INFORMATION TECHNOLOGY

DOI 10.20535/2411-1031.2024.12.2.315731
УДК 004.9

ОЛЕКСАНДР ПУЧКОВ,
ДМИТРО ЛАНДЕ,
ІГОР СУБАЧ

МОДЕЛЬ ПРОСТОРУ ТЕМАТИЧНИХ ТЕЛЕГРАМ-КАНАЛІВ, ЩО БАЗУЄТЬСЯ НА КОНТЕКСНИХ ПОСИЛАННЯХ ТА МЕТОДИКА ВИЯВЛЕННЯ ЇЇ ОСНОВНИХ ЗОН

У роботі проаналізовано існуючи моделі для опису топології новинного вебпростору, які відображають його розподіл на зв'язані компоненти такі, як її цетральна частина й периферійні області та запропоновано нову мережеву модель тематичних Телеграм-каналів, яка базується на ідеї оцінки рівня цитування окремих інформаційних каналів та врахування прямих посилань у повідомленнях з каналів месенджера Телеграм. Вона поєднує в собі змістовний аспект повідомлень з можливістю урахування кількісних параметрів. Дослідження фокусується на каналах, присвячених темі кібербезпеки, і за часовим інтервалом охоплює перший квартал 2024 року. На основі аналізу близько 3500 Телеграм-каналів було виявлено понад 1000 каналів, на які здійснюються гіперпосилання та окреслено ключові зони інформаційного простору, такі як комунікаційна зона, комунікаційне ядро, вхідні та вихідні сегменти джерел. Сформована мережа визначена як безмасштабна, структурна мережа з самоподібними властивостями та степеневим розподілом ступенів вузлів, що підтверджує застосовність закону Парето для опису цього простору. Запропоновано математичну модель, яка дозволяє оцінювати поліноміальну залежність між обсягом комунікаційної області та загальною кількістю джерел. На прикладі системи контент-моніторингу інформаційних джерел "КіберАгрегатор" запропоновано методику автоматизованого розширення бази даних цільових джерел у системі, яка забезпечує динамічне збагачення списку інформаційних ресурсів через аналіз нових контекстних посилань у повідомленнях та алгоритм її реалізації.

Ключові слова: Телеграм-канали, контекстні посилання, інформаційний простір, кібербезпека, мережеві моделі, цитованість, комунікаційна взаємодія, контент-моніторинг, система "КіберАгрегатор".

Постановка проблеми. У сучасному світі, де інформація циркулює з неймовірною швидкістю, моніторинг та аналіз даних стають критично важливими для забезпечення безпеки в кіберпросторі. Одним із новітніх інструментів, що сприяють цьому процесу, є система "КіберАгрегатор" [1], яка надає вебінтерфейс для доступу до функцій пошуку та аналізу інформації в соціальних медіа.

Ця система покликана полегшити процес отримання важливої інформації з різних джерел, що дозволяє аналітикам оперативно реагувати на зміни в інформаційній ситуації. Важливим кроком у розвитку "Кіберагрегатора" стало переведення системи на технологію Big Data [2]. Застосування стандартних програмних засобів цього класу забезпечує мобільність і масштабованість, що дозволяє розглядати систему моніторингу та аналізу соціальних медіа з питань кібербезпеки як сучасну і потужну платформу.

Завдяки цим можливостям, система "КіберАгрегатор" здатна ефективно обробляти великі обсяги даних, виявляючи важливі тенденції та закономірності. У сучасних умовах, одним з основних джерел, з яких збирає інформацію "КіберАгрегатор" з метою їхньої подальшого аналізу, є Телеграм-канали.

Значний обсяг даних, що циркулює у месенджері Телеграм, може містити як достовірну інформацію, так і фейкові новини або пропагандистські матеріали, що створює виклики для

кібербезпеки, зокрема у сфері виявлення загроз, прогнозування атак та аналізу інформаційних операцій. Врахування контекстних повідомлень та взаємодій між Телеграм-каналами дозволяє створювати комплексні моделі інформаційного простору, які можуть бути використані для покращення методів моніторингу та захисту інформаційних систем.

Аналіз останніх досліджень і публікацій. З тематикою цієї статті повязані роботи різних авторів, що стосуються таких напрямів, як топологія новинного вебпростору мережі Інтернет, теорія комплексних (складних) мереж та її застосування, технології моніторингу соціальних мереж і Телеграм-каналів тощо.

Однією з ключових моделей для опису топології вебпростору є модель “галстук-бабочка” (bow-tie), запропонована Ендрю Брьодером [4], [5]. Вона відображає те, як вебпростір поділяється на великі зв’язні компоненти: центральну частину, що містить сильно зв’язний компонент; периферійні області, які вказують на входи до центрального вузла і виходи з нього. Такий підхід може бути застосований до аналізу інформаційної взаємодії у Телеграм-каналах, оскільки її структура, також, має центральні вузли й периферійні канали, що обмінюються інформацією.

Топологія новинного простору активно досліджується в контексті аналізу взаємодій між новинними агентствами, блогами та іншими джерелами інформації [6].

