державного кордону для встановлення осіб, які здійснювали спільні перетини, виявлення періодів перебування за межами України; виокремлення транспортних засобів, які найчастіше використовувалися особами при перетинах.

5. Аналіз податкових накладних.

Power Query допомагає автоматизувати обробку та агрегацію податкових накладних, здійснити очищення даних, встановити підприємства, які подають податкову звітність з однакових IP-адрес та взаєморозрахунки між ними і т.д.

Олейніков Олег Анатолійович,
начальник відділу програмно-технічного забезпечення слідчої та оперативно-розшукової діяльності Управління інформаційних технологій Державного бюро розслідувань

МЕТОДИ GRAPH INTELLIGENCE ТА АНАЛІЗ СХЕМ ЗВ'ЯЗКІВ ПІД ЧАС РОЗСЛІДУВАННЯ КРИМІНАЛЬНИХ ПРАВОПОРУШЕНЬ

У процесі розслідування кримінальних правопорушень збільшується потреба аналізу складних схем зв'язків: між абонентами мобільного зв'язку, учасниками фінансових операцій, суб'єктами у соціальних мережах, особами й об'єктами реального світу.

Зв'язки, які раніше аналізувалися інтуїтивно, зараз мають великий обсяг та складну структуру – надвелика кількість учасників (об'єктів аналізу), історичні дані за значний календарний період, поєднання схем з ваговими або часовими атрибутами. У таких випадках традиційні табличні методи та графічне представлення стають недостатньо ефективними.

Термін Graph Intelligence (GraphINT) почав з'являтись серед офіційних публікацій розробників сучасних програмних продуктів орієнтованих на роботу зі складними зв'язками (Tom Sawyer Software, Graphistry, Locstat). Загалом зі збільшенням складності та розміру досліджуваних даних виникла потреба у застосуванні спеціального програмного забезпечення, пов'язаного з сучасними аналітичними методами та спеціалізованими нейронними мережами, орієнтованих на роботу з графовою структурою. Не менш важливим є

забезпечення ефективної інфраструктури зберігання та обробки графової інформації, засобів її відображення, підтримки редагування аналітичних сценаріїв.

Дослідження схем зв'язків під час розслідування переважно збігається з підходами, які застосовуються при дослідженні зв'язків у комерційній сфері, банківській діяльності, протидії страховому шахрайству, тощо.

Серед інструментів, які підтримують аналіз графових структур, можна виділити кілька категорій:

– візуальні засоби з можливістю базового статистичного аналізу або інтеграцій з іншими сервісами зберігання даних: Gephi, Maltego, Cytoscape, IBM i2 Analyst Notebook – для побудови та візуалізації схем зв'язків;

– публічні сервіси візуалізації з використанням обчислювальних можливостей надавача послуг: Graphistry, Kineviz, Linkurious;

– прикладні програмні бібліотеки для розробки додатків: networkx, igraph, graph-tool, pyvis, DGL, PyTorch Geometric – дають змогу створювати власні аналітичні інструменти;

– системи управління баз даних: Neo4j (Neo4j Bloom в якості інструмента взаємодії), TigerGraph, ArangoDB – оптимізовані для роботи з графами;

– табличні процесори Excel, LibreOffice Calc, pandas – для первинного групування або формування списків вузлів та зв'язків.

У рамках підходів GraphINT застосовується широкий спектр методів, що охоплюють як класичні, так і сучасні техніки аналізу. Насамперед використовуються математичні методи теорії графів, зокрема центральності, кластеризації, підграфи зв'язності та аналіз потоків, які дозволяють виявляти ключові вузли та групи. Доповненням до них виступають статистичні методи – частотний аналіз та виявлення аномальної активності. Для реалізації аналітики використовуються програмні підходи, які передбачають побудову алгоритмів на основі списків зв'язків, матриць суміжності або часових журналів подій. Окрему категорію становлять методи машинного навчання, серед яких слід виокремити графові нейронні мережі (GNN), а також алгоритми node2vec, класифікацію вузлів та embedding-

підходи. Не менш важливою є візуальна складова, яка охоплює інтерактивну навігацію по графу, форматування, динамічну фільтрацію за часом, категоріями зв'язків або сумарними показниками.

У найпростішому випадку зв'язки між об'єктами описуються за допомогою ненаправлених простих графів, які лише фіксують сам факт взаємозв'язку без уточнення його сили чи напряму. Така структура часто використовується для моделювання соціальних відносин. У межах такого підходу можуть бути застосовані базові аналітичні методи: виявлення груп (кластерів), пошук центральних вузлів, аналіз компонент зв'язності. Подібні графи добре підходять для первинного дослідження структури мережі, коли пріоритетом є зрозумілість і наочність.

Більш інформативними є зважені графи, у яких кожне ребро має числове значення – наприклад, кількість з'єднань, тривалості взаємозв'язків або обсяг переданої інформації чи коштів. У таких графах з'являється можливість враховувати не лише наявність зв'язку, але й його інтенсивність.

