УДК 004.724.4(045)

НЕЧАЄВ Олександр Олексійович

ВИЯВЛЕННЯ ТА АНАЛІЗ ЗЛОЧИННИХ МЕРЕЖ, ПОБУДОВАНИХ НА ОСНОВІ СТАТИСТИЧНИХ ДАНИХ ТЕЛЕФОННИХ КОНТАКТІВ

Постановка проблеми. Нині використання багатьох засобів комунікації таких, як Інтернет та мобільний зв'язок, настільки увійшло в наше життя, що без цього переважній більшості людей важко себе уявити. Окрім того, не менше ніж в побуті засоби комунікації використовуються для координації та вчинення злочинних дій. Відомо, що всі перелічені засоби зв'язку залишають гетерогенні сліди в інформаційному просторі та, у встановленим законом порядку, акумулюються правоохоронними органами. У цій роботі показано теоретичну базу для виявлення і класифікації злочинних організацій в мережах, створених на основі даних про їх телефонні контакти.

Як правило, кримінальні організації мають чітку структуру, яку в більшості випадків можливо сміливо назвати ієрархічною. В організованому злочинному угрупованні завжди є лідер та наближені до нього; прості виконавці; ті, що виконують специфічні функції, найбільш цікаві з яких це зв'язки з іншими групами. Також відомо, що різні ОЗУ контактують між собою.

Кожна особа є окремою індивідуальністю, вона має унікальні відбитки пальців, унікальне обличчя та унікальну поведінку. Проте всі люди мають однакову структуру тіла та природою закладені схожі реакції на зовнішні подразники. Те ж саме можна сказати і про спілкування людини. Кожна особа має своє унікальне розподілення контактів. Проте її робочі контакти, близькі соціальні контакти, родинні зв'язки тощо будуть мати свої, у більшості випадків, загальні для всіх показники. В мережі контактів членів злочинного угруповання всі по-різному будуть спілкуватися з особами, які стоять на вищих щаблях ієрархії, та з тими, що з ними рівні. Також при великому складі угруповання завжди виділяються внутрішні субструктури, які виконують схожу роботу та входять, якщо можна сказати, до одних підрозділів.

Метою роботи є підведення теоретичної бази для виявлення і класифікації злочинних організацій в мережах, створених на основі даних про їх телефонні контакти. Також в ході роботи необхідно сформувати прототип експертної системи виявлення в мережах телефонного зв'язку кримінальні угруповання та їх субструктури.

Таким чином, виявлення злочинних організацій включає два формалізовані етапи:

1. Викриття таких спільнот у великих телекомунікаційних мережах можливо на основі достовірно відомої інформації про зв'язки окремих осіб із ОЗУ або на підставі припущення про їхню причетність до угруповання, заснованому на статистичному аналізі гетерогенних медіа-даних.

2. Другий етап включає вивчення взаємин, що існують між членами злочинних угруповань, динаміки їх зв'язку, ієрархічних відносин; з'ясування повної структури організації та ролі кожного її функціонера.

У сучасному світі ці ідеї широко використовуються на комерційній основі, багато розвинутих країн є їхніми споживачами, але, природно, що більшість із реальних досліджень у цій галузі підпадають під гриф державної або комерційної таємниці.

Серед найбільш відомих засобів аналізу візуалізації кримінальних мереж можна назвати наступні: COPLINK, Analyst's Notebook, XanalysisLink Explorer та відомий PalantirGovernment [9].

Виклад основного матеріалу. Початкові дослідження в області аналізу соціальних мереж можна знайти у працях Мільграма і Траверса [1], де аналізуються характеристики реальних складних мереж, здійснюються соціальні експерименти в реальному світі та зображений так званий «ефект малого світу». Згідно цієї моделі доказується, що незважаючи на великий обсяг, складні мережі мають низку загальних властивостей: ступеневий розподіл сформованого соціального графу підкоряється степеневому закону, хоча і асимптотично, тобто відношення вузлів графу, що мають k зв'язків між собою, до кількості усіх вершин для великих значень k визначається як: $P(k) \sim k^{-\gamma}$, де γ – це константа, значення якої знаходиться, зазвичай, у межах $2 < \gamma < 3$, однак інколи значення γ може бути поза цими межами. Також в таких мережах існує відносно короткий шлях, що з'єднує довільну пару вузлів.

