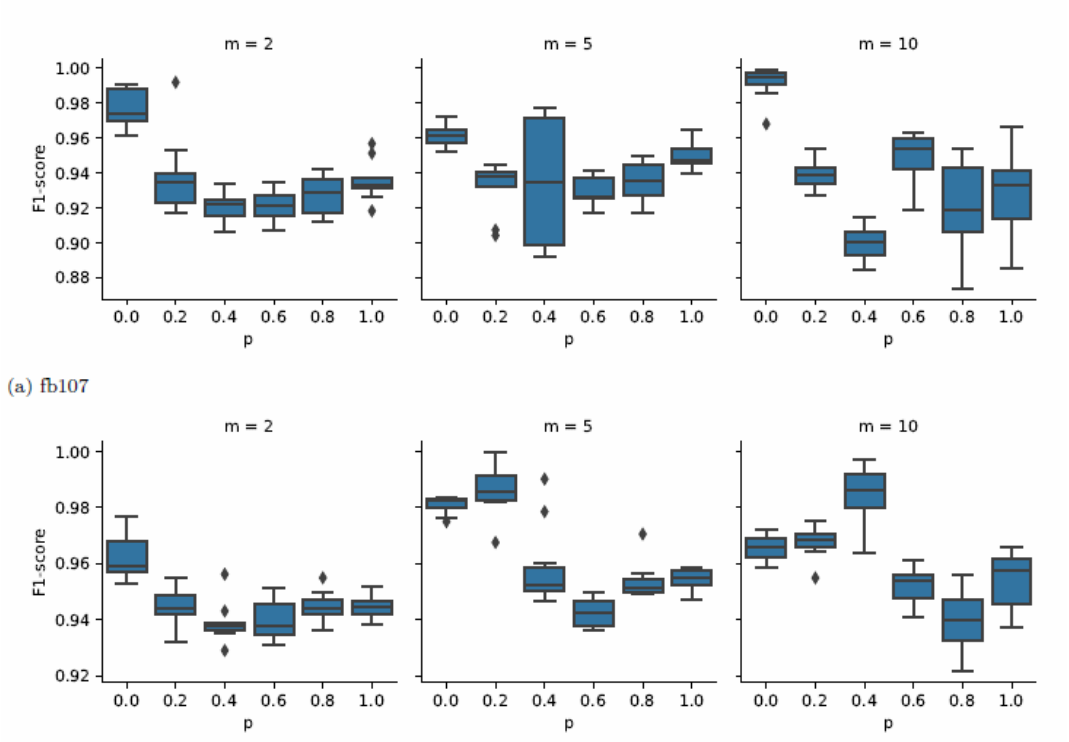


Рис. 2.5. Порівняння F1-оцінки на графах моделі Leader-Follower

Тут ми згенерували низку LF-графів, варіюючи параметри p та m. Параметр p — це ймовірність того, що мережа демонструє тенденцію до анти-преваги приєднання, яка позитивно корелює з асортативністю ступеня (r). Параметр m пропорційний щільності ($\bar{d}$) мережі. Кожна оцінка представляє результати прогнозування 5×2 зразків крос-валідації, які

усереднені для 10 синтетичних графів, згенерованих із заданими параметрами p та m.

2. Пріоритет транзитивності над розподілом ступенів. Як видно на рис. 2.4a – 2.4d, деякі графи, згенеровані за допомогою ERGM, виявляються більш вразливими, ніж графи 1K або 2K, незважаючи на те, що вони не відтворюють оригінальний розподіл ступенів. Традиційно вважалося, що приватність зростає зі збуренням розподілу ступенів. Наші результати вказують на те, що інша метрика, а саме транзитивність, є більш розкривальною, ніж розподіл ступенів.

Зокрема, для мережі soc-anybeat (рис. 2.4c) середня вразливість графів ERGM-cc вища, ніж середня вразливість графів 1K та 2K. Це відбувається попри значну відмінність у розподілі ступенів (таблиця 2.1): оригінальний граф (і, відповідно, 1K та 2K) мав 49.5% вузлів зі ступенем 1, тоді як графи ERGM-cc мають лише 2.57%. Цей результат свідчить про наявність структурних властивостей, незалежних від розподілу ступенів, які ставлять під загрозу приватність вузлів, що ускладнює завдання захисту приватності графів.

3. Підтвердження відомих явищ:

- Оригінальний граф (GS) є найбільш вразливим у всіх випадках, окрім мережі soc-anybeat.

- Графи 0K (Ердеша-Реньї) є, як і очікувалося, найменш вразливими, але й найменш репрезентативними.

- Показники вразливості графів 1K та 2K є найближчими до оригіналу. Це підтверджує раніше відомі результати про те, що графи dK вищого порядку можуть витікати значну структурну інформацію, що обмежує їхню ефективність для цілей анонімізації.

2.4.2. Аналіз причинності на основі пояснювального моделювання

Для виявлення залежностей між показником вразливості та макрорівневими структурними властивостями графа використовувалися

методи пояснювального моделювання. Аналіз ґрунтується на двох метриках важливості: F-тесті (лінійна залежність) та взаємній інформації (MI) (загальна нелінійна залежність). Результати F-тесту та MI подано на рис. 2.6.

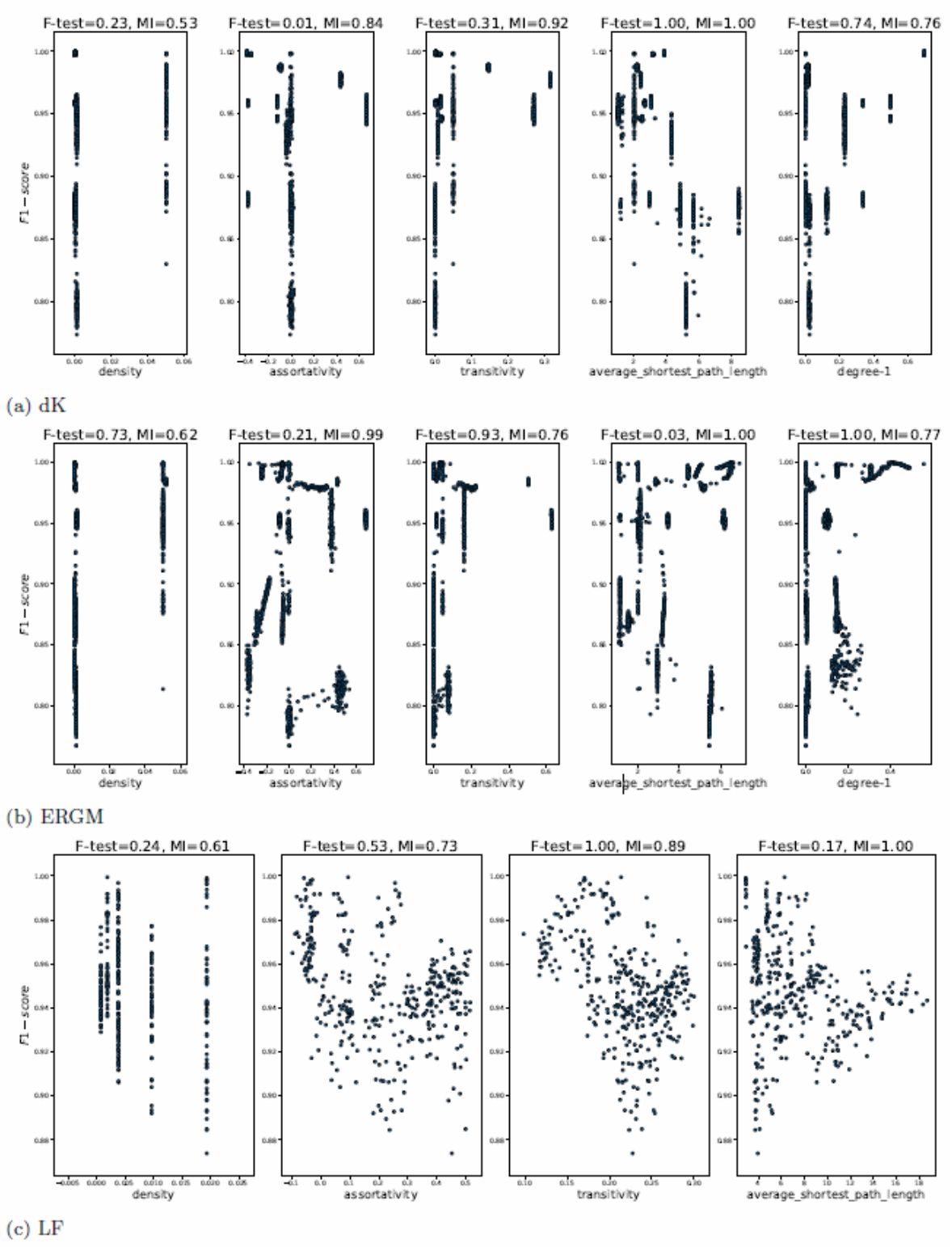


Рис. 2.6. Порівняння між показниками F-тесту та взаємної інформації

Асортативність демонструє відносно низьке значення F-тесту (слабка лінійна залежність), але значно вище значення MI, що вказує на сильну нелінійну залежність з показником вразливості.

Транзитивність також має вищу MI у просторі dK, підкреслюючи нелінійний характер її залежності.

Середня довжина шляху демонструє максимальне значення MI в обох просторах (dK та ERGM).

Частка вузлів зі ступенем 1 є сильним кандидатом на залежність, показуючи високі значення як F-тесту, так і MI. Це частково пояснює низьку вразливість оригінального графа soc-anybeat: велика частка нерозрізненних вузлів зі ступенем 1 зменшує інформаційну розкривність позицій їхніх сусідів.

