КІБЕРБЕЗПЕКА ТА ЗАХИСТ ІНФОРМАЦІЇ

УДК 004.8-9 https://doi.org/10.32515/2664-262X.2025.11(42).2.70-78

О. В. Лозинська, **О. О. Марків**, доц., канд. техн. наук, **В. А. Висоцька**, доц., д-р техн. наук
Національний університет «Львівська політехніка», м. Львів, Україна
e-mail: olha.v.lozynska@lpnu.ua, oksana.o.markiv@lpnu.ua, victoria.a.vysotska@lpnu.ua

Метод виявлення розповсюджувачів дезінформації на основі графового представлення структури соціальної мережі

У роботі розглянуто проблему поширення та виявлення дезінформації в соціальних мережах. На основі проведеного аналізу, досліджено, що проблема виявлення дезінформації тісно пов'язана насамперед з проблемою виявлення розповсюджувачів даної дезінформації. Запропоновано метод виявлення розповсюджувачів дезінформації на основі представлення соціальної мережі у вигляді орієнтованого графа, з поділом користувачів на спільноти. Введено класифікацію вузлів (сусідні, граничні та основні) та показано, яку роль вони відіграють у поширенні інформації. Проведено аналіз україномовного датасету новин, що включає понад 2000 записів, з метою виявлення користувачів, які потенційно можуть стати розповсюджувачами інформації (як правдивої так і неправдивої). Отримані результати демонструють, що граничні та окремі основні вузли є ключовими для передачі фейкової інформації всередині спільнот. Запропонований підхід дає змогу підвищити ефективність виявлення потенційних розповсюджувачів дезінформації.

дезінформація, фейкові новини, датасет, розповсюджувачі дезінформації, графи

Постановка проблеми. Соціальні мережі є невід'ємною частиною повсякденного життя. Люди використовують різні платформи соціальних мереж для спілкування з близькими та рідними, а також використовують їх як основне джерело новин. Проте, не завжди соціальні медіа-платформи містять правдиву інформацію. Дуже часто вони використовуються для масового поширення дезінформації. Науковці з усього світу намагаються розробити ефективні методи та засоби для розпізнавання неправдивої інформації, так званих фейкових новин. Все частіше дослідники зосереджуються лише на методах виявлення достовірності інформації, але важливо не тільки виявляти фейкову інформацію, але й визначати людей, які її читають і відповідно поширюють. Розроблення стратегій виявлення дезінформації може допомогти стримати та запобігти швидкому її поширенню в соціальних мережах. Таким чином, постає проблема не тільки виявлення фейкових новин, а насамперед виявлення їх розповсюджувачів. Серед способів поширення дезінформації також потрібно виділити координовану неавтентичну поведінку користувачів акаунтів. Щоб виявити координовану неавтентичну поведінку, потрібно або використовувати подібність у поведінці користувачів, проаналізовану за допомогою мереж або часову синхронність між користувачами. Авторами статті досліджено відомі методи виявлення розповсюджувачів інформації у соціальних мережах, проведено аналіз цих методів та запропоновано використання графів для подання спільнот користувачів у соціальних мережах. На основі цього методу обчислено кількість основних, граничних та сусідніх вузлів у датасеті україномовних новин. Також досліджено, які вузли можуть стати розповсюджувачами інформації.

Аналіз останніх досліджень і публікацій. Більшість робіт що стосуються виявлення дезінформації зосереджено на виявленні вмісту новини або самої новини.

Малий відсоток наукових робіт, які зосереджені на виявленні користувачів, що найімовірніше поширюють фейкові новини. Перші наукові роботи в цій області використовували загальні функції метаданих користувача, такі як кількість підписників, ім'я профілю користувача, електронну адресу тощо для виявлення підозрілих профілів. Науковці [1] проаналізували понад 100 мільйонів повідомлень, зібраних із Twitter протягом одного місяця та виділили дві категорії спамерів, що відрізняються за поведінкою, і які використовують різні стратегії розсилання спаму.

У роботі [2] досліджено автоматичні методи оцінки достовірності певного набору твітів. Автори аналізують дописи мікроблогів, пов'язані з «трендовими» темами, і класифікують їх як такі, які заслуговують довіри, або не заслуговують. Для цього використовуються особливості поведінки користувачів при розміщенні та повторному розміщенні публікацій («повторне твітування»). Результати досліджень показують, що існують вимірювані відмінності в способах поширення повідомлень, які можна використовувати для автоматичної класифікації їх як достовірних чи недостовірних, з точністю та запам'ятовуванням у діапазоні від 70% до 80%.

