

УДК 004.65:519.17]:316.472.4

DOI: 10.31673/2412-9070.2024.051035

**О. М. ШУШУРА**, доктор техн. наук, професор;

ORCID: 0000-0003-3200-720X

**Б. С. КОКІДЬКО**, аспірант,

ORCID: 0009-0003-2669-3458

Національний технічний університет України «Київський політехнічний університет  
імені Ігоря Сікорського»

## МОДЕЛЮВАННЯ СОЦІАЛЬНИХ МЕРЕЖ З ВИКОРИСТАННЯМ ТЕОРІЇ ГРАФІВ ТА НЕЧІТКОЇ ЛОГІКИ

*Аналіз соціальних мереж є потужним інструментом для розуміння складних взаємодій у сучасному світі. Його застосування знаходить місце у різних галузях, від маркетингу до кримінології, що робить його незамінним для досліджень та практичного використання.*

*У роботі запропоновано підхід до аналізу та моделювання соціальних мереж на основі використання графових баз даних та нечіткої логіки. Сучасні дослідження підтверджують ефективність поєднання графових баз даних та нечіткої логіки для аналізу та моделювання соціальних мереж, що дозволяє враховувати складні взаємозв'язки між об'єктами та покращити точність прогнозів. Однак необхідність розвитку методів та моделей аналізу соціальних мереж, врахування нестаціонарності та невизначеності їх характеристик обумовлюють актуальність подальших досліджень у цьому напрямку.*

*Метою роботи є формалізація соціальних мереж з використанням теорії графів та нечіткої логіки для побудови інформаційної технології аналізу соціальних мереж на основі графових баз даних. Розглянуто аспекти використання графових баз даних та нечіткої логіки для зберігання та аналізу даних соціальних мереж, виділено та формалізовано основні елементи соціальних мереж та їх характеристики. Розроблена модель включає визначення акаунтів соціальних мереж як вершин зваженого графу та їх основні характеристики, опис зв'язків між акаунтами соціальних мереж у вигляді звичайних та нечітких бінарних часово-залежних відношень. Розглянуто відношення підписки, впливу та схожості акаунтів соціальних мереж.*

*Запропонована модель дозволяє аналізувати вплив окремих акаунтів та моделювати поширення інформації, визначати типи користувачів для виявлення аномальної поведінки, проводити аналіз динаміки змін у соціальних мережах та прогнозувати майбутні тренди. Результати роботи можуть бути використані у маркетингу, управлінні кризовими ситуаціями та безпеки, політичних кампаніях, соціальних науках та інших галузях, роблячи аналіз соціальних мереж більш точним та ефективним інструментом для різноманітних завдань.*

**Ключові слова:** соціальні мережі, графова база даних, нечітка логіка, теорія графів, аналіз, моделювання, інформаційна технологія, інформаційна система, аналіз даних.

### Постановка задачі

Аналіз та моделювання соціальних мереж дозволяє виявляти складні взаємозв'язки між їх учасниками та використовувати цю інформацію у задачах маркетингу, соціології, політології та ін. Представлення соціальної мережі у вигляді моделі теорії графів є досить природнім, а нечітка логіка забезпечує більш гнучкий опис зв'язків у мережі. Нечітка логіка надає можливість ефективно моделювати та аналізувати неоднозначність та непевність в даних, що є ключовими властивостями соціальних мереж. Для представлення інформації про соціальні мережі в інформаційних системах можуть використовуватися графові бази даних, які відображають взаємозв'язки між об'єктами у вигляді графа, де вузлами є об'єкти, а зв'язки між ними відобра-

жають взаємодії. Цей підхід дозволяє моделювати складні структури та аналізувати взаємозв'язки великої кількості об'єктів. Поєднання нечіткої логіки з графовими базами даних дозволяє отримати нові можливості для аналізу соціальних мереж, враховуючи нечіткість та складність відносин між учасниками. Такий підхід допомагає краще зрозуміти поведінку користувачів, виявляти приховані патерни та робити більш обґрунтовані висновки. Однак, розробка ефективних методів інтеграції нечіткої логіки та графових баз даних потребує вирішення низки задач. Серед них – формалізація характеристик соціальних мереж з використанням теорії графів та нечіткої логіки, розробка відповідних структур даних та алгоритмів, здатних працювати з великою кількістю взаємозв'язаних об'єктів, а також розробка правил нечіткого виводу, адаптованих для графових структур. Крім того, необхідно враховувати масштабованість рішень, щоб забезпечити їх ефективне застосування в умовах великих обсягів даних соціальних мереж.