А. Барабаши [2] довів модель зростання мережі, яка була декілька разів успішно використана на WorldWideWeb, телекомунікаційних мереж тощо. Автор показав, що реальні складні мережі мають одну і ту ж динаміку зростання, так звану пільгову прихильність: нові вузли, як правило, більше підключаються до вузлів із великим ступенем, ніж до вузлів з нижчим. Ці твердження спричинили виникнення поняття «безмасштабності» мережі (тобто специфічного розподілу вузлів та їх відповідного розвитку), дозволяючи наявність вузлів-концентраторів (вузлів-зірок) та посередників (тих, через які проходить максимальна кількість мінімальних шляхів між довільними парами вузлів). Таким чином, «ефект малого світу», привілейована модель зростання, та степеневий розподіл ступенів характеризують структуру соціальної мережі [5].

Основні роботи із застосування аналізу складних мереж для дослідження саме злочинних організацій, їх вразливостей, а також праці в сфері розвідки належать Малколму Спарроу [3]. Саме він визначив чотири риси, властиві кримінальним мережам:

1) обмежений розмір. Кримінальні мережі мають переважно не більше 1-2 тисячі вузлів;

2) неповнота інформації. Кримінальні мережі неминуче містять порожні фрагменти інформації та помилкову інформацію;

3) невизначені кордони. Важко визначити всі відносини окремих вузлів;

4) динаміка. Нові стійкі з'єднання здійснюються лише з вузлами схожих мережевих структур.

Завдяки М. Спарроу інші автори теж взялись за дослідження кримінальних мереж. Наприклад, Бейкер і Фаулкнер [5] вивчали злочинні мережі, порівнюючи їх розвиток із законами, притаманними електричному обладнанню. Арквілла і Ронфельд [4] узагальнили попереднє дослідження шляхом введення концепції мережевого протистояння тероризму. Вони явно демонструють різницю між звичайними соціальними мережами та кримінальними.

Однак усі останні роботи нехтують важливістю візуалізації, зосереджуючи більше уваги на статистичних характеристиках мереж, або винайдення параметрів, які б допомогли ідентифікувати роль кожного вузла в мережі.

Лише у 2006 році Валдіс Кребс застосував мережевий аналіз із візуалізацією, який розкрив структуру «Аль-Каїди», викритої шляхом аналізу складної мережі, створеної на основі телефонних контактів осіб, які мали місце в подіях 11 вересня 2001 року в США [8].

Слід зазначити, що саме після цих подій перед світом найбільш гостро постало питання необхідності розвитку теорії складних мереж для забезпечення державної безпеки. Це стало відправною точкою для серії наукових робіт, в яких аналіз соціальних мереж застосовується на реальних випадках, на відміну від попередніх, які використовували для дослідів штучні та змодельовані мережі. Тоді американський уряд звернувся до вчених та дослідників за допомогою у виявленні терористичних груп з метою їх попереднього захоплення та уникнення можливості вчинення терористичних актів. Як було відомо із медіа, подальше усунення цієї організації було здійснено саме завдяки методам теорії складних мереж.

Роботи В. Кребса є одним із найбільших відкритих прикладів застосування теорії складних мереж для викриття кримінальних структур.

Зазначені факти доводять ефективність таких технологій та роль, яку вони відіграють в діяльності сучасних правоохоронних органів. Проте лише незначна частина з реальних досліджень не підпадає під гриф державної або комерційної таємниці.