Дослідження взаємозв'язків між новинними сайтами показують, що більші медіа-ресурси концентрують найбільший потік інформації та формують центральні вузли інформаційних потоків. Водночас дрібніші ресурси залишаються на периферії, взаємодіючи з більшими через посилання або репости. Це нагадує структуру Телеграм-каналів, де великі канали домінують в інформаційному просторі, а менші канали їх розширюють.

Теорія комплексних мереж є фундаментом для дослідження складних систем, що складаються з великої кількості взаємодіючих елементів [7]. Однією з найбільш відомих моделей є безмасштабна модель Барабаші-Альберта [8]–[10], яка демонструє, як у багатьох реальних мережах невелика кількість вузлів має значно більше зв'язків порівняно з іншими. Це добре відображає структуру взаємодій в Телеграм-каналах, де певні канали є “центрами” інформаційних потоків, концентруючи велику кількість посилань і впливаючи на периферійні елементи мережі.

Технології моніторингу соціальних мереж [11] –[14] включають автоматизовані системи аналізу контенту, виявлення трендів та пошуку загроз. Сучасні рішення, засновані на машинному навчанні та обробці природної мови, дозволяють не тільки класифікувати повідомлення, але й знаходити нові джерела загроз або важливу інформацію, що впливає на безпеку [15]. Моніторинг Телеграм-каналів є особливо актуальним у контексті кібербезпеки через їхню популярність у політичних дискусіях, розповсюдженні новин та координації інформаційних операцій.

Формулювання цілей статті. Метою статті є розробка моделі та на її основі методики для розширення охоплення Телеграм-каналів як інформаційних джерел та аналізу їх взаємодії з метою виявлення кіберзагроз. Це дозволить дослідити структуру інформаційних взаємодій, ідентифікувати ключові канали поширення інформації та визначити можливі ризики, пов'язані з кіберзагрозами, дезінформацією або іншими небезпечними інформаційними потоками.

Для досягнення цієї мети вирішуються наступні задачі:

1. *Аналіз топології інформаційного простору Телеграм-каналів*. У рамках цієї задачі проводитьс детальний аналіз контекстних зв’язків між повідомленнями у Телеграм-каналах, що дозволяє виявити ключові канали, які відіграють роль у поширенні важливої або потенційно небезпечної інформації. Досліджуються структурні властивості цієї топології, використовуючи моделі, зокрема модель “галстука-бабочки” Ендрю Брьодера, яка описує потоки інформації у вебпросторі.

2. *Розробка мережевої моделі*, що відображає взаємодію між каналами та повідомленнями. Це дозволяє побудувати карту інформаційних потоків і визначити центри впливу, використовуючи методи теорії комплексних мереж. Особлива увага приділяється моделям типу “безмасштабна мережа” за класифікацією Альберт-Ласло Барабаші [6], яка може відображати динаміку поширення інформації.

3. *Розробка та тестування інформаційної технології*, яка дозволяє на практиці ефективно розширювати охоплення Телеграм-каналів. Дана технологія включає інструменти для автоматизованого пошуку і аналізу нових каналів, інтеграцію різних джерел інформації та створення механізмів для підвищення інформованості щодо важливих тем у кіберпросторі. Метою є не лише збільшення кількості залучених джерел, а й підвищення точності й ефективності моніторингу інформаційного простору.

4. *Застосування побудованої моделі* для виявлення кіберзагроз, зокрема для виявлення аномальних або небезпечних інформаційних потоків, пов'язаних із кіберзлочинністю чи дезінформаційними кампаніями. Це дозволяє створити ефективні інструменти моніторингу й аналізу для захисту інформаційного простору.

Вирішення наведених вище задач дозволяє розширити охоплення Телеграм-каналів як важливих джерел інформації, а також сприяє розвитку нових підходів до моніторингу і захисту інформаційного простору на основі моделй і методів теорії комплексних мереж і кібербезпеки.

Виклад основного матеріалу дослідження. В рамках роботи досліджується мережева модель розподілу джерел інформації – Телеграм-кналів, що базується на ідеї оцінки рівня цитування окремих інформаційних каналів та врахування прямих посилань у повідомленнях з каналів месенджера Телеграм. Ця модель поєднує в собі змістовний аспект повідомлень з можливістю урахування кількісних параметрів.

Було проаналізовано розподіл взаємного посилання Телеграм-каналів з питань кібербезпеки за перший квартал 2024 року. Усього досліджувалось близько 3500 Телеграм-каналів, серед яких таких, на які вели гіперпосилання на інші, було виявлено 1037.

Для ранжування каналів за кількістю вхідних і вихідних посилань було використано відсортовані значення ступенів, де для кожного вузла його положення в ранзі визначається кількістю вхідних d^{in} та вихідних d^{out} посилань.

Нехай $R(i)$ – це ранг вузла i за кількістю вхідних посилань, а $y_i = d_i^{in}$ – кількість посилань. Виявлено, що $R(i)$ підкоряється степеневій залежності:

$$y_i \sim R(i)^{-\alpha}, \tag{1}$$

де α – показник степені для розподілу ранжируваних посилань.