### *Аналіз останніх досліджень та публікацій*

Питанням аналізу та моделювання соціальних мереж за допомогою графових баз даних і нечіткої логіки присвячені роботи таких авторів, як M. Bastian, S. Heymann, та M. Jacomy [1], які розробили Gephi для візуалізації графів. Роботи E. Csardi та T. Nepusz [2] зосереджені на застосуванні бібліотеки igraph для аналізу великих графів. Основні методи нечіткої логіки, розроблені Л. Заде та Е. Мамдані [3,4], використовуються для моделювання невизначеності та неоднозначності у даних соціальних мереж. Вони дозволяють оцінювати ступінь належності об'єктів до певних категорій на основі нечітких множин та правил інтерпретації. Існуючі методи аналізу соціальних мереж здебільшого використовують графові бази даних для зберігання та обробки даних про взаємодії між користувачами. Це дозволяє ефективно моделювати складні структури та виявляти патерни поведінки [5]. З іншого боку, інтеграція нечіткої логіки дозволяє більш гнучко описувати складні явища та враховувати невизначеність у взаємодіях [6]. Запропоновані методи інтеграції графових баз даних та нечіткої логіки потребують розробки відповідних структур даних для реалізації бази знань та процедур нечіткого виводу [7]. Робота [8] представляє узагальнену структуру бази знань для цього підходу, але потребує конкретизації для практичного застосування. В останні роки набула розвитку концепція Social Networks Mining (видобуток соціальних мереж), як засіб вирішення проблематики отримання інформації із соціальних мереж. Дана концепція представляється як системна інтеграція комп'ютерного моделювання багатоагентних систем, що включає безліч комп'ютаційних стандартів, методологій, мов моделювання, алгоритмів моделювання та імітаційних моделей платформ та комп'ютерних систем.

Таким чином, сучасні дослідження підтверджують ефективність поєднання графових баз даних та нечіткої логіки для аналізу та моделювання соціальних мереж, що дозволяє враховувати складні взаємозв'язки між об'єктами та покращувати точність прогнозів. Однак, необхідність розвитку методів та моделей аналізу соціальних мереж, врахування нестационарності їх характеристик обумовлюють актуальність подальших досліджень у цьому напрямку.

### *Основна частина*

Метою даної роботи є формалізація соціальних мереж з використанням теорії графів та нечіткої логіки в широкому розумінні для побудови інформаційної технології аналізу соціальних мереж на основі графових баз даних. Для досягнення поставленої мети розглядаються аспекти використання графових баз даних та нечіткої логіки для зберігання та аналізу даних соціальних мереж, виділяються та формалізуються основні елементи та характеристики соціальних мереж.

Використання графових баз даних дозволяє ефективно зберігати та обробляти складні взаємозв'язки між об'єктами, тоді як нечітка логіка забезпечує гнучкість і можливість роботи з неточними або неповними даними, що є характерним для соціальних мереж. Це поєднання технологій сприяє глибшому розумінню динаміки соціальних мереж і прийняттю більш обґрунтованих рішень на основі аналізу даних.

Основи нечіткої логіки було закладено наприкінці 1960-х років у працях вченого Л. Заде. В основі алгебри нечіткої логіки лежать два основні поняття: нечіткої множини та нечітких операцій над ними. При цьому вважається, що логіка мислення при прийнятті рішень людиною відмінна від двійкової та багатозначної логіки, оскільки вона має справу з нечіткими, розмитими класами понять та відносин людської мови.

Інтеграція графових баз даних та нечіткої логіки передбачає кілька ключових етапів. Перш за все, необхідно визначити структуру графа, яка буде використовуватися для моделювання соціальних мереж. Графові бази даних, такі як Neo4j, дозволяють зберігати вузли та ребра, що представляють користувачів та їх взаємодії відповідно. Цей підхід дозволяє легко масштабувати систему та зберігати великі обсяги даних про взаємозв'язки.