Аналіз соціальних мереж

Таким чином, у кримінології та боротьбі з тероризмом аналіз соціальних мереж (далі SNA – social network analysis) став потужним інструментом для викриття структур кримінальних організацій. Він дозволяє зрозуміти позиції кожного вузла в соціальній мережі й особливості взаємозв'язків між

вузлами, а також приналежність окремих вузлів до субструктур. SNA визначає ключові поняття для опису структури мережі та ролей вузлів в ній, основні з яких:

1. Центральність. Відноситься до групи метрик, метою яких є визначення «значущості» або «впливу» певного вузла (або групи вузлів) у мережі. Прикладами загальних методів виміру центральності є визначення центральності з посередництва, за близькістю, центральності власного вектора, альфа-центральності та центральності за ступенем [15].

2. Міст. Вузол, слабкі зв'язки якого заповнюють структурні прогалини, забезпечуючи єдине з'єднання між двома індивідами (або кластерами, субграфами). Він так само включає в себе найкоротший шлях, коли більш довгий шлях неможливий через високий ризик спотворення повідомлення або неможливості доставки.

3. Закритість мережі. Відображає міру повноти реляційних тріад. Присвоєння вузлам ступеня закритості (тобто того факту, що їх зв'язки в свою чергу зв'язані між собою) називається транзитивністю.

4. Щільність. Ставлення прямих зв'язків у мережі до загально можливої кількості зв'язків. Щільним графом називається граф, в якому кількість зв'язків близька до максимальної. Граф із протилежною властивістю, що має малу кількість зв'язків, називається розрідженим графом.

5. Посередництво. Відображає кількісне значення, що дорівнює кількості мінімальних шляхів між двома довільними вузлами, які проходять через вузол, що розглядається. Таким чином, може розраховуватись значення посередництва між двома зв'язками.

Також SNA включає методи виявлення кластерів (або субграфів), визначаючи їх роль та ролі окремих вузлів в субструктурах, які забезпечуються різними методами графічного зображення мереж. Одні з них призначені для виявлення груп всередині мережі, а інші розроблені для виявлення соціальних ролей членів групи. Найбільш поширені методи візуалізації мають матричне представлення і так зване «вузол-зв'язок».

Питання якісної візуалізації, яка б демонструвала всі перелічені вище властивості соціальних мереж, набуває все більшого значення. Розвиток візуалізаційних засобів стикається з такими основними аспектами:

1. Вибір алгоритму візуалізації.

2. Представлення динамічної зміни структури графу в часі.

3. Режими запровадження ітеративності для зменшення візуальної складності.

Зараз існує багато засобів візуалізації графів, які мають ті чи інші недоліки та переваги, однак це питання досі є надактуальним. У цій роботі буде розглянуто найбільш актуальні методи для візуалізації саме кримінальних мереж.

Центральний вузол злочинної мережі може грати ключову роль лідера, віддаючи накази, забезпечуючи та контролюючи інформаційний потік до всіх компонентів злочинної мережі. Видалення таких вузлів може ефективно вплинути на структуру мережі, в деяких випадках аж до припинення нею протиправної діяльності.

Крім того, при аналізі кримінальних мереж слід звернути увагу на підгрупи (або банди), кожні з яких найчастіше грають свою специфічну роль. Взаємодія субструктур між собою дозволяє функціонувати всій злочинній організації як окремому організму. Виходячи з цього, при видаленні окремих вузлів, які зв'язують між собою групи, можна також вчиняти великий деструктивний вплив на всю мережу в цілому.

Важливим аспектом аналізу злочинних мереж є забезпечення поповнення аналітичної системи різноплановою інформацією із різних джерел, що надаватимуть якісні структуровані початкові дані для подальшої аналітичної роботи.

Мобільні телефони та різноманітні інтернет-засоби спілкування постійно використовуються для координації та здійснення протиправної діяльності.