Науковці [3] проводили дослідження психологічного профілю людей, які стають жертвами фейкових новин. Результати їхнього дослідження свідчать про те, що віра у фейкові новини певною мірою може бути зумовлена загальною тенденцією до надмірного сприйняття людиною слабких тверджень. Автори статті [4] проаналізували характеристики та мотиваційні фактори розповсюджувачів фейкових новин у соціальних мережах за допомогою психологічних теорій та досліджень поведінки. У наукових роботах [5, 6] автори створили реальні набори даних, які вимірюють рівень довіри користувачів до фейкових новин, і виділили репрезентативні групи як «досвідчених» користувачів, які здатні розпізнавати фейкові новини як неправдиві, так і «наївних» користувачів, які з більшою ймовірністю повірять фейковим новинам. Після цього провели порівняльний аналіз явних і прихованих характеристик профілю між цими групами користувачів, що розкриває їхній потенціал для розрізнення фейкових новин.

У роботі [7] запропоновано методику для визначення, чи буде автор стрічки Twitter поширювати фейкові новини. Автори показали можливість автоматичного визначення потенційних розповсюджувачів фейкових новин у Twitter і труднощі їх ідентифікації. Авторами [8] запропоновано структуру на основі керованого машинного навчання для профілювання розповсюджувачів фейкових новин. Розроблений метод ґрунтується на поєднанні індивідуальних особливостей великої п'ятірки (привітність, сумлінність, екстраверсія, емоційний діапазон, відкритість) та стилометричних характеристик.

У роботі [9] запропоновано метод визначення чи хоче автор каналу Twitter бути розповсюджувачем фейкових новин з використанням класифікатора методу опорних векторів (SVM) із функціями n-грам символів і слів. Автори [10] також використали символьні n-грами як функції в поєднанні з лінійною SVM для англійської мови та логістичної регресії для іспанської мови. Їхні моделі досягають загальної точності 73% і 79% на офіційних тестах англійською та іспанською мовами відповідно.

У праці [11] запропоновано модель CheckerOrSpreader, яка може класифікувати користувача як такого, що перевіряє факти, або потенційного розповсюджувача фейкових новин. Розроблена модель заснована на згортковій нейронній мережі і поєднує вбудовування слів із функціями, які представляють особистісні риси користувачів і мовні моделі, що використовуються в їхніх твітах.

Інші наукові дослідження продемонстрували переваги графових моделей на основі уваги для моделювання та виявлення чуток [12]. У науковій роботі представлено модель глибокої уваги на основі рекурентних нейронних мереж для ідентифікації чуток. Масштабні експерименти на реальних наборах даних, зібраних із веб-сайтів соціальних мереж, демонструють, що модель на основі глибокої уваги перевершує ті моделі, які покладаються на функції, створені вручну та здатна виявляти чутки швидше і точніше.

Для виявлення фейкових новин автори статті [13] запропонували мережеву модель згортки графів, яка використовує шляхи поширення для виявлення фейкових новин. Авторами розроблено нову модель двонаправленого графа під назвою Bi-Directional Graph Convolutional Networks (Bi-GCN), щоб досліджувати поширення чуток як зверху вниз, так і знизу вгору.

Науковці [14] розробили мережу уваги з урахуванням графів, яка використовує уявлення користувачів і шляхи поширення інформації, щоб передбачити фейкові новини. У роботі [15] запропоновано структуру індуктивного навчання FANG (Factual News Graph), яка використовує графові нейронні мережі для представлення соціальної структури та виявлення фейкових новин. FANG краще вловлює соціальний контекст у високоточному представленні порівняно з іншими графічними та неграфічними моделями. Зокрема, FANG дає значні покращення для завдання виявлення фейкових новин і є надійним у випадку обмежених навчальних даних.