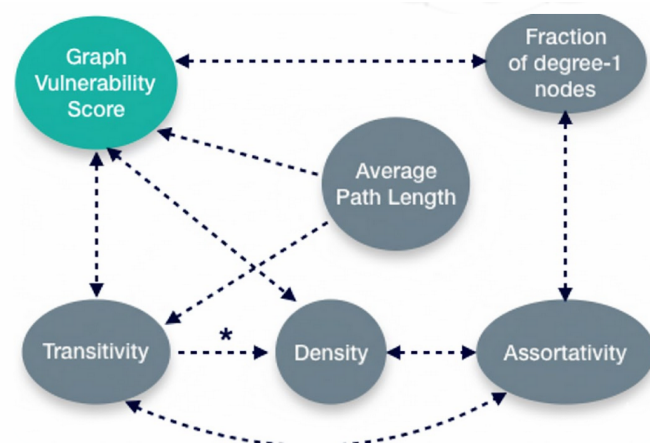


Рис. 2.7. Спрямований ациклічний граф Перла

Ми використовуємо фреймворк Перла [22] для причинного висновку щодо метрик графа та показника вразливості графа. Напрямок ребра позначає відношення причина-наслідок, де стрілка вказує на наслідок. Позначення "*" на ребрі вказує на припущення алгоритму причинного висновку про справжні причинні зв'язки.

Фреймворк Перла (з алгоритмом IC*) був використаний для виведення причинних шляхів (рис. 2.7). Спрямований ациклічний граф (DAG) виявив такі явища:

- Транзитивність, щільність, середня довжина шляху та частка вузлів зі ступенем 1 мають пряму статистичну залежність з показником вразливості. Проте ці залежності не були визначені як справжні причинні відносини алгоритмом Перла, що може бути спричинено неспостережуваними (сплутуючими) змінними.

- Асортативність не має прямої статистичної залежності з вразливістю, але сплутує інші метрики (транзитивність, щільність, ступінь 1) для впливу на вразливість. Цей сплутуючий ефект значною мірою захоплюється транзитивністю, яка потім впливає на вразливість графа.

Таким чином, вразливість можна пояснити як лінійну функцію частки вузлів зі ступенем 1 та нелінійну функцію інших структурних метрик.

2.4.3 Аналіз продуктивності на основі прогностичного моделювання

Прогностичне моделювання дозволяє оцінити рівень узагальнення, підганяючи регресійні моделі до показників вразливості, використовуючи структурні властивості як ознаки. Крос-валідація проводилася двома способами: розділенням за просторами графів (dK, ERGM, LF) та розділенням за сім'ями оригінальних графів (5-кратна крос-валідація).

Таблиця 2.2.

Порівняння точності прогнозування F1-оцінки

Тренувальний набір	**Тестовий набір**	**RMSE**	**EVAR**	**R2S**
Synthetic	Original	6.3156	-0.0020	-2.3627
Synthetic	Synthetic	0.0821	0.0578	-3.9392
dK	dK	0.0533	0.2162	-4.5240
dK	ERGM	0.0753	-0.1867	-0.1994
dK	LF	0.0392	-0.2170	-0.4052
ERGM	ERGM	0.1372	-1.4304	-48.9614
ERGM	dK	0.0646	0.4552	0.3288
ERGM	LF	0.0297	-2.7625	-2.9118

Ми використовуємо модель лінійної регресії для прогнозування F1-оцінки в різних просторах графів для крос-валідації, застосовуючи

структурні властивості, що включають щільність, асортативність, транзитивність, середню довжину найкоротшого шляху та відсоток вузлів зі ступенем.

Простір dK демонструє кращу точність, ніж ERGM, при внутрішньопросторовій крос-валідації. Наприклад, RMSEdK−dK (0.05) нижче, ніж RMSEERGM−ERGM (0.13), а EVARdK−dK>0. Це означає, що dK простір є більш вразливим до атак повторної ідентифікації.

Тренування на просторах ERGM з тестуванням на dK (EVARERGM−dK) дає кращі результати, ніж навпаки (EVARdK−ERGM). Це свідчить про те, що простір ERGM є багатшим навчальним набором через більшу варіативність значень метрик графа та вразливості, тоді як простір dK обмежує значення ознак.

dK є кращим навчальним набором для тестування на LF, ніж ERGM. Це підтверджує близькість цих просторів (оскільки асортативність у LF є агрегатною мірою спільного розподілу ступеня, що визначає dK) і знову вказує на сильний вплив асортативності та кластеризації на вразливість графа.

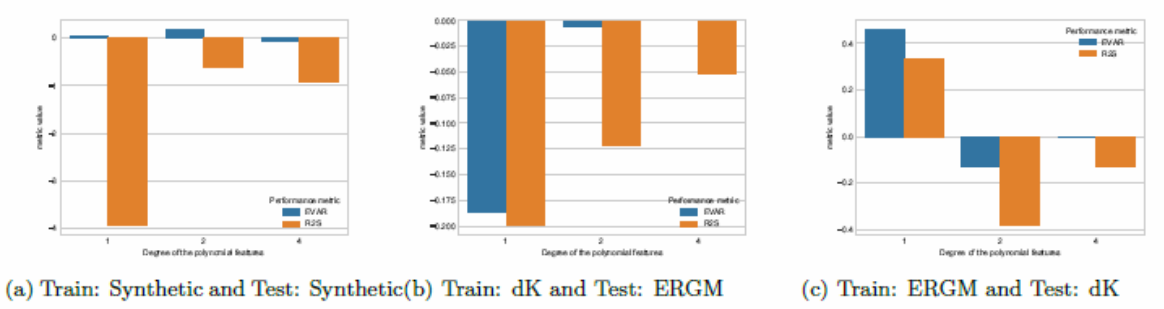
Для обліку нелінійних залежностей структурні властивості були трансформовані в поліноміальний простір ознак (лінійний, квадратичний та квартичний ступені).

У моделі, навченій на всіх синтетичних даних, R2S значно зростає у квадратичному поліноміальному просторі (рис. 2.8 a). Це підтверджує наявність комбінованого ефекту структурних сил на пояснення вразливості графа.

У моделях, навчених окремо на просторах dK (рис. 2.8b) та ERGM (рис. 2.8c), прогностична потужність зростає зі збільшенням ступеня поліноміальних ознак (EVAR та R2S зростають).

У моделі тренування ERGM (рис. 2.8c) прогностична потужність лінійної моделі послаблюється після додавання взаємодійних термінів у квадратичному просторі, на відміну від моделі dK. Це свідчить про те, що

трансформація в новий простір ознак може послаблювати корисність, збережену в ERGM, що призводить до відносно гіршої прогностичної моделі.



(a) Train: Synthetic and Test: Synthetic(b) Train: dK and Test: ERGM (c) Train: ERGM and Test: dK

Рис. 2.8. Метрики продуктивності залежно від степеня поліноміальних ознак

Ми генеруємо поліноміальні ознаки та ознаки взаємодії з оригінального простору ознак, що описують структурні властивості графа, таким чином, що новий простір ознак включає всі поліноміальні комбінації ознак під заданим степенем полінома. Метрики продуктивності базуються на відповідній поліноміальній регресійній моделі.

2.5. Синтез результатів та перспективи фреймворку для оцінки вразливості мереж

Цей розділ присвячений вивченню та наданню відповіді на ключове питання: Які структурні властивості графів визначають підвищену вразливість мережевих наборів даних до атак повторної ідентифікації вузлів? На відміну від попередніх досліджень, ми зосереджуємося на внутрішній вразливості оригінального (неанонімізованого) графа, а не на його модифікованих версіях. Відповідь на це питання є критично важливою для:

- Визначення ризику, пов'язаного з публікацією оригінальних даних.

- Керівництва анонімізацією, тобто надання рекомендацій для вибору методів анонімізації, які забезпечують оптимальний компроміс між корисністю даних та приватністю.

Ми пропонуємо та емпірично досліджуємо фреймворк, призначений для визначення взаємозв'язків між вразливістю графа та його топологічними властивостями. Ключові компоненти фреймворку:

- Кількісна оцінка вразливості вимірюється успішністю атаки повторної ідентифікації.

- Кількісна оцінка взаємозв'язку - вимірює залежність між показником вразливості та обраним набором метрик графа, із застосуванням надійного інструментарію для аналізу причинності.

Використання синтетичних графів із контрольованими властивостями дозволило виявити декілька емпіричних закономірностей:

1. Вплив транзитивності та асортативності.

Застосована модель атаки виявила сильну статистичну залежність між показником вразливості, транзитивністю та асортативністю. Це означає, що успішні алгоритми анонімізації не повинні прагнути зберегти ці метрики. Розробка методів, які цілеспрямовано збурюють асортативність та транзитивність, відкриває нові перспективи для підвищення приватності графа проти сильних атак деанонімізації.

2. Нелінійний характер залежностей.

За винятком частки вузлів зі ступенем 1, лінійний зв'язок між показником вразливості та іншими структурними метриками графа відсутній. Це пояснюється взаємозалежністю найбільш релевантних мережевих метрик. Застосування більш повного набору метрик у причинній моделі Перла може допомогти виявити складніші причинно-наслідкові зв'язки.

3. Недостатність збурення розподілу ступенів.

Попередні дослідження вказували на те, що щільність та збереження розподілу ступенів є важливими детермінантами вразливості, і їхнє збурення вважалося необхідною умовою для анонімізації. Однак наше дослідження демонструє, що ця умова є недостатньою: інші мережеві властивості, незалежні від розподілу ступенів, можуть скомпрометувати приватність вузлів. Це є тривожним висновком, оскільки він вказує на те, що

забезпечення приватності графів є значно складнішим завданням, ніж передбачалося раніше.

Існує ймовірність, що виявлені залежності можуть залежати від використаних інструментів, зокрема від сильної моделі атаки та конкретного представлення ознак вузлів (NDD). Слабка модель атаки або інше представлення ознак може призвести до зміни показників вразливості та, відповідно, до інших виявлених взаємозв'язків.

Запропонований фреймворк є гнучким і може бути використаний для дослідження:

- Причинного зв'язку між параметрами моделі атаки та вразливістю.

- Порівняння сили різних моделей атак.

- Визначення інших розкривальних метрик графа за умови, що інформація про ступінь вузла невідома атакуючому.