На рис. 1 наведено графік ранжируваного розподілу Телеграм-каналів з тематики кібербезпека, на які ведуть посилання інші канали.

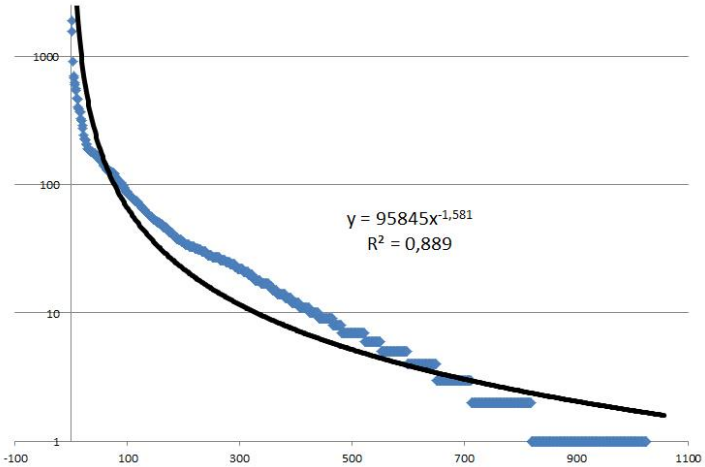

Рисунок 1 – Ранжирувана залежність розподілу Телеграм-каналів у логарифмічному масштабі (номер Телеграм-каналу – вісь X, кількість посилань – вісь Y)

Слід звернути увагу на те, що наведений графік з високою точністю демонструє степеневу залежність, тобто відповідає узагальненому закону Парето (на невелику кількість інформаційних каналів веде найбільша кількість посилань).

У результаті виконання спеціальної процедури для кожного повідомлення, що входить до певного джерела – Телеграм-каналу, були виявлені вхідні посилання на інші канали (посилання на власне джерело виключалися). З'ясувалося, що вхідні контекстні посилання були присутні у 36214 повідомленнях з 1224 Телеграм-каналів.

Побудована мережа взаємних посилань телеграм каналів виявилась безмасштабною, самоподібною, тобто такою, що має степеневий розподіл ступенів вузлів (рис. 2). З цього випливає, що навіть обмежена (але репрезентативна) множина каналів може давати достовірну інформацію щодо повної мережі, тобто простору тематичних Телеграм-каналів.