У праці [16] досліджено розповсюдження всіх підтверджених правдивих і неправдивих новин, поширених у Twitter з 2006 по 2017 рік. Дані містять приблизно 126 000 історій, які твітнули близько 3 мільйонів людей понад 4,5 мільйона разів. Результати досліджень виявили, що неправдиві новини були більш новими, ніж правдиві, що свідчить про те, що люди частіше ділилися новою інформацією. Крім того, фейки поширювались значно далі й швидше, ніж правда, у всіх категоріях інформації. Тоді як неправдиві історії викликали у відповідях страх, огиду та здивування, правдиві історії викликали очікування, смуток, радість та довіру.

Автори [17] запропонували баєсівську непараметричну модель для розуміння ролі контенту в поширенні правдивих і неправдивих новин і відмінностей між ними. Розроблена модель була навчена на базі даних реальної соціальної мережі та здатна краще прогнозувати мітки новин, ніж сучасніші нейронні мережі та моделі Баєса.

Ще одним із способів поширення дезінформації є координована неавтентична поведінка користувачі, яка проявляється у створенні кількох фейкових акаунтів із залученням якнайбільше користувачів та поширенням однакової інформації (новин, постів, твітів) за певний період. При цьому ці аккаунти посилаються один на одного або на одне й те саме джерело. Такі новини найчастіше подаються із заголовком, який одразу привертає увагу (написаний великими літерами, із знаками оклику або ж це якась сенсація). Метою поширення такої інформації є ввести читачів в оману, створити для них інформаційну ілюзію, коли вони будуть читати лише повідомлення, які містять дезінформацію.

Щоб виявити координовану неавтентичну поведінку, потрібно або використовувати подібність у поведінці користувачів, проаналізовану за допомогою мереж або часову синхронність між користувачами. Перший підхід зазвичай включає такі аналітичні кроки, як побудова зваженої мережі подібності користувача, фільтрація такої мережі, виявлення різних скоординованих спільнот у мережі та вивчення їх масштабу та моделей координації. Провівши аналіз наукових досліджень, варто зазначити, що для розрізнення автентичної та неавтентичної поведінки користувачів потрібно розробити методи їх ідентифікації. Більшість вчених просто вважають, що

певні рівні координації завжди вказують на зловмисну поведінку. Інші дослідники натомість запропонували поєднати аналіз координації та пропаганду, як спосіб ідентифікації шкідливих скоординованих спільнот.

На основі проведеного аналізу наукових робіт, які переважно аналізують вміст повідомлення (новини), наш підхід використовує графове представлення структури соціальної мережі для подання зв'язків між користувачами.

Постановка завдання. Метою дослідження є розроблення методу виявлення розповсюджувачів дезінформації з використанням графового представлення структури соціальної мережі. Для вирішення цього завдання потрібно: провести аналіз методів та стратегій виявлення дезінформації, які можуть допомогти стримати та запобігти швидкому її поширенню в соціальних мережах; використати графові структури для подання зв'язків між користувачами у соціальній мережі; на основі датасету україномовних новин здійснити підрахунки та передбачити користувачів, які, швидше за все, стануть розповсюджувачами інформації (правдивої чи неправдивої).

Виклад основного матеріалу. Для того, щоб зрозуміти як поширюються новини у соціальній мережі, представимо цю мережу у вигляді зв'язного орієнтованого графа, де кожна вершина (вузол) являє собою користувача. Користувачі пов'язані між собою ребрами та об'єднані у різні спільноти. Рівень довіри усередині спільноти значно вищий, ніж між користувачами різних спільнот. Таким чином, якщо всередині такої спільноти поширюється дезінформація, ймовірність зараження всіх членів спільноти буде високою. Для прикладу розглянемо граф, зображений на рис. 1.

Граф включає сім спільнот вузлів (вершин), де індекс вузла позначає спільноту, до якої він належить. У контексті будь-якої соціальної мережі орієнтоване ребро $L_1 \rightarrow K_1$ означає, що L_1 стежить за K_1. Таким чином, будь-яка новина (пост) йде від K_1 до L_1. Якщо L_1 вирішує зробити репост новини K_1, це значить, що L_1 схвалив новину K_1 і що L_1 довіряє K_1. Червоним кольором зображені вузли (користувачі), які поширили неправдиву інформацію, синім кольором – інші вузли.

Оскільки довіра усередині спільноти висока, то чим більше L довіряє K, тим вищий шанс, що L зробить репост поста K і таким чином поширить повідомлення K, незалежно від того, чи воно правдиве або неправдиве. На рис. 1 показано поширення фейкових новин, починаючи з N_1, коли вони поширюються мережею через K_3 до K_7.