Аналіз соціальних мереж часто зосереджений на виявленні спільнот та прихованих структур. Для цього використовуються різні алгоритми кластеризації та виявлення спільнот, наприклад, такі як алгоритм Louvain [8]. Інтеграція нечіткої логіки дозволяє покращити точність цих алгоритмів, враховуючи неоднозначність та невизначеність у взаємодіях між користувачами. Це особливо важливо, оскільки багато зв'язків можуть мати різну інтерпретацію та ступінь важливості для учасників мережі. Крім того, застосування нечіткої логіки дозволяє ефективно класифікувати учасників соціальної мережі та визначити їхню важливість. Це допомагає у створенні більш точних та реалістичних моделей взаємодій у мережі. Застосування нечіткої логіки також дозволяє виявляти впливові особи та групи у соціальній мережі. Це важливо для розуміння структури та динаміки мережі, а також для розробки стратегій управління та планування дій. Додатково, нечітка логіка може бути корисною для аналізу динаміки змін у соціальній мережі та прогнозування майбутніх трендів. Це допомагає відстежувати розвиток мережі та адаптувати стратегії управління відповідно до змін. Отже, можна прийти до висновку, що використання нечіткої логіки має значний вплив на покращення аналізу соціальних мереж. Цей підхід дозволяє ефективно враховувати складність та неоднозначність взаємодій між учасниками мережі, що допомагає отримати більш повне та точне розуміння їхньої структури та динаміки.

Однією з перспектив використання нечіткої логіки є розвиток більш точних та ефективних алгоритмів аналізу соціальних мереж. Врахування неоднозначності у визначенні важливих вузлів та груп у мережі дозволяє отримувати більш об'єктивні результати, що є важливим для прийняття рішень у сфері маркетингу, управління та розвитку бізнесу. Крім того, нечітка логіка дозволяє ефективно моделювати динаміку змін у соціальних мережах та прогнозувати їхні майбутні тренди. Це відкриває нові можливості для розробки стратегій управління та планування дій у сферах соціального медіа, маркетингу та комунікацій. Також використання нечіткої логіки на графових базах даних дозволяє ефективно враховувати різноманітність типів зв'язків у соціальних мережах, таких як дружба, співпраця, вплив тощо. Це допомагає покращити якість аналізу та розуміння взаємодій у мережі. Загалом, перспективи використання нечіткої логіки на графових базах даних для аналізу соціальних мереж великі та різноманітні. Цей підхід дозволяє ефективно враховувати неоднозначність та непевність у взаємодіях учасників мережі, що відкриває нові можливості для покращення аналізу та управління соціальними мережами в цілому.

Визначимо множину соціальних мереж  $s_i \in S$ , де  $i$  – номер соціальної мережі. У загальному випадку соціальну мережу можна розглядати як зважений граф, вершинами якого є акаунти, а ребрами – зв'язки між ними. Для формалізації акаунтів у соціальних мережах введемо множину  $A_i$  акаунтів соціальної мережі  $s_i$ . У соціальній мережі можна виділити 4 основні групи акаунтів: оповісники, з великою кількістю послідовників (наприклад: акаунти знаменитостей); знайомі із певною взаємністю у стосунках (наприклад: звичайні користувачі); боти, сторінкою якого керує автоматизована програма для інтеракції з користувачами; послідовники, які стежать за набагато більшою кількістю користувачів, ніж за ними (наприклад: спамери, сталкери). Акаунт представляється як елемент  $a_i^j \in A_i$ , де  $j$  – номер акаунту у соціальній мережі  $s_i$ . У соціальних мережах одному власнику, як фізичній особі чи організації, може належати

кілька акаунтів соціальних мереж. У задачах аналізу соціальних мереж важливо формувати інформацію щодо саме власників акаунтів.

Множину  $V$  власників акаунтів у соціальних мережах можна представити як:

$$V = \bigcup V_i, \quad (1)$$

де  $V_i$  – множина власників акаунтів соціальної мережі  $s_i$ .