Емпіричні дані підтверджують, що для використаної атаки асортативність і транзитивність є високоінформативними щодо ідентичностей вузлів. Це дослідження може бути розширене для оцінки вразливості динамічних графів або графів з атрибутами вузлів та ребер.

Досліджуваний фреймворк долає розрив між теоретичними оцінками приватності та практичним застосуванням, надаючи уніфіковану платформу для:

- Розробки нових методологій анонімізації та деанонімізації.

- Кількісної оцінки вразливості графів.

- Емпіричного калібрування теоретичних оцінок (наприклад, методів на основі диференціальної приватності).

Власники даних можуть використовувати цей зворотний зв'язок для вибору компромісу між прийнятною вразливістю та необхідною корисністю (вираженою через метрики графа) для цілеспрямованого перепроектування алгоритмів анонімізації.

Висновки до розділу

У цьому розділі було розроблено та емпірично оцінено фреймворк для кількісної оцінки вразливості графів до атак деанонімізації із застосуванням машинного навчання. Фреймворк включає модель загрози, яка імітує поведінку зловмисника, та алгоритм атаки, що використовує машинне навчання для ідентифікації вузлів. Проведено детальний аналіз взаємозв'язку між традиційними топологічними метриками графа та його реальною вразливістю, використовуючи як пояснювальне, так і прогностичне моделювання. Результати причинно-наслідкового аналізу підтвердили, що певні топологічні характеристики корелюють із високим рівнем вразливості, дозволяючи спрогнозувати ризик ще до проведення повноцінної атаки. Оцінка на емпіричних та синтетичних наборах даних продемонструвала високу продуктивність прогностичної моделі. Синтез результатів підтвердив життєздатність фреймворку як потужного інструменту для оцінки ризиків та розробки контрзаходів у соціальних мережах.

РОЗДІЛ 3. МОДЕЛЮВАННЯ АКТИВНОСТІ В СОЦІАЛЬНИХ МЕРЕЖАХ З УРАХУВАННЯМ ЕКЗОГЕННИХ ФАКТОРІВ

3.1. Розробка модульної архітектури для імітації процесів в соціальних мережах

Платформи соціальних мереж функціонують як віртуальні простори для онлайн-комунікації та, як правило, зазнають впливу екзогенних факторів, що походять із зовнішнього світу. Ці зовнішні впливи (наприклад, середовищні, політичні чи культурні події) часто не реєструються безпосередньо в системі соціальної мережі. Наприклад, реакція користувачів X на стихійні лиха (як-от землетрус на Гаїті у 2010 році чи ураган Ірма) або на політичні рішення, як-от деплатформізація чи модерація контенту, демонструє цей зв'язок. Проте, кількісна оцінка та відокремлення впливу цих зовнішніх сил від інших факторів є складним завданням. Виділення та аналіз цих екзогенних впливів є ключовим для глибшого розуміння процесів поширення інформації на платформах соціальних мереж.

Основна гіпотеза полягає в тому, що інкорпорування екзогенних подій дозволить підвищити точність прогнозування активності користувачів у соціальних мережах. Ми досліджуємо можливість точного моделювання активності (наприклад, у Twitter (X)) за допомогою сигналів, отриманих з інших платформ. Особливий інтерес становить прогнозування в контексті несподіваних подій (криз), коли користувачі реагують на неочікувані новини непередбачуваними способами.

Мета – прогнозування активності X з використанням екзогенних джерел (рис. 3.1). Наш підхід передбачає симуляцію активності з високою деталізацією (дрібнішою гранулярністю), а саме прогнозування:

- Часу публікації твіту.
- Тематики повідомлення.
- Ідентичності автора.

- Ідентичності ретвітера.

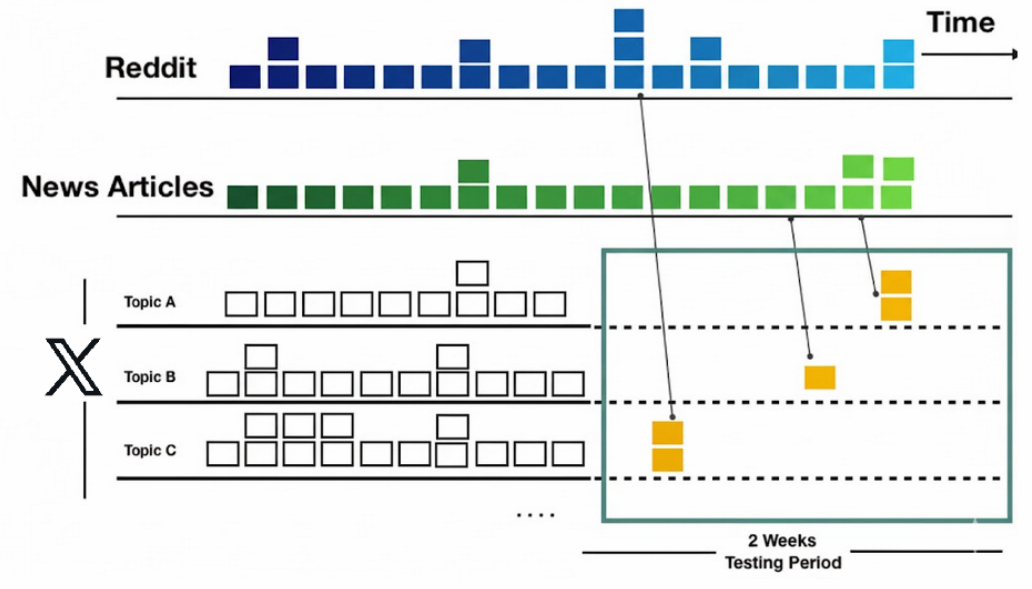


Рис. 3.1. Прогнозування тематичної активності мережі X (X) з використанням екзогенних даних

Це відрізняє наше завдання від простого прогнозування загального щоденного обсягу активності.

Моделювання здійснюється з щоденною гранулярністю протягом двотижневого інтервалу. На відміну від більшості підходів машинного навчання, ми прагнемо прогнозувати активність дня d без використання даних активності дня d−1.

Емпірична оцінка проводиться на прикладі прогнозування активності X, пов'язаної з політичною кризою у Венесуелі з січня до кінця лютого 2019 року. Періоди криз є складними для прогнозування, оскільки вони містять мало повторюваних патернів поведінки. Як екзогенні дані використовуються новинні статті та відповідний сабреддит.

В даному розділі представлено дизайн симулятора, здатного імітувати пікові значення реальної активності. Успішний дизайн передбачає модуляризацію для спеціалізованого прогнозування конкретних підзадач

(наприклад, прогнозування кількості інформаційних каскадів, а також їхнього розміру та зростання).

Доведено, що прогнози з використанням екзогенних даних поточного дня перевершують моделі, які базуються на даних попереднього дня. Це ставить під сумнів усталену практику використання лише історичної інформації та доводить, що користувачі соціальних мереж швидко реагують на поточні події та новини в режимі реального часу.

Обговорюється, як різні екзогенні джерела даних є корисними для прогнозування різних тем у межах однієї широкої розмови.

3.2. Дизайн модульного симулятора для прогнозування активності соціальної мережі

Активність користувачів у X характеризується взаємодією з різноманітними темами та часто залежить від екзогенних подій. Отже, точна симуляція активності X вимагає інтеграції сигналів про реальні події, отриманих із зовнішніх (екзогенних) джерел даних.

Цей розділ представляє архітектуру симулятора, розробленого для точного прогнозування активності X з використанням екзогенних даних. Основна мета полягає у прогнозуванні двотижневого інтервалу активності X, не використовуючи при цьому фактичні дані активності за цей період. Прогноз повинен надавати деталізовану інформацію: тип дії (твіт чи ретвіт), тематику, автора та часову мітку повідомлення.

3.2.1. Представлення модульної архітектури

Симулятор реалізовано на основі двомодульної архітектури (рис. 3.2):

- Модуль початкових подій (Seed Module) - приймає історичну активність X та дані з інших платформ, прогнозуючи щоденну кількість початкових твітів (які ми називаємо початковими подіями) для кожної теми.

- Модуль каскадів (Cascade Module) - використовує вихідні дані модуля початкових подій для генерації каскадів ретвітів у відповідь на початкові події. Цей модуль розширює функціональність попередніх рішень для прогнозування інформаційних каскадів X, призначаючи кожному повідомленню у каскаді користувача та день ретвіту.

3.2.2 Модуль прогнозування початкових подій

Емпіричний досвід показав, що точне прогнозування щоденного обсягу твітів є критичним для забезпечення загальної точності симуляції. Отже, цей спеціалізований модуль відповідає за прогнозування щоденної кількості твітів за темою та їхнього автора.

Спочатку була спроба прогнозувати загальний обсяг твітів з подальшим розподілом за темами. Однак прогнозування щоденної частки твітів для тем, які зазнають впливу непередбачуваних подій (наприклад, масові протести чи політичні інтервенції), виявилося надзвичайно складним. Тому було прийнято рішення безпосередньо прогнозувати щоденну кількість твітів за темою.

Модуль реалізовано за допомогою нейронної мережі, навченої на історичній активності X (кількість твітів на день) та відповідному екзогенному сигналі.

Початково розглядалося використання активності X дня d для прогнозування дня d+1, але це призводило до накопичення помилок і не могло прогнозувати спалахи активності. Тому було прийнято рішення використовувати виключно екзогенну активність для прогнозування.

Вектор ознак включає one-hot кодування теми та щоденну кількість екзогенних подій (новостих статей та повідомлень Reddit), пов'язаних із темою в день d. Експерименти проводилися з окремими екзогенними джерелами (лише новини або лише Reddit).

Цільова змінна - кількість твітів, пов'язаних із темою в день d+1.

Архітектура мережі наступна. Нейронна мережа з 3 прихованими шарами (розміри: 15, 10 та 5 нейронів), оптимізатором Adam та функцією втрат середньоквадратичної помилки. Найкращі гіперпараметри визначалися за допомогою 5-кратної крос-валідації.