Як приклад, розглянемо два сценарії виявлення розповсюджувача фейкових новин:

1. Інформація досягає границь спільноти. Розглянемо приклад, коли повідомлення поширюється через N_1, який є вузлом, що з'єднує спільноти 1 і 3. Вузол K_3 піддається впливу та, ймовірно, поширить інформацію, таким чином розпочавши поширення інформації в заданій пов'язаній спільноті. Таким чином, важливо передбачити вузли на границі спільнот, які, ймовірно, стануть розповсюджувачами інформації.

2. Інформація проникає в спільноту. Розглянемо приклад, коли вузол K_3 вирішує поширити повідомлення. Вузли L_3, N_3 і O_3, які є безпосередніми послідовниками K_3, тепер доступні для інформації. Через їхню близькість вони вразливі для того, щоб повірити даному вузлу. Подібним чином, для спільноти 7, коли повідомлення досягло вузла K_7, вузли N_7 і R_7 знаходяться на відстані одного кроку. Інтуїтивно зрозуміло, що в тісній структурі спільноти, якщо один із вузлів вирішить поширити частину інформації, ймовірність її швидкого поширення серед усієї спільноти дуже висока. Таким чином, важливо виявляти вузли в межах спільноти, які, ймовірно, стануть розповсюджувачами інформації, щоб захистити здоров'я всієї спільноти.

Розглянемо основні принципи оцінки здоров'я спільноти [18]. Соціальна мережа має характерну властивість демонструвати структури спільноти, які формуються на

основі взаємодії між вершинами. Спільноти, як правило, є модульними групами, де члени всередині групи тісно пов'язані, а учасники міжгрупових зв'язків слабкі. Таким чином, члени спільноти, як правило, мають вищий рівень довіри один до одного, ніж між членами різних спільнот. Якщо такі спільноти піддаються фейковим новинам, що поширюються поблизу нього, ймовірність зараження всіх членів спільноти буде високою. Керуючись ідеєю легкого поширення в спільноті, використано модель оцінки стану здоров'я в спільноті [18]. Модель визначає три типи вузлів по відношенню до спільноти: сусідні, граничні та основні вузли:

1. Сусідні вузли: ці вузли безпосередньо приєднані принаймні до одного вузла спільноти. Набір сусідніх вузлів позначається $N_{сп}$. Вони не є частиною спільноти.

2. Граничні вузли: це вузли спільноти, які безпосередньо приєднані принаймні до одного сусіднього вузла. Множину граничних вузлів позначимо $B_{сп}$. Важливо зазначити, що лише вузли спільноти, які мають граничне ребро до сусідніх вузлів, знаходяться в $B_{сп}$.

3. Основні вузли: це вузли спільноти, які пов'язані лише з членами спільноти. Набір основних вузлів позначається $C_{сп}$.

У табл. 1 зазначено сусідні, граничні та основні вузли для спільнот, що зображені на рис. 1.

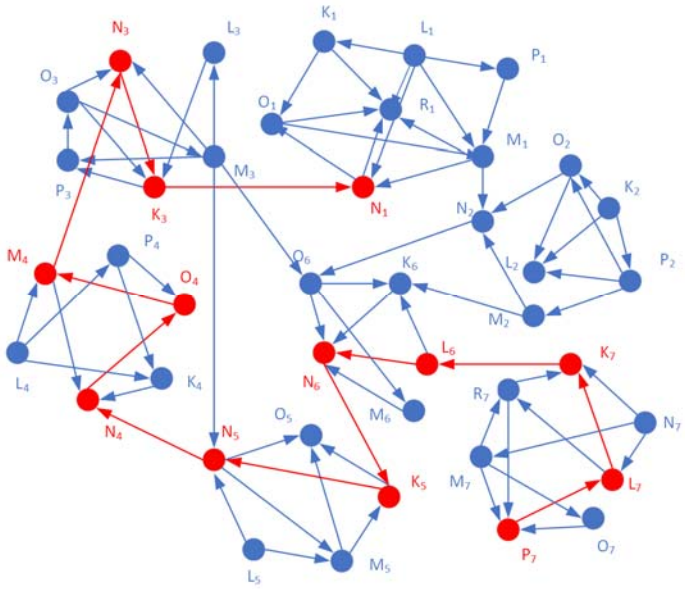

Рисунок 1 – Приклад графового представлення структури мережі спільнот для виявлення розповсюджувачів фейкових новин