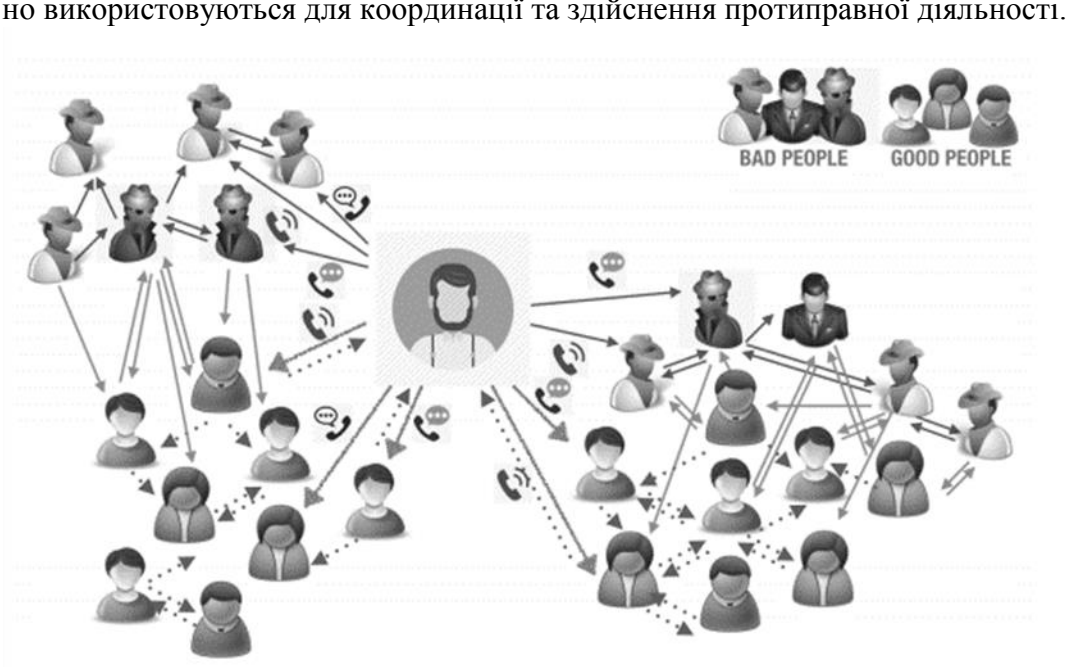

Рис. 1. Телефонні дзвінки підозрюваного

На рис. 1 засобом візуалізації зображено зв'язки підозрюваного. Завдяки схемі контактів можна зробити аргументоване припущення, що підозрюваний грає роль посередника між двома субструктурами злочинної організації [5].

Алгоритми візуалізації

Найбільш типовий алгоритм візуалізації графів FDA (force-directed algorithm) відображає структуру графу як фізичної системи, в якій на позицію вузла (його координати) впливають різні псевдофізичні сили. Система моделює положення вузлів, при якому всі сили перебувають в стані найбільшої рівноваги. Цей алгоритм зручно використовувати, коли необхідно продемонструвати об'єднані групи вузлів або кластерів відповідно до рівня з'єднаності один з одним. Метод BHA (Barnes-Hutalgorithm) симулює гравітаційну модель в сенсі постійного оновлення позицій вузлів.

Для настройки візуалізації залежно від того чи іншого завдання виникає можливість ітеративно змінювати параметри відношень між вузлами. У цьому випадку можна уявити, що вузли з'єднані між собою пружинами однакової напруженості, що координує позиції відображення вузлів. Ітеративна зміна параметрів являє собою зміну напруженості пружин. Відповідно, вузли з невеликим ступенем (кількістю зв'язків) будуть розташовані в периферійній області, а вузли-хаби в центральній.

Нижче наведемо методи візуалізації специфічного налаштування властивостей FDA для відображення саме злочинних мереж.