Користувачі призначаються прогнозованим твітам випадковим чином з імовірністю, пропорційною до їхнього балу поширення [202]. Ця евристика краще відображає активність впливових користувачів.

3.2.3. Модуль генерації дерев ретвітів (каскадів)

Цей модуль використовується для генерації дерев ретвітів (інформаційних каскадів) у відповідь на прогнозовані твіти, забезпечуючи дрібногранулярні прогнози та імітуючи шаблони взаємодії користувачів.

Через обмеження X API, які не дозволяють визначити, чи є ретвіт відповіддю на інший ретвіт, реконструкція істинного дерева ретвітів є нетривіальною. Ми використовуємо алгоритм дифузії, заснований на часі, який наближає структуру каскаду шляхом з'єднання ретвіту з попереднім повідомленням на основі мережі підписників X.

Генерація каскаду здійснюється використовуючи ймовірнісний підхід для генерації структури каскаду з користувачами та часовими мітками.

Оскільки модуль початкових подій не враховує нових користувачів, а вони становлять більшість населення X, ми вводимо окремий підмодуль для прогнозування щоденної частки нової залученості користувачів.

Підмодуль використовує ті ж вектори екзогенних ознак, що й модуль початкових подій, і навчає окремі нейронні мережі для прогнозування щоденної кількості нових користувачів за темою.

Нові користувачі призначаються листовим вузлам каскадів, оскільки більшість нових користувачів беруть участь у каскадах саме як кінцеві ретвіти (тобто їхні ретвіти не перепощуються далі). Цей процес повторюється до досягнення прогнозованої щоденної кількості нових користувачів.

3.3. Опис набору даних для відображення подій та тематична класифікація активності в соціальних мережах

Протягом останніх двох десятиліть у венесуельському суспільстві спостерігається глибока соціополітична фрагментація, зумовлена значними розбіжностями в інтересах, ідентичностях та політичних орієнтаціях. Політичний спектр країни переважно поляризований на дві ключові фракції: чавізм (прихильники політичної ідеології покійного президента Уго Чавеса) та античавізм (рішуча опозиція спадщині Чавеса). Незважаючи на контроль політичної системи з боку чавізму під керівництвом Ніколаса Мадуро, країна переживає значний економічний колапс, спричинений неефективним управлінням глобалізацією, недостатніми інвестиціями в інфраструктуру та загальною неефективністю державного управління. Наслідками цього є зростання рівня злочинності та насильства, дефіцит основних товарів, гостра нестача ліків і продовольства, а також безпрецедентна гуманітарна криза.

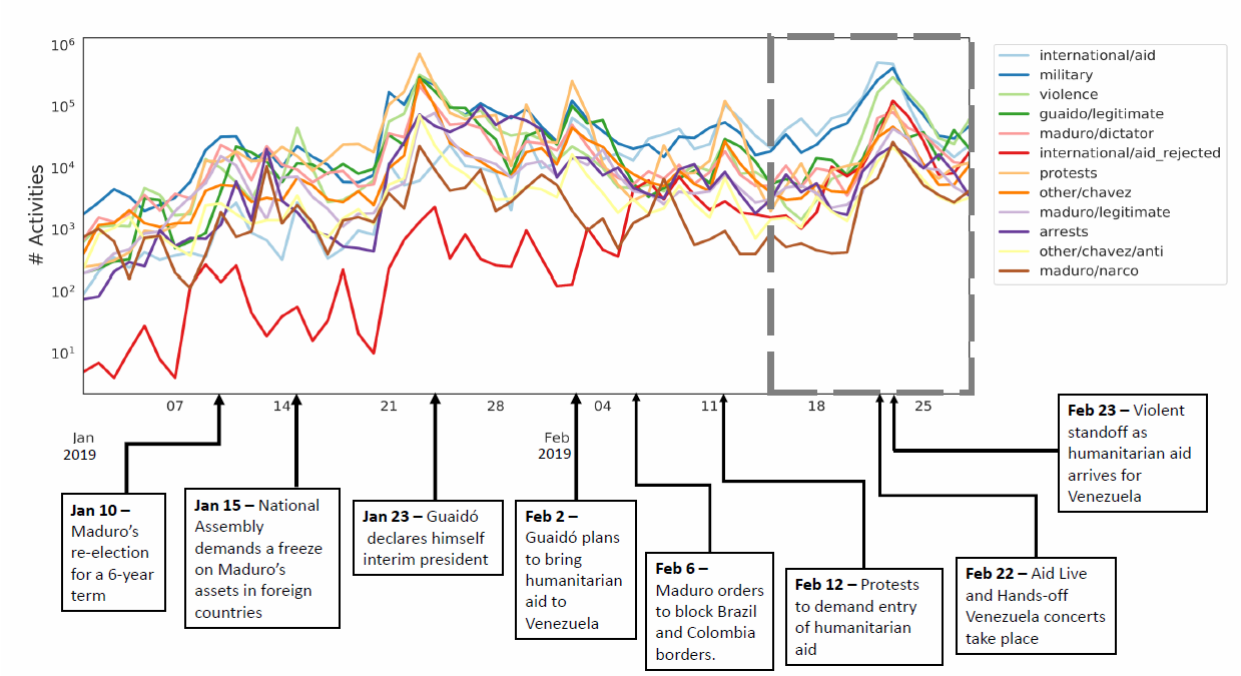


Рис. 3.3. Часові ряди повідомлень X для 12 тем у наборі даних політичних подій

Це дослідження зосереджується на даних, специфічних для періоду високої політичної напруги, що розгорталася у Венесуелі на початку 2019 року. Цей період характеризується загальнонаціональними протестами, військовими операціями та масовими інцидентами насильства й арештів. Стисла хронологія ключових подій представлена на рис. 3.3.

Деталізуємо методологію збору та попередньої обробки даних X, а також екзогенних джерел.

Дані набору даних збиралися протягом двох місяців за допомогою інструменту GNIP API на основі вичерпного списку ключових слів, пов'язаних із політичною кризою (рис. 3.4).

#23Ene, #23Feb, 23 de Enero, 23 de Febrero, Aid Venezuela, #BravoPueblo, Caracas, Maturin, Maracaibo, #Chavismo, #Chavistas, FANB, #FreeVenezuela, #FueraDictadura, Fuerza Venezuela, GNB, #GritemosConBrio, #GuaidoPresidente, #JGuaido, Juan Guaido, #LasCallesSonDelChavismo, Leales siempre traidores nunca, Libertad para Venezuela, Freedom for Venezuela, #VamosBien, #MaduroDictador, #MaduroUsurpador, Nicolas Maduro, #SOSVenezuela, #VenezolanosEnElMundo, Venezuela Aid Live, #WeAreMaduro, Yankee go Home, #HandsOffVenezuela, #FebreroRebelde, #NoMasDictadura, Maduro, #AbajoCadenas, Venezuela Crisis Humanitaria, Maduro Ilegitimo, Guaido, Chacao

Рис. 3.4. Ключові слова, використані для збору даних

Набір даних включає майже 1105000 початкових повідомлень (твіти, відповіді, цитати) від 274 тисяч користувачів та 11000000 ретвіти від 890 тисяч користувачів.

Мовами переважно є іспанська (86%) та англійська (6%). Ідентифікатори користувачів є анонімізованими. Кожен запис містить унікальний ідентифікатор, анонімізований ID користувача, часову мітку, зміст повідомлення та його тип.

Для автоматизації тематичного анотування мільйонів повідомлень було проведено напівконтрольоване завдання класифікації у два етапи:

- Ручне позначення. Спочатку частину набору даних вручну анотували початковий корпус. Після аналізу узгодженості (виміряної коефіцієнтами Каппа Коена та Флісса) початковий список із 49 підтем було звужено до 12

найбільш надійних тем (наприклад, міжнародна/допомога, військові, гуайдо/легітимний). Ці 12 тем продемонстрували зважений середній коефіцієнт Каппа Коена 0.64 та Каппа Флісса 0.7.

На основі ручно розміченого корпусу було навчено багатомовну модель BERT для автоматичного призначення однієї або кількох тем кожному повідомленню. Модель була навчена на 10 тисячах унікальних документах і продемонструвала на тестовому наборі: точність 67%, повноту 66% та F1-оцінку 66%.

Для оцінки взаємодії між активністю X та зовнішніми подіями були зібрані дані з двох екзогенних джерел:

- Reddit. Зібрано обговорення з сабреддиту. Цей ресурс надає альтернативну перспективу політичних обговорень, які можуть не висвітлюватися традиційними ЗМІ. Зібрано майже 5000 постів.

- Новинні статті. Зібрано дані через базу даних геополітичних подій GDELT, яка агрегує машинно-кодовані події з новинних звітів (оновлення кожні 15 хвилин). Запит за терміном "Venezuela" повернув 138,000 URL-адрес джерел.