Джерело: розроблено на підставі [18]

Таблиця 1 – Сусідні, граничні та основні вузли для спільнот користувачів соціальної мережі

№ Сп	$N_{сп}$	$B_{сп}$	$C_{сп}$
1	N_2	M_1	K_1, L_1, N_1, O_1, P_1, R_1
2	K_6, O_6	M_2, N_2	K_2, L_2, O_2, P_2
3	N_1, N_5, O_6	K_3, M_3	L_3, N_3, O_3, P_3
4	N_3	M_4	K_4, L_4, N_4, O_4, P_4
5	N_4	N_5	K_5, L_5, M_5, O_5
6	K_5	N_6	K_6, L_6, M_6, O_6
7	L_6	K_7	L_7, M_7, N_7, O_7, P_7, R_7

Джерело: розроблено авторами

Для даного завдання було проаналізовано датасет [19], що містить 2174 записів новин з різною інформацією. Кожен запис містить дату, час, текст самого повідомлення, мітку (правда-фейк), пост/репост, джерело, автор повідомлення, мову, якою написано повідомлення, кількість поширень, кількість вподобайок, та покликання на саме джерело. Таким чином, можна прослідкувати коли різні користувачі аккаунтів почали поширювати одну й ту саму неправдиву інформацію. Кількість записів із міткою «правда» становить 1283, а з міткою «фейк» – 891 запис.

Запропонований метод виявлення розповсюджувачів інформації полягає у наступному. Припустимо, що є певна соціальна мережа (ω), яка складається зі спільнот користувачів ($com \in \omega$) і кожна спільнота має чітко визначені сусідні вузли ($N_{сп}$), граничні вузли ($B_{сп}$) і основні вузли ($C_{сп}$). Нам потрібно передбачити граничні вузли b, які, швидше за все, стануть розповсюджувачами інформації ($b_{роз}$) у спільноті користувачів, оскільки вони найбільш дотичні до всіх решта вузлів спільноти. Таким же ж чином потрібно зробити обрахунки, щоб передбачити основні вузли c, які, швидше за все, стануть розповсюджувачами інформації усередині самої спільноти ($c_{роз}$). Статистику датасету україномовних новин із зазначенням кількості різних типів вузлів, ребер та підрахунку розповсюджувачів інформації наведено у табл. 2. Таким чином, якщо дезінформація потрапить до граничних вузлів $b_{роз}$ (у нашому датасеті таких є 11), вона розповсюдиться по всій спільноті. Кількість розповсюджувачів $b_{роз}$ правдивої інформації згідно обрахунків становить 14. Крім того, виявлено, що найчастіше поширюють фейкову інформацію користувачі, в яких більшість постів на сторінці є фейковими. Це може свідчити, що даний аккаунт є фейковим.

Таблиця 2 – Статистика зібраного датасету україномовних новин

	Фейк	**Правда**
Кількість вузлів	451	645
Кількість ребер	1532	1967
Кількість розповсюджувачів	58	97
Кількість вузлів у N	115	182
Кількість вузлів $n_{роз}$ у N	39	73
Кількість вузлів у B	102	204
Кількість вузлів $b_{роз}$ у B	11	14
Кількість вузлів у C	234	386
Кількість вузлів $c_{роз}$ у C	8	10

Джерело: розроблено авторами

Подяка. Дана стаття підготована завдяки грантової підтримки Національного Фонду Досліджень України, реєстраційний номер проєкту 33/0012 від 3/03/2025 (2023.04/0012) «Розроблення інформаційної системи автоматичного виявлення джерел дезінформації та неавтентичної поведінки користувачів чатів» за конкурсом «Наука для зміцнення обороноздатності України».

Висновки. У результаті дослідження було розроблено метод виявлення розповсюджувачів дезінформації у соціальних мережах з урахуванням внутрішньої структури спільнот. Соціальну мережу подано у вигляді зв'язного орієнтованого графа, де кожен вузол являє собою користувача. Користувачі пов'язані між собою ребрами та об'єднані у різні спільноти. Досліджено, що граничні вузли $b_{роз}$ мають вирішальне значення для поширення інформації між спільнотами, а основні вузли – для розповсюдження всередині спільноти. Аналіз реального датасету підтвердив, що фейкова інформація здебільшого поширюється акаунтами з великою часткою фейкових

постів, що свідчить про ймовірність їх фейковості. Результати дослідження можуть бути використані для підвищення ефективності систем автоматичного виявлення дезінформації і забезпечення інформаційної безпеки.