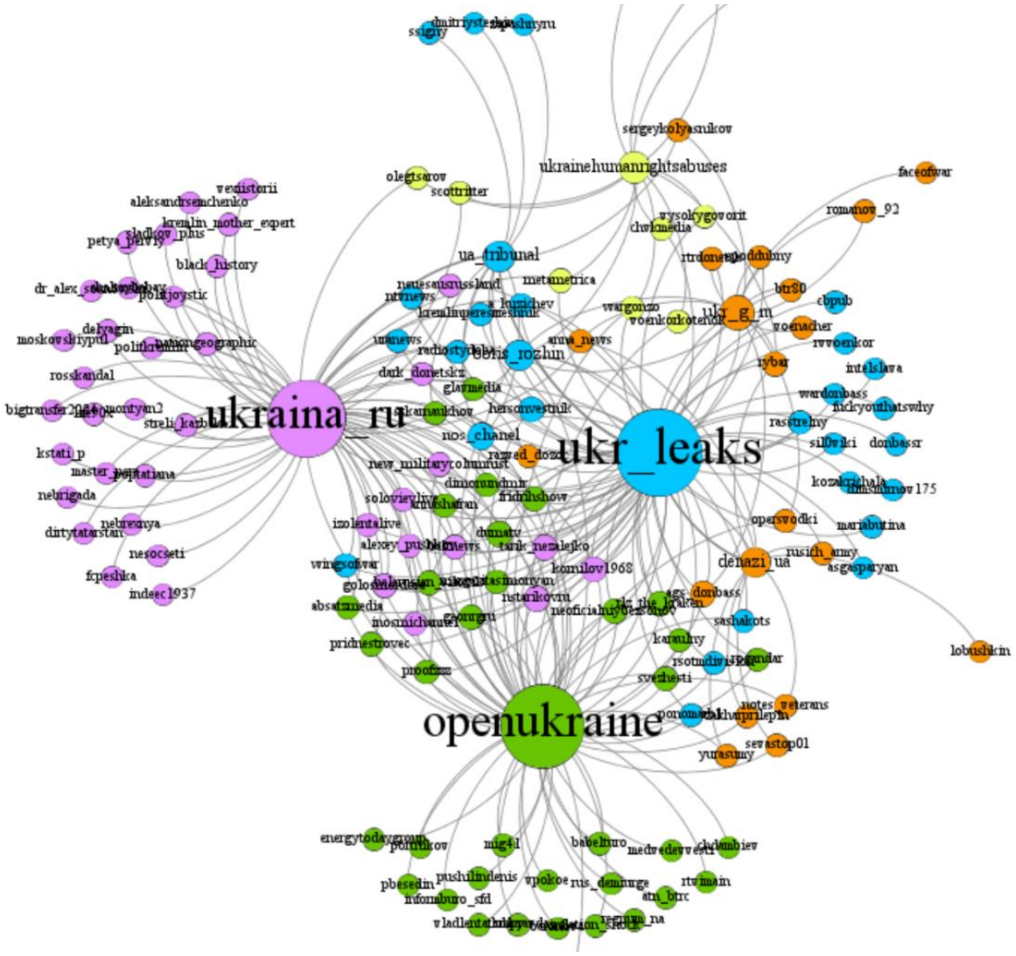

Рисунок 2 – Фрагмент комунікаційного ядра

За отриманими результатами була сформована модель простору тематичних Телеграм-каналів [12], яку наведено на рис. 2. Вона базується на аналізі контекстних гіперпосилань між повідомленнями і дозволяє виділяти різні зони взаємодії між каналами.

Математичні співвідношення між кількістю каналів у кожній зоні та кількістю посилань дають змогу проводити структурний аналіз і визначати ключові вузли інформаційної мережі.

Мережеву модель Телеграм-каналів можна представити у вигляді орієнтованого графа

$$G = (V, E),$$

де V – це множина вузлів (Телеграм-каналів);

E – множина орієнтованих ребер (гіперпосилань між каналами).

Граф G базується на гіперпосиланнях між повідомленнями каналів, що дає змогу моделювати інформаційні потоки у вигляді топології взаємодій між джерелами.

Нехай d_i^{in} і d_i^{out} – це ступені вхідних і вихідних посилань для вузла $v_i \in V$,

де d_i^{in} – кількість каналів, які посилаються на канал i;

d_i^{out} – кількість каналів, на які посилається канал i.

Як було визначено, мережа з високою точністю відповідає степеневому закону розподілу вхідних і вихідних посилань (1) та може бути представленою у вигляді (2):

$$P(k) \sim k^{-\gamma}, \tag{2}$$

де $P(k)$ – ймовірність того, що вузол має ступінь k, а γ – це показник степеня, що визначає характер розподілу, який у розглянутому випадку дорівнює 2,13.

Нехай загальна кількість каналів у системі позначається як N_{all}, та задано наступні зони:

NC – некомунікаційна зона – множина каналів, на які не посилаються інші канали і які не мають посилань на інші канали:

$$NC = N_{all} - N_{CZ}. \tag{3}$$

CZ – комунікаційна зона, яка складається з вхідних і вихідних сегментів:

$$N_{CZ} = N_{Input} + N_{Output} - N_{CK}, \tag{4}$$

де N_{Input} – кількість каналів, що мають вихідні посилання;

N_{Output} – кількість каналів, що мають вхідні посилання;

N_{CK} – комунікаційне ядро, що є перетином вхідного і вихідного сегментів.

Мережа є безмасштабною та самоподібною, тобто такою, що підпорядковується степеневому закону розподілу ступенів вузлів. Це означає, що навіть для обмеженої множини каналів можна отримати достовірні висновки про загальну структуру мережі [6].

Якщо взяти підмножину $S \subset V$ мережі і зібрати її статистику за кількістю посилань (зв'язків), отримаємо ту ж функцію степеневого розподілу (2), причому $P(k)$ – це ймовірність ступеня k у підмножині каналів S. Це дозволяє екстраполювати дані на всю мережу.

У наведеному прикладі, модель охоплює всі каналі (All) та включає наступні зони:

– множина Телеграм-каналів, які не мають посилань на інші Телеграм-канали і на які не посилаються інші Телеграм-канали (NC – некомунікаційна зона) – для множини досліджуваних каналів, приблизно 2500;

– вхідний сегмент джерел (Input). Телеграм-канали, з яких ведуть гіперпосилання на інші Телеграм-канали з виборки, що досліджується. Для досліджуваних каналів – близько 1000;

– вихідний сегмент джерел (Output). Телеграм-канали, на які ведуть гіперпосилання з інших Телеграм-каналів з виборки, що досліджується. Для досліджуваних каналів – також близько 1000;

– комунікаційне ядро – CK (Input та Output), що складається з перетину вхідного і вихідного сегменту джерел, для досліджуваних каналів. Їх виявилося близько 700. Фрагмент цієї області наведено на рисунку 3;

– комунікаційна зона – CZ (Input або Output), що складається з об'єднання вхідного і вихідного сегменту джерел. Для досліджуваних каналів їх виявилося близько 1300.

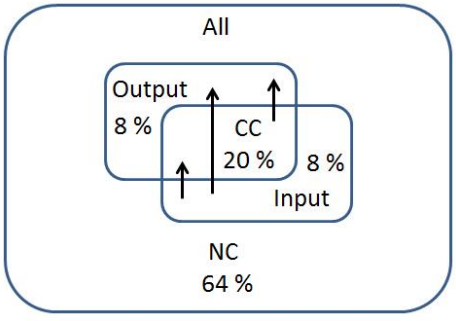

Рисунок 3 – Розподіл зон тематичних Телеграм-каналів для наведеного прикладу

Залежність обсягу комунікаційної області N_{CZ}, що охоплює сукупність вхідних і вихідних джерел (Input або Output) від загальної кількості джерел N_{all}, виявилася поліноміальною, що з високою точністю описується рівнянням другого порядку (дивись рис. 4):

$$N_{CZ} \sim aN_{all}^2 + bN_{all} + c\,, \tag{5}$$

де $a,\ b,\ c$ – це параметри, що визначають точність опису залежності.