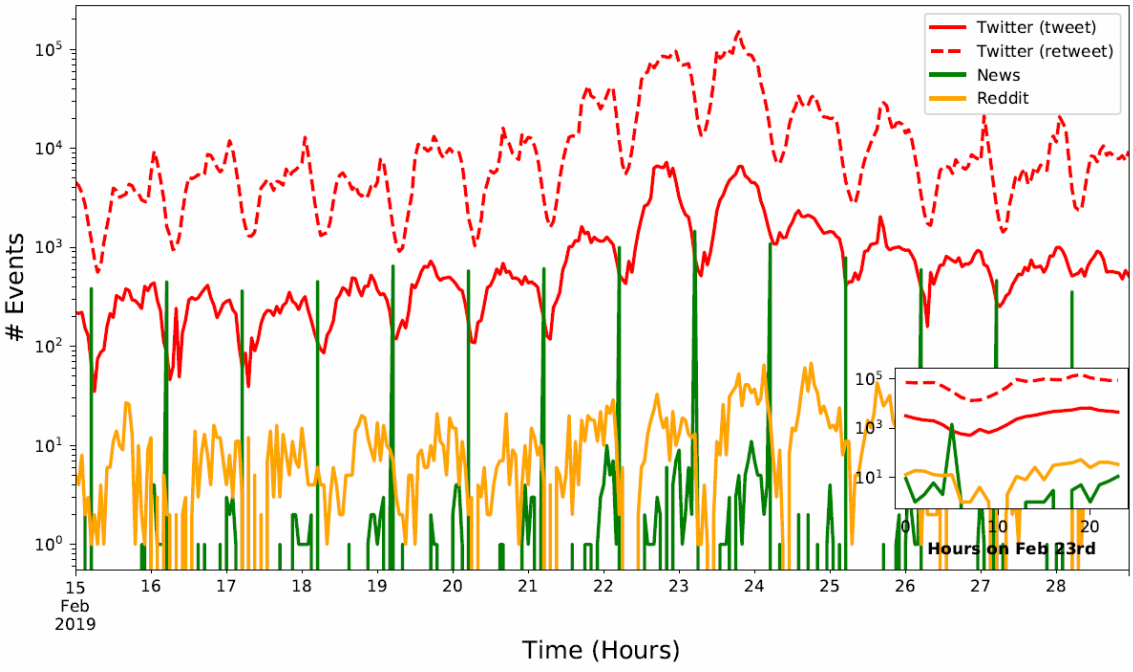


Рис. 3.5. Часові ряди твітів, новинних статей та повідомлень Reddit

Навчена модель BERT була використана для автоматичного призначення 12 тем інтересу вмісту новинних статей та повідомлень Reddit. Після перевірки надійності класифікації було ідентифіковано більше 2 тисяч постів та 31 тисячі коментарів на Reddit, а також 82 тисячі новинних статей, пов'язаних принаймні з однією темою. Рисунок 3.5 ілюструє погодинну активність X у порівнянні з відповідними екзогенними джерелами.

3.4. Оцінка продуктивності симулятора процесів в соціальних мережах

Точність згенерованої активності X оцінювалася за темами шляхом порівняння з реальними даними та двома базовими моделями. Продуктивність вимірювалася за трьома основними метриками:

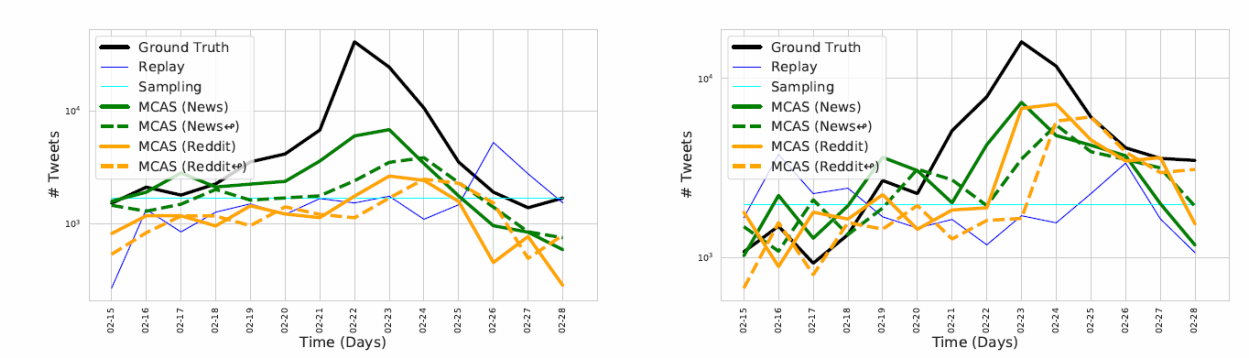
- Щоденний обсяг активності: кількість твітів та ретвітів.
- Кількість нової залученості користувачів щодня.
- Розподіл PageRank мережі взаємодії користувачів.

Через високу складність проблеми прогнозування (яка включає "хто кому відповідає, на яку тему і коли"), пряме порівняння з іншими пов'язаними роботами було неможливим. Натомість, ми використовували дві базові моделі, виведені з навчальних даних:

- Replay: модель просто відтворює всі події, що відбулися за останні два тижні навчального періоду. Змінюються лише часові мітки, нові користувачі відсутні.

- Sampling: модель випадковим чином вибирає повні каскади Twitter, щоб відповідати середньому щоденному обсягу активності за темою, спостережуваному за останні два тижні. Хоча вона ігнорує щоденні варіації обсягу (рис. 3.6), вона наближає загальний обсяг за двотижневий період.

Перевершити такі базові моделі в симуляції активності користувачів є досить складним завданням, як показано в [27].



(а) тема міжнародна допомога (б) тема військова, політика

Рис. 3.6. Кількість твітів за темами

Моделі MCAS (News) та MCAS (Reddit) використовують відповідні екзогенні ознаки з дня прогнозування, тоді як моделі MCAS (News') та MCAS (Reddit') використовують відповідні екзогенні ознаки з попереднього дня перед прогнозуванням. Ми візуалізуємо лише часові ряди для двох найпопулярніших тем, щоб зменшити візуальне перевантаження.

3.4.1. Прогнозування кількісних поділів твітів

Прогнозування щоденної кількості поділів (твітів) за темою є складним через спалахоподібну динаміку та домінування різних тем у різні часові проміжки. Наприклад, піки активності наприкінці лютого (рис. 3.3) були спричинені концертом "Venezuela Aid Live" (домінувала тема "міжнародна допомога") та насильницькими протистояннями 23 лютого (домінували теми "військові", "насильство" та "протести"). Екзогенні дані, особливо з новинних звітів, є цінними для відображення варіацій популярності тем.

Ми оцінювали прогнозування щоденної кількості твітів за темою, використовуючи:

- Нормалізована середньоквадратична помилка (NRMSE) - оцінює часові шаблони (незалежно від масштабу).

- Симетрична середня абсолютна відсоткова помилка (SMAPE) - враховує масштаб помилки.

Ключові висновки щодо прогнозування твітів наступні:

1. Перевага моделі.

Наші варіанти рішення ближче захоплюють тренд кількості твітів до реальних даних, ніж базові моделі, майже для всіх тем. За NRMSE наше рішення перевершує базові моделі для всіх тем (рис. 3.5a). Ми успішно прогнозуємо великі піки активності (рис. 3.6a та 3.6б), що доводить необхідність використання сучасних екзогенних даних для точного прогнозування.

2. Точність обсягу.

Наше рішення демонструє кращі результати не лише в тренді, але й у фактичному обсязі твітів (SMAPE) для більшості тем (рис. 3.5b). Наприклад, для найпопулярнішої теми ("міжнародна допомога") мінімальне значення SMAPE становить 61, тоді як для базової моделі Replay — 99. Найгірший результат отримано для теми "арешти", де піки активності слабо корелювали з екзогенними платформами.

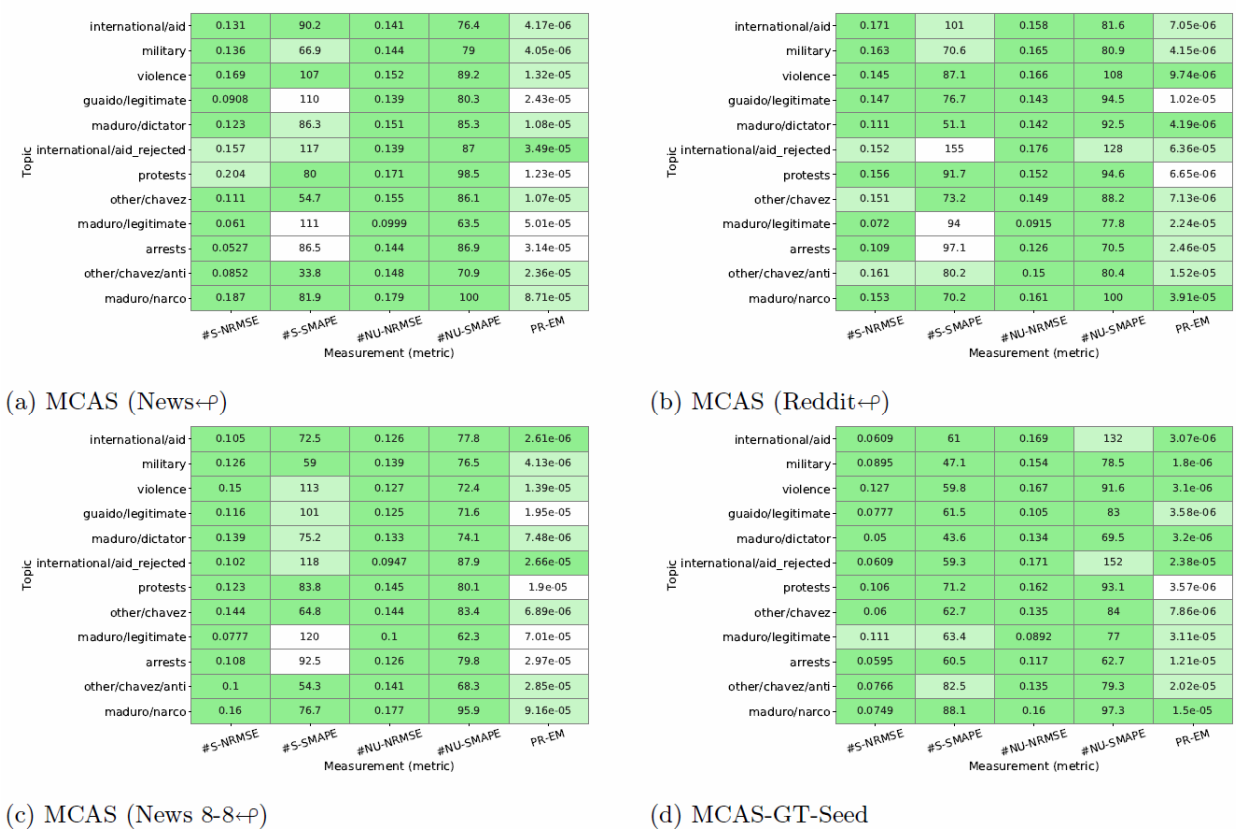
Модель, що використовує ознаки Reddit, прогнозує тренд твітів краще, ніж модель, що використовує лише новини, для більшості тем (рис. 3.5a), за винятком теми "міжнародна допомога". Це свідчить про те, що дискусії в спільноті Reddit пропонують унікальні сигнали щодо політичної кризи.