В подальших дослідженнях планується здійснити обрахунки пари оцінок «довіра-надійність» для кожного вузла мережі, використовуючи алгоритм довіри до соціальних медіа (TSM).

Список літератури

1. Almaatouq A., Shmueli E., Nouh M., Alabdulkareem A., Singh V.K., Alsaleh M., Alarifi A., Alfaris A., Pentland A. If it looks like a spammer and behaves like a spammer, it must be a spammer: analysis and detection of microblogging spam accounts. International Journal of Information Security. 2016. Vol. 15, № 5. P. 475-491. URL: https://doi.org/10.1007/s10207-016-0321-5 (дата звернення: 10.03.2025).
2. Castillo C., Mendoza M., Poblete B. Information credibility on twitter. 20th international conference on World wide web, 2011. P. 675–684. URL: https://doi.org/10.1145/1963405.1963500 (дата звернення: 10.03.2025).
3. Pennycook G., Rand D.G. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Journal of Personality. 2020. Vol. 88(2). P.185–200. URL: https://doi.org/10.1111/jopy.12476 (дата звернення: 10.03.2025).
4. Karami M., Nazer T.H., Liu H. Profiling fake news spreaders on social media through psychological and motivational factors. 32nd ACM conference on hypertext and social media, 2021. P. 225–230. URL: https://doi.org/10.1145/3465336.347509 (дата звернення: 10.03.2025).
5. Shu K., Wang S., Liu H. Understanding user profiles on social media for fake news detection. 2018 IEEE conference on multimedia information processing and retrieval (MIPR), 2018. P. 430–435. DOI: 10.1109/MIPR.2018.00092 (дата звернення: 10.03.2025).
6. Shu K., Wang S., Liu H. Beyond news contents: the role of social context for fake news detection. Twelfth ACM international conference on web search and data mining, 2019. P. 312–320 URL: https://doi.org/10.1145/3289600.3290994 (дата звернення: 10.03.2025).
7. Rangel F., Giachanou A., Ghanem B.H.H., Rosso P. Overview of the 8th author profiling task at pan 2020: profiling fake news spreaders on Twitter. 11th International Conference of the CLEF Association, 2020. Vol. 2696. DOI: 10.1007/978-3-030-58219-7_25 (дата звернення: 10.03.2025).
8. Cardaioli M., Cecconello S., Conti M., Pajola L., Turrin F. Fake news spreaders profiling through behavioural analysis. Working Notes of Conference and Labs of the Evaluation Forum, 2020. URL: https://ceur-ws.org/Vol-2696/paper_113.pdf (дата звернення: 10.03.2025).
9. Pizarro J. Using n-grams to detect fake news spreaders on Twitter. Working Notes of Conference and Labs of the Evaluation Forum, 2020. URL: https://ceur-ws.org/Vol-2696/paper_181.pdf (дата звернення: 10.03.2025).
10. Vogel I., Meghana M. Fake news spreader detection on Twitter using character n-grams. Working Notes of Conference and Labs of the Evaluation Forum, 2020. URL: https://ceur-ws.org/Vol-2696/paper_59.pdf (дата звернення: 10.03.2025).
11. Giachanou A., Ríssola E.A., Ghanem B., Crestani F., Rosso P. The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. International conference on applications of natural language to information systems. Springer. 2020. P. 181–192. DOI: 10.1007/978-3-030-51310-8_17. (дата звернення: 10.03.2025).
12. Chen T., Li X., Yin H., Zhang J. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. Lecture Notes in Computer Science. 2018. Vol. 11154. P. 40-52. URL: https://doi.org/10.48550/arXiv.1704.05973 (дата звернення: 05.04.2025).
13. Bian T., Xiao X., Xu T., Zhao P., Huang W., Rong Y., Huang J. Rumor detection on social media with bi-directional graph convolutional networks. AAAI 2020. 2020. URL: https://doi.org/10.48550/arXiv.2001.06362 (дата звернення: 10.03.2025).
14. Lu Y., Li C. GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media. 58th Annual Meeting of the Association for Computational Linguistics, 2020. P. 505–514. URL:

https://doi.org/10.48550/arXiv.2004.11648 (дата звернення: 05.04.2025).