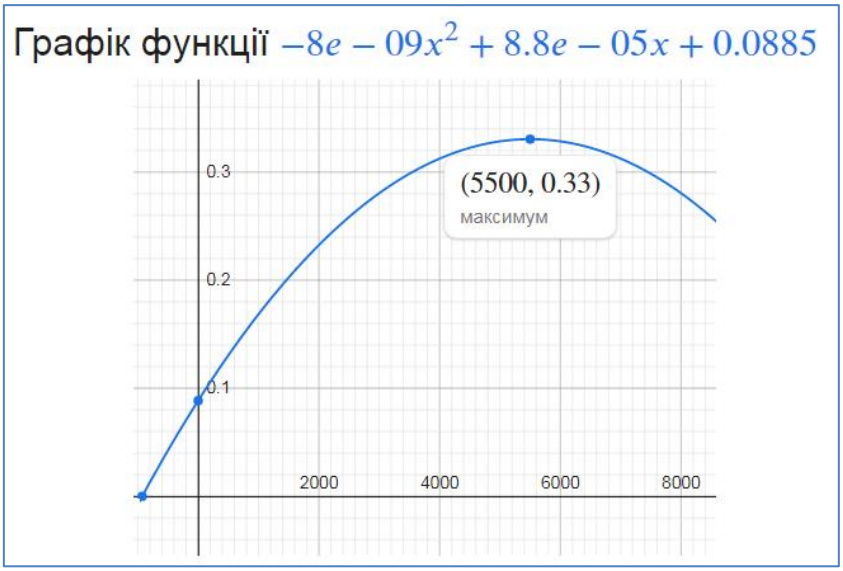

Рисунок 4 – Залежність обсягу комунікаційної області від загальної кількості джерел

На основі інформації щодо топології інформаційного простору Телеграм-каналів, для розширення його охоплення в рамках системи “КіберАгрегатор” було розроблено та реалізовано на практиці методику, яка забезпечує автоматизоване розширення переліку цільових інформаційних джерел за рахунок поступового “збагачення“ бази даних (БД) системи під час експлуатації.

Передбачається, що з початку здійснюється сканування деякого списку джерел з мережі Інтернет або соціальних мереж та месенджерів. Протягом певного часу в інформаційних повідомленнях, що поступають на вхід системи, з’являються посилання на деякі інформаційні джерела, що не входять до списку сканування. Саме ці посилання аналізуються інженерами зі знань і, при потребі, у випадку їх релевантності тематиці з кібербезпеки, включаються до списка джерел для сканування. Нерелевантні джерела можуть блокуватись (приховуватись в інтерфейсі користувача), а вже інсуючі в системі джерела можуть відображатися опціонально.

Для підключення нового джерела в систему контент-моніторингу користувачу достатньо ввести ключове слово, або його фрагмент у вікно пошуку, після чого йому буде запропоновано перелік знайдених нових джерел, які після перегляду можна підключити до системи, обравши функцію “Додати”.

На рис. 5 наведено фрагмент інтерфейсу програмного застосунку для знаходження релевантних джерел на прикладі Телеграм-каналів та їх перегляду у вебінтерфейсі месенджера Телеграм.

Рисунок 5 – Фрагмент інтерфейсу програмного застосунку знаходження релевантних запиту Телеграм-каналів

Нехай: S_0 – початковий список каналів або інформаційних джерел (вузлів мережі);

t – час функціонування системи;

$N(t)$ – кількість нових джерел, знайдених на момент часу t;

$R(t) \subset S$ – підмножина релевантних джерел на момент часу t;

$I(t) \subset S$ – підмножина нерелевантних джерел на момент часу t;

$P(t) \subset S$ – підмножина джерел, що вже існують у системі, але можуть бути відключені (опціонально).

Процес розширення охоплення джерел базується на аналізі нових інформаційних потоків. Припустимо, що потік нових посилань на інформаційні джерела визначається як функція $L(t)$:

$$L(t) = \{l_1, l_2, ..., l_n\}, \tag{6}$$

де l_i – посилання на джерело, яке не було присутнє у початковому списку S_0.

Система перевіряє кожне нове посилання l_i та визначає його релевантність через функцію (7):

$$f(l_i) = \begin{cases} 1, & \text{якщо } l_i \text{ релевантно тематиці з кібербезпеки} \\ 0, & \text{інакше} \end{cases}. \tag{7}$$

Тоді новий список релевантних джерел на момент часу t можна представити наступною множиною:

$$R(t) = \{l_i \in L(t) \mid f(l_i) = 1\}. \tag{8}$$

Далі, для оновлення БД, на кожному кроці часу (моменту) t, коли з'являються нові посилання $L(t)$, БД системи оновлюється за допомогою додавання нових релевантних джерел:

$$S(t+1) = S(t) \cup R(t). \tag{9}$$

Якщо $R(t) = \varnothing$, то система не оновлюється на момент часу t.