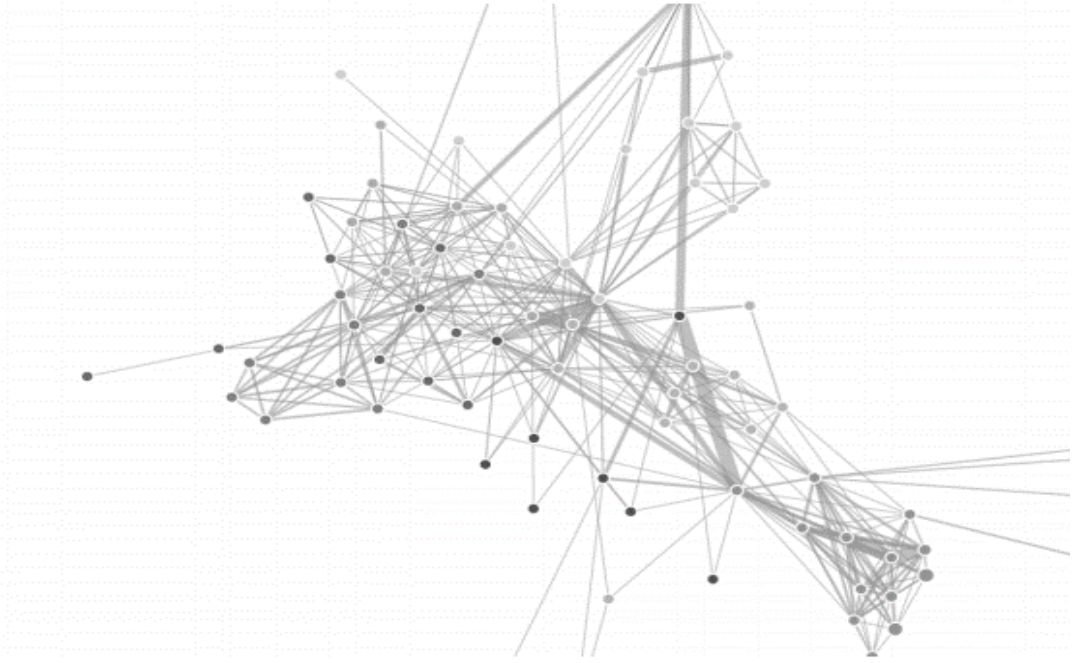

Рис. 2. Приклад force-directed algorithm на невеликій мережі з 75 вузлів з середньою щільністю зв'язків

Фокус та контекст-орієнтована візуалізація

Кількість зв'язків в мережі іноді швидше росте ніж кількість вузлів. Як наслідок, схема мережі може стати нечитаємою через надлишкову щільність зв'язків. Така проблема відома як «візуальне перезавантаження» [10]. Одним

із способів її уникнення є використання функції масштабування тієї частини графу, яка становить найбільший інтерес. Але цей метод має основний недолік – глобальна структура мережі при масштабуванні окремої її частини залишається без уваги. Поміж тим, така методика захисту від надлишковості може бути застосована в злочинних мережах через їх, по-перше, квазіієрархічну структуру, а, по-друге, порівняно з іншими мережами помірного розміру.

Також у правоохоронних органів регулярно виникає необхідність перевірити зв'язки конкретної людини (підозрюваного). Такий вид візуалізації демонструє високі показники при візуалізації невеликих мереж. Основне завдання аналізу мережі у більшості випадків полягає у виявленні нечітких (або латентних) зв'язків особи, яка вивчається, з кримінальними або іншими колами. У такому разі слід зосереджуватися на окремих контактах, а не на всій мережі в цілому і її структурі.

У класичному варіанті FDA не передбачає застосування методики фокусування на окремій частині графу. Але наявні спеціальні модифікації цього алгоритму дозволяють це робити: «риб'яче око» (Fisheye FDA) та «фокус» (Foci FDA) [12–13].

Модифікація Fisheye FDA

Цей алгоритм відображає фокус залежно від контексту.