Власниками акаунтів соціальних мереж можуть бути фізичні особи або організації.

Групи або спільноти, які формуються за вподобаннями, за місцезнаходженням, за релігією тощо, є невід'ємною складовою соціальних мереж. Позначимо  $G$  множину груп у соціальній мережі  $s_i$ .

Акаунт  $a_i^j \in A_i$  в соціальній мережі має множину характеристик  $X_i$ , які у загальному випадку є функціями часу  $t$ . До характеристик, які використовуються для аналізу акаунту  $a_i^j \in A_i$  в соціальній мережі, можна віднести (в загальному випадку не обмежуючись цим переліком):

- вік власника у роках  $y_i^j(t)$ ;
- стаття  $sx_i^j(t) \in SX$ , де  $SX$  – множина можливих статей;
- географічне розташування  $g_i^j(t) = (lat_i^j(t), lon_i^j(t))$ , яке представляється як набір GPS координат широти  $lat_i^j(t)$  та довготи  $lon_i^j(t)$ ;
- професія  $pr_i^j(t) \in PR$ , де  $PR$  – множина усіх професій;
- маркетингові вподобання  $\overline{mr}_i^j(t)$ , де кожен елемент цього вектору визначає рівень зацікавленості акаунту у відповідному маркетинговому продукту, представлений числом з відрізка  $[0,1]$ ;
- політичні погляди  $p_i^j(t) = (lr_i^j(t), dt_i^j(t))$ , що є двоосьовим політичним спектром, де  $lr_i^j \in [0,1]$  означає лівий-правий вимір та  $dt_i^j \in [0,1]$  означає вимір демократії-тоталітаризму в певний проміжок часу  $t$ ;
- множини  $WP_i^j(t)$ ,  $RP_i^j(t)$ ,  $LP_i^j(t)$  відповідно написаних, прочитаних та схвалених («лайкнутих») постів акаунту в соціальній мережі на момент часу  $t$ ;
- оцінки авторитетності акаунту  $C_i^j(t) \in [0,1]$  та важливості його як хабу  $H_i^j(t) \in [0,1]$ , які визначають його важливість та впливовість в мережі на момент часу  $t$ ;
- добова активність  $da_i^j(t) \in [0,24]$ , яка показує час активності акаунту в соціальній мережі за певну добу.

В залежності від постановки задачі аналізу соціальної мережі та типу зв'язків, що досліджується, її граф може бути зваженим, а також орієнтованим чи неорієнтованим. Кожному типу зв'язків може бути співставлено звичайне або нечітке бінарне відношення на множині акаунтів  $A_i$ . Розглянемо деякі приклади зазначених відношень для соціальних мереж, до яких можна віднести підписки, вплив одного акаунта на інший та їх схожість.

Одним з найбільш розповсюджених типів зв'язків між акаунтами в соціальній мережі є підписка. Позначимо  $RS(a_i^j, a_i^k, t)$  часово-залежне бінарне відношення, яке визначає, що акаунт  $a_i^j$  є підписником акаунта  $a_i^k$  в соціальній мережі  $s_i$ . Це відношення дозволяє оцінити кількість підписників акаунту та відповідного його популярність. Однак не завжди наявність підписки означає, що підписант ознайомлюється з інформацією акаунта, й так підписка може бути просто формальною. Тому доцільно розглянути нечітке часово-залежне бінарне відношення  $RFS(a_i^j, a_i^k, t)$ , яке відображає ступінь функціонування підписки. Функція належності такого відношення має вид:

$$\mu_{RFS}(a_i^j, a_i^k, t) = \frac{|WP_i^k(t) \cap RP_i^j(t)|}{|WP_i^k(t)|} \quad (2)$$

Відповідно до формули (2), функція належності  $\mu_{RFS}(a_i^j, a_i^k, t)$  прийме значення 1 тільки в тому випадку, коли акаунт  $a_i^j$  прочитає всі пости, написані акаунтом  $a_i^k$ .