У випадку, коли джерела не є нерелевантними, їх можна приховати (відфільтрувати) за допомогою функції (10):

$$I(t) = \{ l_i \in L(t) \mid f(l_i) = 0 \}. \quad (10)$$

Ці джерела блокуються і не включаються до подальшого моніторингу.

Для опційного відображення вже наявних в системі джерел застосовується функція $g(j)$, яка визначає їх видимість:

$$g(j) = \begin{cases} 1, & \text{якщо джерело } j \text{ відображається} \\ 0, & \text{якщо джерело } j \text{ приховане} \end{cases}. \quad (11)$$

Тоді підмножина джерел, які можуть бути видимими в системі:

$$P(t) = \{ j \in S(t) \mid g(j) = 1 \}, \quad (12)$$

а прихованими, відповідно:

$$Z(t) = \{ j \in S(t) \mid g(j) = 1 \} \quad (13)$$

На підставі наведеного, можна запропонувати наступний алгоритм охоплення Телеграм-каналів, який забезпечує автоматизоване розширення переліку джерел, поступове збагачення БД та інтерактивне управління списком інформаційних джерел через додавання релевантних каналів і блокування нерелевантних.

Крок 1. *Ініціалізація*: ввести початковий список каналів S_0 та встановити часовий крок Δt для сканування нових джерел.

Крок 2. *Сканування нових джерел*: на момент часу t отримати новий потік інформації з посиланнями $L(t)$.

Для кожного нового джерела $l_i \in L(t)$ визначити релевантність $f(l_i)$.

Крок 3. *Оновлення БД*:

Якщо $f(l_i) = 1$, додати l_i до бази даних: $S(t+1) = S(t) \cup R(t)$.

Інакше, якщо $f(l_i) = 0$, відправити l_i до списку блокованих джерел $I(t)$.

Крок 4. *Опціональне відображення джерел*: визначити видимість існуючих джерел $P(t)$ за допомогою функції $g(j)$ та дозволити користувачу приховувати або відображати джерела за допомогою інтерфейсу.

Крок 5. *Додавання нового джерела*: вручну ввести ключове слово або фрагмент для пошуку та отримати від системи згенерований список відповідних нових джерел $R_{search}(t)$.

Користувач переглядає список і, у разі необхідності, додає джерела до БД системи.

Висновки. Таким чином, у результаті проведеного дослідження було побудовано модель простору тематичних Телеграм-каналів, що базується на контекстних посиланнях, а також запропоновані підходи до виявлення основних зон моделі простору тематичних Телеграм-каналів та розраховані числові співвідношення різних зон моделі.

У процесі розробки методики розширення охоплення Телеграм-каналів та реалізації її на практиці в рамках системи “КіберАгрегатор”, було досягнуто ряд результатів, що відповідають поставленим задачам, а саме:

– створена система **автоматизованого сканування інформаційних джерел**, яка дозволяє автоматично сканувати їх початковий список, а також оперативно виявляти нові релевантні джерела, що з’являються в інформаційному просторі. Це значно підвищує ефективність збору даних для моніторингу актуальної інформації у сфері кібербезпеки;

– запропоновано нову функцію аналізу релевантності нових джерел, яка дозволяє адаптувати систему до змін в інформаційному середовищі, при цьому, нерелевантні джерела

фільтруються, що забезпечує підвищення якості отриманої інформації і зменшує "інформаційний шум";

– реалізовано можливість безпосереднього підключення нових джерел за ключовими словами, що дозволяє користувачеві активно впливати на наповненість системи. Це дає змогу користувачам працювати з системою в інтерактивному режимі;

– забезпечено опціональне відображення існуючих джерел, що дозволяє користувачам самостійно керувати інформаційним контентом у межах інтерфейсу, тим самим підвищуючи, завдяки гнучкості у налаштуваннях, ефективність виконання ним завдань.

Таким чином, наукова новизна отриманого результату, полягає у застосуванні нових підходів до автоматизованого моніторингу інформаційних джерел на базі Телеграм-каналів з використанням концепції інформаційного простору. Створена модель, методика та алгоритм автоматичного розширення бази даних, які базуються на моделях і методах теорії комплекних мереж і дозволяють встановлювати нові зв'язки між джерелами інформації.

Практична значимість отриманого результату полягає у можливості розробки інтерактивної платформи для моніторингу Телеграм-каналів, що забезпечує автоматизоване та ручне управління інформаційними джерелами та може бути корисною для різних користувачів, зокрема аналітиків, фахівців з кібербезпеки та дослідників, що займаються вивченням інформаційних тенденцій.