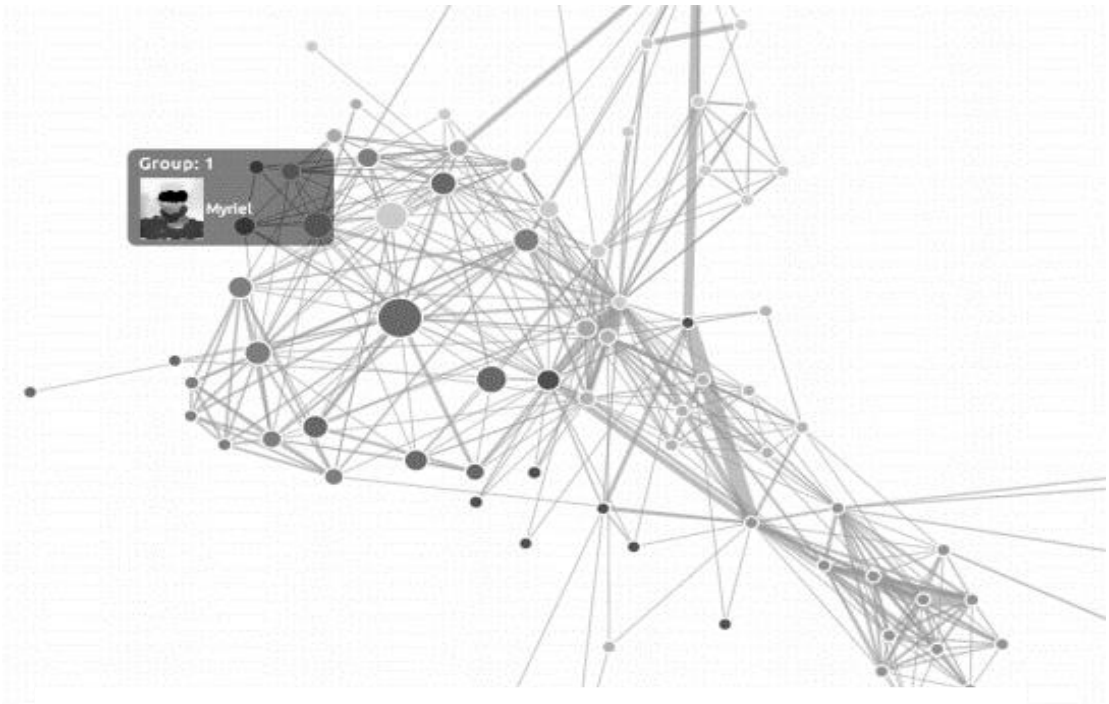

Рис. 3. Приклад Fisheye FDA на тій же мережі

Алгоритм дозволяє користувачу зосередитись на одній або декількох областях соціального графу та динамічно настроювати фокус. Вперше алгоритм був запропонований для візуалізації самоорганізаційних карт Кохонена (Self-organizing map – SOM) В. Фурнасом [10]. Метод відомий тим, що вносить певне якісне спотворення класичної візуалізації. Це метод розширення, який дозволяє не збільшуючи розмір самого простору відображення, значно розширити область навколо фокуса та стиснути райони віддалені від нього, зберігаючи при цьому загальну структуру мережі.

Модифікація Foci FDA

Цей тип візуалізації об'єднує FDA, семантичний та кластерний методи відображення. Він оснований на алгоритмі виявлення субструктур Ловіана, який підтримує багаторівневий аналіз з плаваючим гравітаційним центром.