Вплив одного акаунту на інший в соціальній мережі є одним з найважливіших показників для її аналізу. Враховуючи наявність невизначеності, для формалізації цієї характеристики доцільно використати апарат нечітких відношень. Позначимо  $RFI(a_i^j, a_i^k, t)$  нечітке часово-залежне бінарне відношення, яке визначає ступінь впливу акаунта  $a_i^j$  на акаунт  $a_i^k$  в соціальній мережі  $s_i$ . Вплив акаунта  $a_i^j$  на акаунт  $a_i^k$  визначається багатьма факторами, серед яких можна виділити авторитетність акаунту  $a_i^j$ , метрики залученості акаунту  $a_i^k$  до інформації акаунту  $a_i^j$  (лайки, коментарі, репости тощо), показники тематичного аналізу та сентимент-аналізу постів акаунту  $a_i^k$  у зв'язку з постами акаунту  $a_i^j$ , оцінки швидкості та обсягу поширення контенту від акаунта  $a_i^j$  до акаунта  $a_i^k$ . Для формування значень цих факторів доцільно використати відомі методи, серед яких можна зазначити NITS для визначення авторитетності акаунту, гібридні методи та метод ВТМ (Biterm Topic Model) для сентимент-аналізу та тематичного аналізу відповідно, машинне навчання. Вказані фактори можна представити у вигляді лінгвістичних змінних та реалізувати розрахунок функції належності  $\mu_{RFI}(a_i^j, a_i^k, t)$  відношення  $RFI$  в певний момент часу  $t$  методом Мамдані на основі нечіткої моделі, що містить експертні правила оцінки впливу акаунту.

Оцінка схожості акаунтів є важливою для виявлення акаунтів з однаковим власником як в межах однієї соціальної мережі, так в різних соціальних мережах. Позначимо  $RFA(a_i^j, a_m^k, t)$  нечітке часово-залежне бінарне відношення, яке визначає схожість акаунта  $a_i^j$  на акаунт  $a_m^k$  в соціальних мережах  $s_i$  та  $s_m$  відповідно. Розрахунок функції належності даного відношення з використанням дельта-функції Кронекера та косинусної схожості можна виконувати так:

$$\begin{aligned} \mu_{RFA}(a_i^j, a_m^k, t) = & w_1 \delta_{y_i^j(t), y_m^k(t)} + w_2 \delta_{sx_i^j(t), sx_m^k(t)} + w_3 \delta_{pr_i^j(t), pr_m^k(t)} + \\ & + w_4 \frac{\overline{mr}_i^j(t) \cdot \overline{mr}_m^k(t)}{\|\overline{mr}_i^j(t)\| \cdot \|\overline{mr}_m^k(t)\|} + w_5 \frac{g_i^j(t) \cdot g_m^k(t)}{\|g_i^j(t)\| \cdot \|g_m^k(t)\|} + w_6 \frac{p_i^j(t) \cdot p_m^k(t)}{\|p_i^j(t)\| \cdot \|p_m^k(t)\|} + \\ & + w_7 \frac{|RS^{a_i^j}(t) \cap RS^{a_m^k}(t)|}{|RS^{a_i^j}(t) \cup RS^{a_m^k}(t)|}, \end{aligned} \quad (3)$$

де  $w_1, w_2, w_3, w_4, w_5, w_6, w_7$  – вагові коефіцієнти, що визначають важливість кожного доданку.

### Висновки

У роботі запропоновано підхід до аналізу та моделювання соціальних мереж, що базується на використанні графових баз даних та нечіткої логіки. Виконано формалізацію соціальних мереж з використанням теорії графів та нечіткої логіки, яка включає визначення акаунтів соціальних мереж як вершин зваженого графу та виділення їх основних характеристик, необхідних для аналізу, формування підходів до представлення зв'язків між акаунтами соціальних мереж у вигляді звичайних та нечітких бінарних часово-залежних відношень. Детально розглянуто відношення підписки, впливу та схожості акаунтів соціальних мереж. Відмінність результатів роботи від класичних соціологічних концепцій полягає у врахуванні нестационарності інформації та невизначеності за рахунок використання нечіткої логіки.