15. Nguyen V., Sugiyama K., Nakov P., Kan M. FANG: Leveraging Social Context for Fake News Detection Using Graph Representation. 29th ACM International Conference on Information and Knowledge Management, 2020. P. 1165 – 1174. URL: https://doi.org/10.1145/3340531.3412046 (дата звернення: 15.03.2025).
16. Vosoughi S., Roy D., Aral S. The spread of true and false news online. Science. 2018. Vol. 359, No. 6380. P. 1146–1151. URL: https://www.science.org/doi/10.1126/science.aap9559(дата звернення: 15.03.2025).
17. Kim J., Kim D., Oh A. Homogeneity-based transmissive process to model true and false news in social networks. 12th ACM International Conference on Web Search and Data Mining, 2019. URL: https://doi.org/10.48550/arXiv.1811.09702 (дата звернення: 20.03.2025).
18. Rath B., Gao W., Srivastava J. Evaluating vulnerability to fake news in social networks: a community health assessment model. 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2019. DOI: 10.1145/3341161.3342920 (дата звернення: 20.03.2025).
19. Лозинська О., Марків О., Висоцька В., Романчук Р., Назаркевич М. Інформаційна технологія розроблення та наповнення датасету дезінформації з використанням інтелектуального пошуку дипфейків та клікбейтів. *Herald of Khmelnytskyi National University. Technical Sciences*. 2024. № 343, т. 6(1). С. 158-167. URL: https://doi.org/10.31891/2307-5732-2024-343-6-24 (дата звернення: 20.03.2025).

References

1. Almaatouq, A., Shmueli, E., Nouh, M., Alabdulkareem, A., Singh, V.K., Alsaleh, M., Alarifi, A., Alfaris, A., & Pentland, A. (2016). If it looks like a spammer and behaves like a spammer, it must be a spammer: analysis and detection of microblogging spam accounts. International Journal of Information Security. 15 (5), 475-491. https://doi.org/10.1007/s10207-016-0321-5.
2. Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on Twitter. 20th international conference on World wide web (675–684). https://doi.org/10.1145/1963405.1963500.
3. Pennycook, G., & Rand, D.G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Journal of Personality. 88(2), 185–200. https://doi.org/10.1111/jopy.12476.
4. Karami, M., Nazer, T.H., & Liu, H. (2021). Profiling fake news spreaders on social media through psychological and motivational factors. 32nd ACM conference on hypertext and social media (225–230). https://doi.org/10.1145/3465336.347509.
5. Shu, K., Wang, S., & Liu, H. (2018). Understanding user profiles on social media for fake news detection. 2018 IEEE conference on multimedia information processing and retrieval (430–435). DOI: 10.1109/MIPR.2018.00092.
6. Shu, K., Wang, S., & Liu, H. (2019). Beyond news contents: the role of social context for fake news detection. Twelfth ACM international conference on web search and data mining (312–320). URL: https://doi.org/10.1145/3289600.3290994.
7. Rangel, F., Giachanou, A., Ghanem, B.H.H., & Rosso, P. (2020). Overview of the 8th author profiling task at pan 2020: profiling fake news spreaders on Twitter. 11th International Conference of the CLEF Association. DOI: 10.1007/978-3-030-58219-7_25.
8. Cardaioli, M., Cecconello, S., Conti, M., Pajola L., & Turrin F. (2020). Fake news spreaders profiling through behavioural analysis. Working Notes of Conference and Labs of the Evaluation Forum. https://ceur-ws.org/Vol-2696/paper_113.pdf .
9. Pizarro, J. (2020). Using n-grams to detect fake news spreaders on Twitter. Working Notes of Conference and Labs of the Evaluation Forum. https://ceur-ws.org/Vol-2696/paper_181.pdf.
10. Vogel, I., & Meghana, M. (2020). Fake news spreader detection on Twitter using character n-grams. Working Notes of Conference and Labs of the Evaluation Forum. https://ceur-ws.org/Vol-2696/paper_59.pdf.
11. Giachanou, A., Ríssola, E.A., Ghanem, B., Crestani, F., & Rosso, P. (2020). The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. International conference on applications of natural language to information systems (181–192). DOI: 10.1007/978-3-030-51310-8_17.
12. Chen, T., Li, X., Yin, H., & Zhang, J. (2018). Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. Lecture Notes in Computer Science. 11154, 40-52. URL: https://doi.org/10.48550/arXiv.1704.05973.

13. Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., & Huang, J. (2020). Rumor detection on social media with bi-directional graph convolutional networks. AAAI 2020. https://doi.org/10.48550/arXiv.2001.06362.
14. Lu, Y., & Li, C. (2020). GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media. 58th Annual Meeting of the Association for Computational Linguistics (505–514). https://doi.org/10.48550/arXiv.2004.11648 .
15. Nguyen, V., Sugiyama, K., Nakov, P., & Kan, M. (2020). FANG: Leveraging Social Context for Fake News Detection Using Graph Representation. 29th ACM International Conference on Information and Knowledge Management (1165 – 1174). https://doi.org/10.1145/3340531.3412046.
16. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science. 359 (6380), 1146–1151. https://www.science.org/doi/10.1126/science.aap9559.
17. Kim, J., Kim, D., & Oh, A. (2019). Homogeneity-based transmissive process to model true and false news in social networks. 12th ACM International Conference on Web Search and Data Mining. https://doi.org/10.48550/arXiv.1811.09702.
18. Rath, B., Gao, W., & Srivastava, J. (2019). Evaluating vulnerability to fake news in social networks: a community health assessment model. 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). DOI: 10.1145/3341161.3342920.
19. Lozynska, O., Markiv, O., Vysotska, V., Romanchuk, R., & Nazarkevych, M. (2024). Information technology for developing and populating a disinformation dataset using intelligent deepfakes and clickbait search. *Herald of Khmelnytskyi National University. Technical Sciences*, 343(6(1)), 158-167. DOI: https://doi.org/10.31891/2307-5732-2024-343-6-24 [in Ukrainian].

Olga Lozynska, Oksana Markiv, Assoc. Prof., PhD tech. sci., **Victoria Vysotska**, Assoc. Prof., Doct. tech. sci.
Lviv Polytechnic National University, Lviv, Ukraine

A Method for Detecting Disinformation Spreaders Based on a Graph Representation of the Social Network Structure

The paper considers the problem of disinformation dissemination and detection in social networks. Based on the analysis, it was found that the problem of disinformation detection is closely related, first of all, to the problem of detecting disinformation distributors. A method for detecting disinformation spreaders is proposed based on the representation of a social network in the form of a directed graph, with the division of users into communities.

In the study, the social network is presented as a connected directed graph, where each node represents a user. Users are connected to each other by edges and are united into different communities. The level of trust within a community is much higher than between users of different communities. Thus, if misinformation spreads within such a community, the probability of infection of all members of the community will be high. We introduce a classification of nodes (neighboring, border, and main) and show the role they play in the dissemination of information. An analysis of a Ukrainian-language news dataset, including over 2000 records, is conducted to identify users who can potentially become disseminators of information (both true and false). The number of items with the label “true” is 1283, and with the label “fake” is 891. We need to predict the boundary nodes, which are most likely to become information spreaders in the user community, since they are most tangential to all other nodes in the community. In the same way, calculations need to be made to predict the main nodes, which are most likely to become information distributors within the community itself. Thus, if disinformation reaches the boundary nodes (there are 11 of them in our dataset), it will spread throughout the community. In addition, it was found that users who have fake posts on the page are most likely to spread fake information. This may indicate that this account is fake.

This study introduces a method for identifying misinformation spreaders in online social networks. The findings highlight that boundary nodes serve as conduits for inter-community dissemination, while core nodes play a significant role within their respective communities. The result of research using a real-world Ukrainian news dataset demonstrates that users who predominantly share fake posts are more likely to spread disinformation, potentially indicating the use of fake accounts. These insights can inform the design of advanced detection algorithms and support efforts to safeguard the information integrity of digital communication platforms. Further studies are planned to calculate a pair of “trust-reliability” scores for each network node using the Trust in Social Media (TSM) algorithm.

disinformation, fake news, dataset, disinformation spreaders, graphs

Одержано (Received) 23.04.2025 *Прорецензовано (Reviewed) 30.04.2025*
Прийнято до друку (Approved) 06.05.2025