Ознаки новин (GDELT) можуть бути більш своєчасними для тем, що швидко висвітлюються (наприклад, гуманітарна допомога).

Використання екзогенних даних поточного дня призводить до точніших прогнозів порівняно з даними попереднього дня (рис. 3.5a та 3.5b). Також, зміщення початку "дня" з опівночі на 8 ранку для новинних статей призвело до кращого прогнозування тренду, оскільки це дозволяє захопити пік новинних записів GDELT (близько 5-6 ранку) та зменшити інтервал між подією та початком активності X.

Ключові висновки щодо прогнозування загальних поділів (твіти + ретвіти):

1. Модуль каскадів також захоплює тренд загальних поділів ближче до реальних даних, ніж базові моделі (рис. 3.7 a-c).



(a) MCAS (News↩℘)

(b) MCAS (Reddit↩℘)

(c) MCAS (News 8-8↩℘)

(d) MCAS-GT-Seed

Рис. 3.7. Огляд точності прогнозування активності X

Ми повідомляємо про продуктивність за п'ятьма метриками (показані вздовж осі x) після порівняння з реальними даними (ground-truth):

- Щоденна кількість твітів/ретвітів з часом (#S), виміряна за NRMSE та SMAPE.

- Щоденна кількість новозалучених користувачів з часом (#NU), виміряна за NRMSE та SMAPE.

- Розподіл PageRank мережі взаємодії користувачів (PR), виміряний за метрикою відстані Earth Movers (EM).

Кольори комірок представляють порівняння з базовими моделями: найтемніший колір показує кращу продуктивність, ніж обидві базові моделі; незабарвлений колір показує гіршу продуктивність, ніж обидві базові моделі.

- Модель MCAS (News 8−8′) використовує відповідні екзогенні ознаки за останні 24 години до 8 ранку кожного дня для прогнозування твітів протягом наступних 24 годин.

- Модель MCAS-GT-Seed використовує реальні твіти (ground truth) для генерації ретвітів.

2. Помилки прогнозування в модулі початкових подій (твіти) переносяться на прогнози в модулі каскадів (ретвіти). Найгірші теми в прогнозуванні твітів ("міжнародна/відхилена_допомога", "арешти") також є найгіршими в прогнозуванні загального обсягу (рис. 3.7a, 3.7b).

3. При використанні реальних твітів як вхідних даних для модуля каскадів (нереалістичний сценарій), точність прогнозування ретвітів очікувано зростає (рис. 3.7d), підтверджуючи, що помилки модулів є взаємозалежними.

3.4.2. Прогнозування залучення користувачів

Хоча наше рішення не прогнозує, який саме користувач опублікує повідомлення, ми оцінюємо точність призначення користувачів за часткою нової залученості користувачів та структурою мережі взаємодії.

Прогнозування нових користувачів:

- Наші моделі перевершують базові моделі за NRMSE та SMAPE у прогнозуванні кількості нових користувачів для всіх 12 тем (рис. 3.7a, 3.7b).

- Диференційована корисність джерел, тобто моделі, що використовують ознаки Reddit, краще прогнозують нових користувачів у темах, що виражають невдоволення урядом ("інші/чавес", "мадуро/нарко"). Це свідчить про те, що різні екзогенні джерела пропонують унікальні сигнали для певних груп тем.

Для прогнозування структури мережі (PageRank) була застосована наступна методологія. Для кожної теми була створена спрямована мережа ретвітів (ребро від ретвітера до автора). Розподіли PageRank порівнювалися з реальними даними за допомогою метрики відстані Earth Movers (EM).

Розподіл PageRank, прогнозований нашим рішенням, ближчий до реальних даних, ніж базова модель Sampling для більшості тем (рис. 3.7a, 3.7b). Найнижчі значення відстані EM досягнуто в трьох найпопулярніших темах.

Базова модель Replay була складною для перевершення в цьому вимірюванні. Наше рішення залежить від правильного призначення початкових користувачів (авторів первинних твітів). Оскільки ми випадково обираємо раніше бачених (довгострокових) користувачів як початкових, ми прогнозуємо впливових користувачів, що є реалістичним. Використання реальних початкових користувачів у Модулі Каскадів значно покращує точність розподілу PageRank (рис. 3.7d), підтверджуючи, що подальші вдосконалення у призначенні початкових авторів можуть покращити результати структури мережі.

3.5. Представлення узагальненої архітектури фреймворку та потоку даних представлення активності в соціальній мережі

Фреймворк реалізує модульну архітектуру, призначену для високогранулярного прогнозування активності в соціальній мережі X (Twitter) в умовах інтенсивних, непередбачуваних подій, використовуючи екзогенні джерела даних.

Фреймворк складається з двох ключових, послідовно з'єднаних модулів, що забезпечують спеціалізацію прогнозів та їхню оптимізацію:

- Модуль початкових подій (Seed Module) - Відповідає за кількісне прогнозування первинних подій.

- Модуль каскадів (Cascade Module) - відповідає за генерацію вторинної активності (ретвітів) та моделювання структури поширення.

Вхідні дані (Training/Testing):

1. Екзогенні дані - новинні статті, повідомлення Reddit;

2. Історичні дані - активність X (за темою).

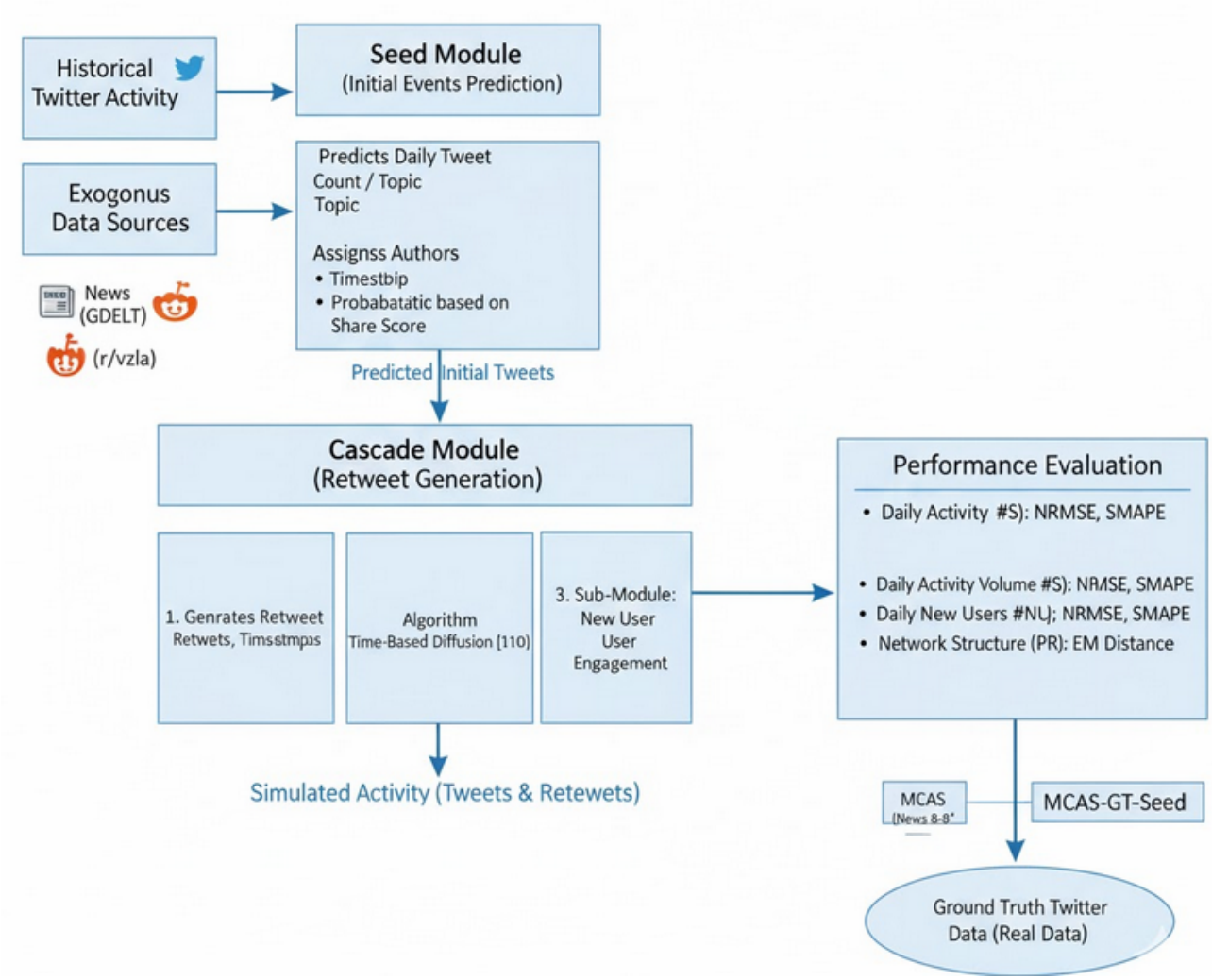


Рис. 3.8. Узагальнена архітектура фреймворку представлення активності в соціальній мережі

Розглянемо потік даних та функціональність модулів

А. Модуль початкових подій (Seed Module)

Вхід: Історична активність X + екзогенні ознаки (кількість подій за темою).

Методологія: нейронна мережа, навчена на екзогенних даних дня d для прогнозування активності дня d+1.

Основна функція: прогнозує щоденну кількість твітів (початкових подій) для кожної теми.

Додаткова функція: призначає автора кожному прогнозованому твіту, використовуючи евристику, пропорційну балу поширення користувача.

Вихід: прогнозований список початкових твітів (включно з темою, часом та автором).

Б. Модуль каскадів (Cascade Module)

Вхід: вихідні дані модуля початкових подій (прогнозовані твіти).