Запропонована модель дозволяє аналізувати вплив окремих акаунтів та моделювати поширення інформації, визначати типи користувачів для виявлення аномальної поведінки, проводити аналіз динаміки змін у соціальних мережах та прогнозувати майбутні тренди. Результати роботи можуть бути використані у маркетингу, управлінні кризовими ситуаціями та

безпеки, політичних кампаніях, соціальних науках та інших галузях, роблячи аналіз соціальних мереж більш точним та ефективним інструментом для різноманітних завдань.

### Список літератури

1. Bastian, M., Heymann, S., Jacomy, M. *Gephi: An open source software for exploring and manipulating networks*. 2009, URL: <https://dx.doi.org/10.13140/2.1.1341.1520>.
2. Csardi, G., & Nepusz, T. *The igraph software package for complex network research*. 2006, URL: [https://www.researchgate.net/publication/221995787\\_The\\_Igraph\\_Software\\_Package\\_for\\_Complex\\_Network\\_Research](https://www.researchgate.net/publication/221995787_The_Igraph_Software_Package_for_Complex_Network_Research).
3. Zadeh, L. A. *Fuzzy sets // Information and Control*. – 1965. – Vol. 8. – P. 338-353, URL: [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
4. Mamdani, E. H., Assilian, S. *An experiment in linguistic synthesis with a fuzzy logic controller // International Journal of Man-Machine Studies*. – 1975. – Vol. 7. – No. 1. – P. 1-13, URL: [https://doi.org/10.1016/S0020-7373\(75\)80002-2](https://doi.org/10.1016/S0020-7373(75)80002-2).
5. Guidara, A. (2021). *Artificial Intelligence and Fuzzy Logic*. In: *Policy Decision Modeling with Fuzzy Logic. Studies in Fuzziness and Soft Computing*, vol 405. Springer, Cham. [https://doi.org/10.1007/978-3-030-62628-0\\_5](https://doi.org/10.1007/978-3-030-62628-0_5).
6. Wang, L. *Fuzzy Logic and Graph Theory for Social Network Analysis // Proceedings of the IEEE*. – 2018. – Vol. 106. – No. 5. – P. 789-802, DOI: 10.1109/ICC34414.2018
7. Peixoto, T. P. *Graph-tool: efficient network analysis with python*. 2014, URL: <https://graph-tool.skewed.de>.
8. Staudt, C. L., et al. *NetworKit: A toolkit for large-scale complex network analysis*, *Network Science*, vol. 4, no. 4, pp. 508–530, 2016. DOI: 10.1017/nws.2016.20.

O. Shushura, B. Kokidko

### MODELING OF SOCIAL NETWORKS USING GRAPH THEORY AND FUZZY LOGIC

*Social network analysis is a powerful tool for understanding complex interactions in today's world. Its application finds a place in various fields, from marketing to criminology, making it indispensable for research and practical use.*

*The paper proposes an approach to the analysis and modeling of social networks based on the use of graph databases and fuzzy logic. Modern research confirms the effectiveness of the combination of graph databases and fuzzy logic for the analysis and modeling of social networks, which allows taking into account the complex relationships between objects and improving the accuracy of forecasts. However, the need to develop methods and models for the analysis of social networks, taking into account the non-stationarity and uncertainty of their characteristics determine the relevance of further research in this direction.*

*The aim of the paper is to formalize social networks using graph theory and fuzzy logic to build information technology for analyzing social networks based on graph databases. Aspects of the use of graph databases and fuzzy logic for the storage and analysis of social network data are considered, the main elements of social networks and their characteristics are highlighted and formalized. The developed model includes the definition of social network accounts as vertices of a weighted graph and their main characteristics, the description of connections between social network accounts in the form of regular and fuzzy binary time-dependent relations. Considered the relationship of subscription, influence and similarity of social network accounts.*

*The proposed model allows you to analyze the influence of individual accounts and simulate the spread of information, identify types of users to detect abnormal behavior, analyze the dynamics of changes in social networks and predict future trends. The results of the work can be used in marketing, crisis management and security, political campaigns, social sciences and other fields, making social network analysis a more accurate and effective tool for a variety of tasks.*

**Keywords:** social networks, graph database, fuzzy logic, graph theory, analysis, modeling, information technology, information system, data analysis.