Перспективи подальших досліджень. Запропонована модель припускає подальше вдосконалювання в наступних напрямках: більш точної ідентифікації контекстних посилань, удосконалювання критерію визначення зон на основі повного урахування структури посилань та методів кластерного аналізу, а також удосконалювання механізму визначення змістовного дублювання інформації.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] D. Lande, I. Subach, and O. Puchkov, "System of Analysis of Big Data from Social Media, *Information & Security: An International Journal*, no. 47 (1), pp. 44-61, 2020. doi: https://doi.org/10.11610/isij.4703.

[2] Д.В. Ланде, І.Ю. Субач, та А.Я. Гладун, *Оброблення надвеликих масивів даних* (*Big Data*): *навч. пос.* Київ, Україна: ТОВ "Інжиніринг", 2021.

[3] D. Boyd, and K. Crawford, "Critical questions for Big Data", *Journal Information, Communication & Society*, vol. 15, iss. 5, pp. 662-679, 2012, doi: https://doi.org/10.1080/1369118X.2012. 678878.

[4] A. Broder, R. Kumar, F. Maghoul, et al., "Graph structure in the Web", *Computer Networks*. vol. 33, iss. 1-6, 2000, pp. 309-320. doi: https://doi.org/10.1016/S1389–1286(00)00083-9.

[5] R. Meusel, S. Vigna, O. Lehmberg, and C. Bizer, "Graph structure in the web – revisited: a trick of the heavy tail", in *Proc. 23rd International Conference on World Wide Web* (*WWW'14 Companion*), Seoul, Korea, 2014, pp. 427-432. doi: https://doi.org/10.1145/2567948.2576928.

[6] О.Г. Додонов, Д.В. Ланде, та В.Г. Путятин, *Інформаційні потоки в глобальних комп'ютерних мережах*. Київ, Україна: Наукова думка, 2009.

[7] А.О. Снарський, Д.В. Ланде, та І.Ю. Субач, *Основи теорії складних мереж*: *навч. пос.* Київ, Україна: ТОВ "Інжиніринг", 2023.

[8] A. Réka, and A.L. Barabási, "Statistical mechanics of complex networks", *Reviews of Modern Physics*, no. 74 (47), pp. 47-97, 2002. doi: https://doi.org/10.1103/RevModPhys.74.47.

[9] A. Réka, J. Hawoong, and A.L. Barabási, "Error and attack tolerance of complex networks", *Nature*, vol. 406, pp. 378-382, 2000. [Online]. Available: https://www.nature.com/articles/35019019. Accessed on: July 19, 2024.

[10] A.L. Barabási, and A. Réka, "Emergence of scaling in random networks", *Science*, vol. 286, iss. 5439, pp. 509-512, 1999. doi: https://doi.org/10.1126/science.286.5439.509.

[11] G. Szabo, G. Polatkan, P.O. Boykin, and A. Chalkiopoulos, *Social Media Data Mining and Analytics*. Chichester, England: John Wiley & Sons Inc., 2018.

[12] K. Cherven, *Mastering Gephi Network Visualization*. Birmingham, England: Packt Publishing, 2015.
[13] I.V. Bezsudnov, and A.A. Snarskii, “From the time series to the complex networks: The parametric natural visibility graph”, *Physica A: Statistical Mechanics and its Applications*, vol. 414, pp. 53-60. doi: https://doi.org/10.1016/j.physa.2014.07.002.
[14] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J.C. Nuno, “From time series to complex networks: The visibility graph”, *Proceedings of the National Academy of Sciences*, vol. 105 (13), pp. 4972-4975, 2008. Doi: https://doi.org/10.1073/pnas.0709247105.
[15] D. Lande, E. Shnurko-Tabakova, “OSINT as a part of cyber defense system”, *Theoretical and Applied Cybersecurity*, no. 1, pp. 103108, 2019. [Online]. Available: http://tacs.ipt.kpi.ua/article/view/169091/168863. Accessed on: July 10, 2024.

Стаття надійшла 27.10.2024.