Ключова функція: генерація дерев ретвітів (інформаційних каскадів) у відповідь на початкові твіти. Використовує імовірнісний підхід та алгоритм дифузії, заснований на часі, для реконструкції структури каскаду в мережі підписників.

Підмодуль нових користувачів прогнозує щоденну частку нової залученості користувачів (#NU) на основі тих самих екзогенних ознак. Нові користувачі призначаються як листові вузли в прогнозованих каскадах (кількостях твітів).

Вихід: деталізована симульована активність (ретвіти, часові мітки, автор) для двотижневого періоду.

Симулятор генерує активність із високою гранулярністю (час, дія, тема, користувач) і оцінюється за наступними метриками, що представлені в таблиці 3.1.

Таблиця 3.1.

Метрики, що використовуються у фреймворку

Метрика	**Вимірювання**
Обсяг активності (#S)	NRMSE, SMAPE (щоденна кількість твітів та ретвітів)
Залучення користувачів (#NU)	NRMSE, SMAPE (щоденна кількість новозалучених користувачів)
Структура мережі (PR)	Відстань Earth Movers (EM) між розподілами PageRank

Цей модульний підхід дозволяє оптимізувати окремі аспекти симуляції (наприклад, точність прогнозування обсягу твітів у Seed Module) та уникнути недоліків наскрізних рішень, які компрометують точність у виняткових випадках (наприклад, піки активності).

3.6. Висновки щодо використання методології симуляції процесів в соціальних мережах

Цей розділ представляє дизайн та оцінку модульного симулятора, спроможного генерувати реалістичну активність X в умовах інтенсивних подій реального світу, що спричиняють піки активності та зміну тематичної структури.

Симулятор був навчений на даних Twitter та екзогенних джерелах (Reddit, новинні статті GDELT) і згенерував двотижневу активність Twitter (включно з деталями про автора, час і тему твітів/ретвітів) за умови доступу лише до сучасних екзогенних даних. Емпірична оцінка, проведена на даних, зібраних під час кризи у Венесуелі, підтверджує, що згенерована симулятором активність відповідає часовим рядам реальних даних за обсягом повідомлень та показниками залучення користувачів.

Включення екзогенних даних є необхідною умовою для точного моделювання активності деяких соціальних платформ, зокрема X. Репрезентативність екзогенних джерел залежить від конкретних тем:

- Reddit виявився точнішим предиктором активності X, пов'язаної з політично чутливими або емоційно зарядженими темами, які рідше висвітлюються у традиційних новинах.

Потенційне врахування семантики може підвищити прогностичну спроможність моделі для складних тем.

Хоча наше рішення точно вловило час настання піків активності для багатьох тем, прогнозування коректного обсягу залишалося складним завданням. Здатність точно прогнозувати пікові обсяги активності має важливі прикладні застосування, наприклад, у виявленні групової активності типу "Pump and Dump". Через надзвичайно короткий час реакції користувачів на події реального світу дослідникам необхідно переоцінити визначення "минулого" у контексті соціальних мереж. Обмеження екзогенних даних лише попереднім днем є необґрунтованим. "Минуле" у X може відбутися

лише кілька хвилин тому. Ми емпірично довели, що зсув часового вікна екзогенних даних (наприклад, використання даних GDELT за останні 24 години до 8 ранку) значно покращує точність прогнозування порівняно з використанням даних до опівночі.

Модульна архітектура, розроблена після численних експериментів (включно з підходами типу Long Short-Term Memory для відстеження трендів), виявилася переважною. Наскрізні (end-to-end) рішення схильні до компромісу між різними метриками продуктивності та ігнорують виняткові випадки (піки активності/залучення нових користувачів). Модульність дозволяє оптимізувати ключові виміри симульованих даних (кількість твітів) та забезпечує механізми для корекції малоймовірних результатів.

Інтеграція семантичної інформації є перспективним напрямком для майбутніх досліджень, оскільки це може значно підвищити точність симулятора.

Висновки до розділу

Цей розділ присвячений розробці та оцінці модульного симулятора для імітації активності в соціальних мережах, що зазнає впливу екзогенних факторів. Представлена модульна архітектура поділена на модуль початкових подій (прогнозує первинні твіти) та модуль каскадів (генерує ретвіти та структуру поширення). як екзогенні дані використовувалися новинні статті та дані Reddit, що дозволило моделі точно прогнозувати піки активності під час інтенсивної політичної кризи. Оцінка продуктивності підтвердила, що симулятор перевершує базові моделі за метриками обсягу поділів та залучення нових користувачів, а також за точністю розподілу PageRank мережі Узагальнена архітектура фреймворку, що відображає цей потік даних, підтверджує його потенціал як реалістичного інструменту для симуляції динамічних соціальних процесів.

ВИСНОВКИ

У магістерській роботі вирішено актуальне науково-прикладне завдання, що полягає у підвищенні ефективності моделювання процесів у соціальних мережах шляхом застосування комплексу "data-driven" методів. Дослідження охоплює два критичні аспекти функціонування соціальних графів: забезпечення приватності даних та точне прогнозування користувацької активності під впливом зовнішніх чинників.

На основі проведеного аналізу літературних джерел та існуючих підходів встановлено, що ключовим викликом у сучасній аналітиці соціальних мереж є пошук балансу між корисністю графових даних та збереженням конфіденційності користувачів. Визначено, що існуючі моделі атак деанонімізації постійно еволюціонують, що вимагає розробки нових, адаптивних метрик оцінки вразливості, які б враховували не лише атрибутивні, але й структурні властивості графів.

Спроектовано та реалізовано фреймворк для оцінки стійкості соціальних графів до атак деанонімізації із застосуванням методів машинного навчання. Запропонована архітектура дозволяє моделювати атаки як задачу класифікації, що дає змогу кількісно оцінити ймовірність успішної ідентифікації вузлів мережі зловмисником.

Встановлення взаємозв’язку між топологією та вразливістю. Шляхом пояснювального та прогностичного моделювання виявлено статистично значущі кореляції між топологічними метриками графа (такими як ступінь вузла, коефіцієнт кластеризації, центральність) та його вразливістю. Емпірична оцінка на реальних та синтетичних наборах даних довела, що певні структурні патерни роблять соціальні мережі більш схильними до витоку приватної інформації, що дозволяє прогнозувати ризики ще на етапі проектування архітектури даних.

Розроблено модульну архітектуру симулятора процесів у соціальних мережах, яка, на відміну від класичних моделей поширення інформації,

враховує вплив зовнішніх (екзогенних) подій. Реалізовано модулі прогнозування початкових подій та генерації каскадів ретвітів, що дозволяє відтворювати реалістичну динаміку інформаційних потоків.

Оцінка продуктивності розробленого симулятора продемонструвала високу точність у задачах прогнозування кількісних показників активності (обсягу твітів) та рівня залучення користувачів. Використання тематичної класифікації активності дозволило підвищити гранулярність моделювання, забезпечуючи адекватне відображення реакції соціуму на різнотипні інформаційні приводи.

Результати дослідження підтверджують, що інтеграція методів оцінки вразливості графів та симуляції активності з урахуванням екзогенних факторів дозволяє створити комплексний інструментарій для аналізу соціальних мереж. Запропоновані підходи можуть бути використані для розробки більш безпечних платформ соціальної взаємодії та побудови прогностичних моделей поширення інформації в умовах гібридних інформаційних впливів.

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Improved Achievability and Converse Bounds for Erd ″os-Rényi Graph Matching / Daniel Cullina // https://arxiv.org/pdf/1602.01042
2. SecGraph: A Uniform and Open-source Evaluation System for Graph Data Anonymization and De-anonymization / Shouling Ji / https://scispace.com/pdf/secgraph-a-uniform-and-open-source-evaluation-system-for-3wtdis00b3.pdf
3. On the Relative De-anonymizability of Graph Data: Quantification and Evaluation. – Shouling Ji. – https://www.princeton.edu/~pmittal/publications/relative-deanonymizability-infocom16
4. Quantification of De-anonymization Risks in Social Networks / Wei-Han Lee // https://www.scitepress.org/papers/2017/61925/61925.pdf
5. De-anonymizing Social Networks / Arvind Narayanan and Vitaly Shmatikov // https://arxiv.org/pdf/0903.3276
6. Graph Data Anonymization, De-anonymization Attacks, and De-anonymizability Quantification: A Survey / Shouling Ji // https://nesa.zju.edu.cn/download/Graph%20Data%20Anonymization,%20De-anonymization%20Attacks,%20and%%20A%20Survey.pdf
7. Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography / Lars Backstrom / https://www.cs.cornell.edu/~lars/www07-anon.pdf
8. A Bayesian method for matching two similar graphs without seeds / Pedram Pedarsan // https://www.cos.ufrj.br/~daniel/ie/slides/GraphMatching-IE.pdf
9. On the Privacy of Anonymized Networks / Pedram Pedarsan // https://scispace.com/pdf/on-the-privacy-of-anonymized-networks-1fwx07dom8.pdf
10. When Can Two Unlabeled Networks Be Aligned Under Partial Overlap? // Ehsan Kazemi / https://infoscience.epfl.ch/server/api/core/bitstreams/d6bdbe50-f82b-456a-b54c-8f7791902fb1/content

11. Reza Shokri / Privacy Games: Optimal User-Centric Data Obfuscation . – https://petsymposium.org/popets/2015/popets-2015-0024.pdf
12. Hamilton, W. L., Ying, R., and Leskovec, J. "Representation Learning on Graphs: Methods and Applications," IEEE Data Engineering Bulletin, vol. 40, no. 3, pp. 52–74, 2017.
13. Kipf, T. N. and Welling, M. "Semi-Supervised Classification with Graph Convolutional Networks," in Proceedings of the 5th International Conference on Learning Representations (ICLR). Toulon, France: OpenReview.net, 2017.
14. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Yu, P. S. "A Comprehensive Survey on Graph Neural Networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 4–24, 2021.
15. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. "Graph Attention Networks," in Proceedings of the 6th International Conference on Learning Representations (ICLR). Vancouver, Canada: OpenReview.net, 2018.
16. Easley, D. and Kleinberg, J. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge, UK: Cambridge University Press, 2010.
17. Narayanan, A. and Shmatikov, V. "De-anonymizing Social Networks," in Proceedings of the 30th IEEE Symposium on Security and Privacy (S&P). Berkeley, CA, USA: IEEE Computer Society, 2009, pp. 173–187.
18. Backstrom, L., Dwork, C., and Kleinberg, J. "Wherefore Art Thou R3579x? Anonymized Social Networks, Hidden Patterns, and Structural Steganography," in Proceedings of the 16th International Conference on World Wide Web (WWW). Banff, Alberta, Canada: ACM, 2007, pp. 181–190.
19. Ji, S., Li, W., Srivatsa, M., and Beyah, R. "Structural Data De-anonymization: Quantification, Practice, and Implications," in Proceedings

of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS). Scottsdale, AZ, USA: ACM, 2014, pp. 1040–1053.