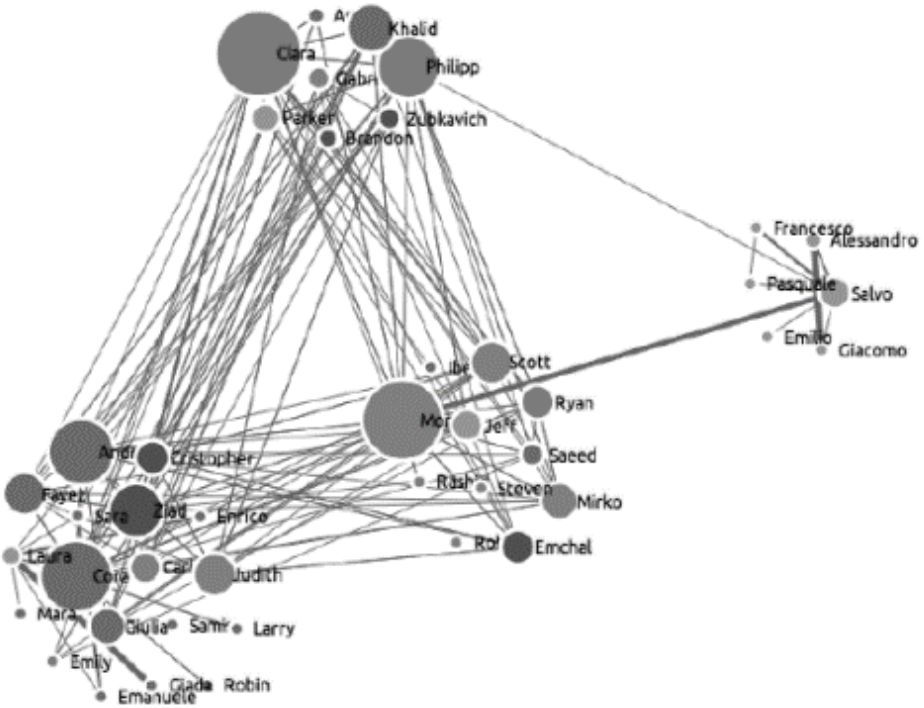

Рис. 4. Приклад Foci FDA на тій же мережі

Запропонований метод візуалізації дозволяє об'єднати щільно зв'язані субграфи в один вузол, що демонструє непогані показники при необхідності зображення повної структури всієї злочинної мережі. Водночас він дає можливість розкрити окрему субструктуру та детально проаналізувати взаємини між членами конкретної групи.

На рис. 4 зображена кластеризація, яка відтворює територіальний поділ між групами, що належать одній організації.

Візуалізація мереж з використанням гео-маппінгу

Також необхідно торкнутись методів візуалізації з прив'язкою до території, оскільки відомо, що всі телефонні контакти здійснюються через базові станції мобільного зв'язку, на яких зберігаються координати. Кожна базова станція має декілька секторів у радіусі своєї дії та запам'ятовує радіус (або направлення), з якого абонент здійснює телефонний або інший контакт, який визначає кут охоплення сектора антени. Ці дані не дозволяють точно локалізувати місцезнаходження абонента на момент зв'язку, однак можна приблизно локалізувати користувачів, які потрапляють в конкретну зону дії базової станції.

Рис. 5. Приклад візуалізації з використанням гео-маппінгу

Використання такого типу відображення кримінальних мереж надає унікальні аналітичні можливості зі спостереження за місцем знаходження членів ОЗУ або підозрюваних у певний момент часу (наприклад, під час вчинення злочину). Необхідно вказати, що ця методика візуалізації може не тільки скоротити час аналітичної обробки великих масивів інформації та представити її в зручному вигляді, але у багатьох випадках зробити такий аналіз можливим.

Висновки. Протягом останніх 10–12 років спостерігається активне залучення вчених та дослідників до вивчення злочинних та терористичних мереж з метою виконання завдань із забезпечення громадської та державної

безпеки розвинутих країн світу. В цьому аспекті аналіз соціальних мереж надає гнучкі та ефективні інструменти для виявлення злочинних мереж на основі їх цифрових слідів в інформаційному просторі, що активно розвивається; виявлення структури злочинних організацій та розбиття їх на підгрупи, які, як правило, зосереджені на виконанні тих чи інших завдань; встановлення взаємозв'язків між окремими функціонерами ОЗУ, а також між окремими бандами та групами; вивчення інформаційних потоків в кримінальному середовищі; виявлення таких функціонерів груп або зв'язків між окремими з них, при знешкодженні яких злочинна мережа може припинити свою протиправну діяльність або зовсім перестати існувати.

З урахуванням викладеного дослідження теорії складних мереж в контексті забезпечення державної та громадської безпеки набуває все більшої актуальності та необхідності запровадження її методів у повсякденну роботу правоохоронних органів.