REFERENCE

[1] D. Lande, I. Subach, and O. Puchkov, “System of Analysis of Big Data from Social Media, *Information & Security: An International Journal*, no. 47 (1), pp. 44-61, 2020. doi: https://doi.org/10.11610/isij.4703.
[2] D. Lande, I. Subach, and A. Gladun, *Processing of ultra-large data arrays (Big Data): a textbook.* Kyiv, Ukraine: LTD “Engineering”, 2021.
[3] D. Boyd, and K. Crawford, “Critical questions for Big Data”, *Journal Information, Communication & Society*, vol. 15, iss. 5, pp. 662-679, 2012, doi: https://doi.org/10.1080/1369118X.2012. 678878.
[4] A. Broder, R. Kumar, F. Maghoul, et al., “Graph structure in the Web”, *Computer Networks*. vol. 33, iss. 1-6, 2000, pp. 309-320. doi: https://doi.org/10.1016/S1389–1286(00)00083-9.
[5] R. Meusel, S. Vigna, O. Lehmberg, and C. Bizer, “Graph structure in the web – revisited: a trick of the heavy tail”, in *Proc. 23rd International Conference on World Wide Web* (*WWW'14 Companion*), Seoul, Korea, 2014, pp. 427-432. doi: https://doi.org/10.1145/2567948.2576928.
[6] O. Dodonov, D. Lande, and V. Putiatyn, *Information flows in global computer networks*. Kyiv, Ukraine: Naukova dumka, 2009.
[7] A. Snarskyi, D. Lande, and I. Subach, *Fundamentals of the theory of complex networks*: *textbook.* Kyiv, Ukraine: LTD “Engineering”, 2023.
[8] A. Réka, and A.L. Barabási, “Statistical mechanics of complex networks”, *Reviews of Modern Physics*, no. 74 (47), pp. 47-97, 2002. doi: https://doi.org/10.1103/RevModPhys.74.47.
[9] A. Réka, J. Hawoong, and A.L. Barabási, “Error and attack tolerance of complex networks”, *Nature*, vol. 406, pp. 378-382, 2000. [Online]. Available: https://www.nature.com/articles/35019019. Accessed on: July 19, 2024.
[10] A.L. Barabási, and A. Réka, “Emergence of scaling in random networks”, *Science*, vol. 286, iss. 5439, pp. 509-512, 1999. doi: https://doi.org/10.1126/science.286.5439.509.
[11] G. Szabo, G. Polatkan, P.O. Boykin, and A. Chalkiopoulos, *Social Media Data Mining and Analytics*. Chichester, England: John Wiley & Sons Inc., 2018.
[12] K. Cherven, *Mastering Gephi Network Visualization*. Birmingham, England: Packt Publishing, 2015.
[13] I.V. Bezsudnov, and A.A. Snarskii, “From the time series to the complex networks: The parametric natural visibility graph”, *Physica A: Statistical Mechanics and its Applications*, vol. 414, pp. 53-60. doi: https://doi.org/10.1016/j.physa.2014.07.002.
[14] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J.C. Nuno, “From time series to complex networks: The visibility graph”, *Proceedings of the National Academy of Sciences*, vol. 105 (13), pp. 4972-4975, 2008. Doi: https://doi.org/10.1073/pnas.0709247105.
[15] D. Lande, E. Shnurko-Tabakova, “OSINT as a part of cyber defense system”, *Theoretical and Applied Cybersecurity*, no. 1, pp. 103108, 2019. [Online]. Available: http://tacs.ipt.kpi.ua/article/view/169091/168863. Accessed on: July 10, 2024.

OLEXANDR PUCHKOV,
DMYTRO LANDE,
IHOR SUBACH

A MODEL OF THE SPACE OF THEMATIC TELEGRAM CHANNELS BASED ON CONTEXTUAL LINKS

The paper analyzes the existing models for describing the topology of the news web space, which reflect its division into coherent components such as its central part and peripheral areas, and proposes a new network model of thematic Telegram channels based on the idea of assessing the level of citation of individual information channels and taking into account direct links in messages from Telegram channels. It combines the content aspect of messages with the ability to take into account quantitative parameters. The study focuses on channels dedicated to cybersecurity and covers the first quarter of 2024. Based on the analysis of about 3,500 Telegram channels, more than 1,000 hyperlinked channels were identified and key areas of the information space, such as the communication zone, the communication core, and incoming and outgoing source segments, were outlined. The formed network is defined as a scale-free, structural network with self-similar properties and a power law distribution of node degrees, which confirms the applicability of the Pareto law to describe this space. A mathematical model is proposed that allows estimating the polynomial relationship between the volume of the communication area and the total number of sources. On the example of the CyberAggregator content monitoring system for information sources, a methodology for automated expansion of the database of target sources in the system is proposed, which provides dynamic enrichment of the list of information resources through the analysis of new contextual references in messages and an algorithm for its implementation.

Keywords: Telegram channels, contextual links, information space, cybersecurity, network models, citation, communication interaction, content monitoring, CyberAggregator system.

Пучков Олександр Олександрович, кандидат філософськихнаук, професор, начальник, Інститутут спеціального зв'язку та захисту інформації Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна, ORCID 0000-0002-8585-1044, iszzi@iszzi.kpi.ua.

Ланде Дмитро Володимирович, доктор технічних наук, професор, завідувач кафедри інформаційної безпеки, Навчально-науковий фізико-технічний інститут Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", Київ,Україна, ORCID 0000-0003-3945-1178, dwlande@gmail.com.

Субач Ігор Юрійович, доктор технічних наук, професор, завідувач кафедри кібербезпеки і застосування інформаційних систем і технологій, Інститут спеціального зв'язку та захисту інформації Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна, ORCID 0000-0002-9344-713X, igor_subach@ukr.net.

Puchkov Oleksandr, PhD in philosophy, professor, head of the Institute of special communication and information protection at the National technical university of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

Lande Dmytro, doctor of technical sciences, professor, chair of the academic department of the information security, Educational and scientific physico-technical institute at the National technical university of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

Subach Ihor, doctor of technical sciences, professor, chair of the academic department of the cyber security and application of information systems and technologies, Institute of special communication and information protection at the National technical university of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.