20. Nilizadeh, S., Kapadia, A., and Ahn, Y.-Y. "Community-Enhanced De-anonymization of Online Social Networks," in Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS). Scottsdale, AZ, USA: ACM, 2014, pp. 537–548.
21. Pedarsani, P., Grossglauser, M., and Chhatwal, N. "On the Privacy of Social Graphs under Uncertainty," in Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT). Vail, CO, USA: IEEE, 2018, pp. 206–210.
22. Qian, J., Li, Y., Zhang, C., and Chen, H. "De-anonymizing Social Networks and Inferring Private Attributes Using Knowledge Graphs," in Proceedings of the IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications. San Francisco, CA, USA: IEEE, 2016, pp. 1–9.
23. Gong, N. Z. and Liu, B. "You Are Who You Connect With: Attribute Inference in Online Social Networks," in Proceedings of the 25th International Conference on World Wide Web (WWW). Montreal, Canada: ACM, 2016, pp. 985–995.
24. Zhou, B. and Pei, J. "Preserving Privacy in Social Networks Against Neighborhood Attacks," in Proceedings of the 24th International Conference on Data Engineering (ICDE). Cancún, Mexico: IEEE, 2008, pp. 506–515.
25. Liu, K. and Terzi, E. "Towards Identity Anonymization on Graphs," in Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. Vancouver, Canada: ACM, 2008, pp. 93–106.
26. Bhagat, S., Cormode, G., Krishnamurthy, B., and Srivastava, D. "Class-based Graph Anonymization for Social Network Data," in Proceedings of the 35th International Conference on Very Large Data Bases (VLDB). Lyon, France: VLDB Endowment, 2009, pp. 766–777.

27. Dwork, C. and Roth, A. "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.
28. Myers, S. A., Zhu, C., and Leskovec, J. "Information Diffusion and External Influence in Networks," in Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). Beijing, China: ACM, 2012, pp. 33–41.
29. Kempe, D., Kleinberg, J., and Tardos, É. "Maximizing the Spread of Influence through a Social Network," in Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, DC, USA: ACM, 2003, pp. 137–146.
30. Gomez-Rodriguez, M., Leskovec, J., and Krause, A. "Inferring Networks of Diffusion and Influence," in Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, DC, USA: ACM, 2010, pp. 1019–1028.
31. Farajtabar, M., Gomez-Rodriguez, M., Valera, I., Wang, N., and Song, L. "Co-evolutionary Dynamics of Information Diffusion and Network Structure," in Proceedings of the 24th International Conference on World Wide Web (WWW). Florence, Italy: ACM, 2015, pp. 619–630.
32. Rizoiu, M.-A., Xie, L., Sanner, S., Cebrian, M., Yu, H., and Van Hentenryck, P. "Expecting to be HIP: Hawkes Intensity Processes for Social Media Popularity," in Proceedings of the 26th International Conference on World Wide Web (WWW). Perth, Australia: ACM, 2017, pp. 1261–1270.
33. De, A., Valera, I., Ganguly, N., Bhattacharya, S., and Gomez-Rodriguez, M. "Learning and Forecasting Opinion Dynamics in Social Networks," in Advances in Neural Information Processing Systems (NIPS). Barcelona, Spain: Curran Associates, Inc., 2016, pp. 397–405.
34. Zhao, Q., Erdogdu, M. A., He, H. Y., Rajaraman, A., and Leskovec, J. "SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity," in Proceedings of the 21st ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining (KDD). Sydney, Australia: ACM, 2015, pp. 1513–1522.

35. Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., and Leskovec, J. "Can Cascades be Predicted?" in Proceedings of the 23rd International Conference on World Wide Web (WWW). Seoul, Korea: ACM, 2014, pp. 925–936.
36. Goel, S., Anderson, A., Hofman, J., and Watts, D. J. "The Structural Virality of Online Diffusion," Management Science, vol. 62, no. 1, pp. 180–196, 2016.
37. Guille, A., Hacid, H., Favre, C., and Zighed, D. A. "Information Diffusion in Online Social Networks: A Survey," ACM SIGMOD Record, vol. 42, no. 2, pp. 17–28, 2013.
38. Matsubara, Y., Sakurai, Y., Prakash, B. A., Li, L., and Faloutsos, C. "Rise and Fall Patterns of Information Diffusion: Model and Implications," in Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). Beijing, China: ACM, 2012, pp. 6–14.
39. Lerman, K. and Ghosh, R. "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks," in Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM). Washington, DC, USA: AAAI Press, 2010, pp. 90–97.
40. Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. "The Role of Social Networks in Information Diffusion," in Proceedings of the 21st International Conference on World Wide Web (WWW). Lyon, France: ACM, 2012, pp. 519–528.
41. Kwak, H., Lee, C., Park, H., and Moon, S. "What is Twitter, a Social Network or a News Media?" in Proceedings of the 19th International Conference on World Wide Web (WWW). Raleigh, NC, USA: ACM, 2010, pp. 591–600.
42. Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. P. "Measuring User Influence in Twitter: The Million Follower Fallacy," in Proceedings of

the 4th International AAAI Conference on Weblogs and Social Media (ICWSM). Washington, DC, USA: AAAI Press, 2010, pp. 10–17.

43. Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.
44. Vosoughi, S., Roy, D., and Aral, S. "The Spread of True and False News Online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.
45. Yang, J. and Leskovec, J. "Modeling Information Diffusion in Implicit Networks," in Proceedings of the 10th IEEE International Conference on Data Mining (ICDM). Sydney, Australia: IEEE, 2010, pp. 599–608.

ДОДАТОК

Додаток А

Фрагменти програмних кодів

In [ ]:

```
import pandas as pd
import pickle
import json
import numpy as np
import glob
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings("ignore")

import dtw
import os
```

In [ ]:

```
from xgboost import XGBRegressor
import itertools
from sklearn.model_selection import GridSearchCV
import pickle
import cloudpickle
import copy
import scipy.stats as stats
from datetime import datetime, timedelta, date
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
```

In [ ]:

```
'''
Metrics
'''

def myerrsq(x,y):
    return((x-y)**2)

### s2 predictions, s1 ground truth
def dtw_(s1, s2):
    window=2

    s1= pd.DataFrame(s1)
    s2 = pd.DataFrame(s2)

    z1=(s1-s1.mean())/(s1.std(ddof=0).apply(lambda m: (m if m > 0.0 else 1.0)))
    z2=(s2-s2.mean())/(s2.std(ddof=0).apply(lambda m: (m if m > 0.0 else 1.0)))

    ### first value simulation second value GT
    dtw_metric = np.sqrt(dtw.dtw(z2[0], z1[0], dist_method=myerrsq, window_type='slantedband',
                                 window_args={'window_size':window}).normalizedDistance)

    return dtw_metric

def ae(v1,v2):
    v1=np.array(v1)
    v2 = np.array(v2)
    return np.abs(v1 - v2)

# Scale-Free Absolute Error
def sfae(v1,v2):

    v1=np.array(v1)
    v2 = np.array(v2)

    return ae(v1, v2) / np.mean(v1)

def MAD_mean_ratio(v1, v2):
    """
    MAD/mean ratio
    """
    return np.mean(sfae(v1, v2))
```

In [ ]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import glob
import pickle
import os
```

In [ ]:

```
def create_dir(x_dir):
    if not os.path.exists(x_dir):
        os.makedirs(x_dir)
        print("Created new dir. %s"%x_dir)
    else:
        print("Dir. already exists")
```

In [ ]:

```
domain = "domain_name"
domain_out = domain

platform = "platform_name"

models = ["MCAS-BLENDED-INTERNAL", "MCAS-BLENDED-EXOG", "MCAS-META-ENSEMBLE", "MCAS-ENSEMBLE"]
```

In [ ]:

```
for model_id in models:
    file_name = model.lower()
    domain_ = domain_out.split("_")[0]
    domain_id = "_".join(domain_out.split("_")[1:])

    # domain_name = domain_+"_"+domain_id
    domain_name = domain
    orig_path = "../output/{3}/{0}/{1}/{2}/".format(domain_, domain_id, model_id, platform)

    save_path = "../newuser_module/Simulation_cascade_output/{0}/{2}/{1}/".format(domain_name, model_id, platform)
    print(orig_path)
    print(save_path)

    create_dir(save_path)
    sim_datas=[]
    for i in range(1,2):
        sim_data=[]
        for x in glob.glob(orig_path+"cascade_vV%d-*pkl.gz"%i):
            print(x)
            sim_data.append(pd.read_pickle(x))
        sim_data=pd.concat(sim_data)
        sim_data['nodeTime']=pd.to_datetime(sim_data['nodeTime'],unit='s')
        sim_data.sort_values(by='nodeTime',inplace=True)
        sim_data['version']=i
        sim_data = sim_data.reset_index(drop=True)
        file_id = file_name+"_v"+str(i)+".pkl.gz"
        print(i,sim_data.shape[0])
        sim_datas.append(sim_data.reset_index(drop=True))

        sim_data.to_pickle(save_path+file_id)
        print("Saved!",save_path+file_id)
```