Список використаних джерел

1. Travers, Jeffrey & Stanley Milgram (1969) «An Experimental Study of the Small World Problem», Sociometry, Vol. 32, No. 4, pp. 425–443.

2. Barabási, Albert-László and Réka Albert, «Emergence of scaling in random networks», Science, 286:509-512, October 15, 1999.

3. M. K. Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. Social Networks, 13(3):251–274, 1991.

4. J. Arquilla and D. Ronfeldt. Networks and netwars: The future of terror, crime, and militancy. Survival, 44(2):175–176, 2001.

5. W. Baker and R. Faulkner. The social organization of conspiracy: illegal networks in the heavy electrical equipment industry. Am. Social. Rev., 58, 1993.

6. E. Ferrara, P. De Meo, S. Catanese, and G. Fiumara. Detecting criminal organizations in mobile phone networks. Expert Systems with Applications, 41(13):5733–5750, 2014.

7. P. Klerks and E. Smeets. The network paradigm applied to criminal organizations: Theoretical nitpicking or a relevant doctrine for investigators.Connections, 24:53–65, 2001.

8. V. Krebs. Mapping networks of terrorist cells. Connections, 24(3):43–52, 2002.

9. C. Morselli. Assessing vulnerable and strategic positions in a criminal network. Journal of Contemporary Criminal Justice, 26(4):382–392, 2010.

10. F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger. Understanding online social network usage from a network perspective. In Proc. 9th SIGCOMM conference on Internet measurement conference, pages 35–48. ACM, 2009.

11. H. Zang, F. Baccelli, and J. Bolot. Bayesian inference for localization in cellular networks. In 2010 Proceedings IEEE INFOCOM, pages 1–9. IEEE, 2010.

12. J. Xu and H. Chen. Criminal network analysis and visualization. Comm. ACM, 48(6):100–107, 2005.

13. C. Yang, H. Chen, and K. Hong. Visualization of large category map for internet browsing. Decis. Support Syst., 35(1):89–102, Apr. 2003.

14. C. Yang, N. Liu, and M. Sageman. Analyzing the terrorist social networks with visualization tools. InIntelligence& securityinformatics. 2006.

15. L. Freeman. A set of measures of centrality based on betweenness. Sociometry, pages 35–41, 1977.

Рецензенти:
кандидат технічних наук Н. Григоренко,
кандидат технічних наук Ю. Іващенко

Аннотация Рассмотрены методы выявления преступных и террористических организаций, изучения их структуры и слабых мест на основе анализа сложных сетей, смоделированных на базе данных их телефонных или других контактов, которые оставляют гетерогенные следы в информационном пространстве. Проведено исследование эффективности использования ряда новейших алгоритмов визуализации для анализа криминальных сетей. Представлена теоретическая база для выявления и классификации преступных организаций в сетях, созданных на основе данных об их телефонных контактах.

Ключевые слова: идентификация преступных организаций, анализ преступных сетей, сети телефонных контактов.

Abstract: The article considers criminal and terrorist organizations detecting, their structure and weak points on the basis of the analysis of complex networks, built on the data related to their phone and other contacts which leave heterogeneous traces in information space. This paper examines efficiency of using newest algorithms of visualization for criminal networks analysis. Theoretical basis for detecting and classification of criminal organizations in the networks, based on their phone contacts is introduced.

Key words: terrorist organizations detecting, analysis of criminal networks, mobile phone networks.

УДК 004.912

САВЧЕНКО Денис Сергійович

ТЕОРЕТИЧНІ ОСНОВИ ФОРМУВАННЯ МОДЕЛЕЙ СИСТЕМ АВТОМАТИЗОВАНОЇ ОБРОБКИ НЕСТРУКТУРОВАНИХ ТЕКСТІВ

Постановка проблеми. Неструктуровані тексти природною мовою є найбільш поширеною формою представлення знань: вони легко породжуються, сприймаються, змінюються і розповсюджуються людиною. Однак інтенсивне зростання корпусу неструктурованих текстів, доступних на сьогодні в електронному вигляді завдяки глобальним інформаційним мережевим