Найбільш гнучким інструментом для кримінального аналізу є множинні направлені графи з часовими позначками та вагами, які здатні відображати складні сценарії взаємодії – зокрема, банківські транзакції, податкові зобов'язання, історію взаємодії. Кожне ребро в таких графах має напрямок, вагу (наприклад, суму переказу) і часову мітку, причому між одними й тими ж вузлами можуть існувати множинні зв'язки. Це дозволяє досліджувати динаміку взаємодій, аналізувати часові шаблони, виявляти піки активності, будувати агрегати за періодами («ковзне вікно») та виявляти аномалії у часовому контексті.

Зі збільшенням складності графу – зростає як навантаження на обчислювальні ресурси, так і глибина доступної аналітики. Саме складні графи з напрямками, вагами та часом надають найбільшу цінність у кримінальному аналізі, але водночас вимагають гнучких інструментів і формалізованих методик.

У цій статті наводяться деякі початкові складності, які виникають при аналізі графів без спеціального програмного

забезпечення. Їх вирішення та автоматизація може стати початком для застосування більш складніших методів або розробки власних аналітичних модулів.

Проблема побудови графа з «готових» даних. Наявність структурованих масивів, таких як журнали подій чи транзакцій, не означає, що вони готові до відображення у вигляді схеми зв'язків. Перетворення таких даних потребує фільтрації, агрегування й очищення. Очікується, що результат роботи аналітика – це не візуальна копія наявних даних, а саме візуальне узагальнення, досягнуте в результаті опрацювання даних в контексті розслідування. Таким чином, може існувати хибне уявлення про тотожність вилученим даним до аналітичного результату.

Проблема абсолютних метрик. Оцінювання вузлів через звичні метрики (кількість зв'язків, суму транзакцій) не враховує індивідуальний контекст. Фільтрація вузлів за абсолютними вагами може призвести до втрати інформації про менш активні об'єкти схеми. Пропонується вводити оцінки близькості між об'єктами, які будуть враховувати особливості кожного, наприклад – формуванням кластерів близькості серед сусідів або застосування алгоритмів масштабування «скейлерів», нормалізації даних, тощо.

Проблема шумових вузлів. У графах часто з'являються технічні або сервісні об'єкти (сервісні номери, рекламні облікові записи, групи типових одержувачів коштів, тощо). Такі вузли мають високу активність, але не несуть цінності для розслідування. Вони формують штучні зв'язки, які можуть спотворити аналіз схеми або вказувати на хибно позитивні зв'язки між учасниками провадження.

Проблема фіксації цільових вузлів. На відміну від «шумових» вузлів, на схемах зв'язків може бути втрачена важлива інформація щодо об'єктів, які становлять інтерес для розслідування, якщо інтенсивність їх участі замала, хоча сам факт їх участі вже є важливим.

Проблема обмеженої видимості. У більшості випадків граф будується лише навколо об'єктів, які викликають інтерес (наприклад, щодо яких прийнято рішення про виїмку). При цьому решта мережі лишається невідомою, і базовий граф не

відображає реальні зв'язки та множину учасників. Через відсутність повноти графу звичайні метрики завищуються, що призводить до спотвореної інтерпретації аналізу графу та неможливості їх використання (наприклад модуль Social Network Analysis IBM i2 Analyst Notebook).

Проблема редагування та доповнення. У графах, які були очищені від слабких зв'язків, нові дані часто не мають можливості інтеграції. Для подолання цього доцільно зберігати первинні схеми та формувати окремі аналітичні зрізи під час роботи з графом.

Проблема візуального перевантаження. Аналіз та узагальнення схем отриманих з великих наборів інформації створює обов'язковість компромісу між втратою інформації та здатністю графа бути інформативним та зрозумілим. Доцільним є перехід від друкованих примірників до використання у роботі інтерактивних схем, з можливістю перегляду спрощених та деталізованих варіантів інтерпретації, використання тривимірних схем для відображення графів з великою кількістю об'єктів чи кластерів.

Сучасні розслідування дедалі частіше базуються на аналізі графових структур – від простих соціальних зв'язків до складних транзакцій з часовими ознаками. Проте більшість доступних інструментів залишаються занадто універсальними або негнучкими, що обмежує їх ефективність у реальних криміналістичних сценаріях або призводить до повторюваної втрати часу на адаптацію.

У цьому контексті особливо актуальною є розробка спеціалізованих рішень у межах внутрішньої інфраструктури в залежності від потреб підрозділу або типових напрямків розслідування. Власні інструменти дослідження графів дозволяють автоматизувати типові аналітичні дії, зменшити залежність від вартісних ліцензій, а також використовувати сучасні підходи – від машинного навчання до інтеграції з зовнішніми джерелами.