

Т. І. Смірнов<sup>1</sup>, Ю. В. Шабатура<sup>2</sup>

<sup>1</sup> аспірант кафедри КН, Національний лісотехнічний університет України, Львів, Україна, smirnov.t@nltu.lviv.ua

<sup>2</sup> д.т.н., проф кафедри електромеханіки та електроніки, Національна академія сухопутних військ імені гетьмана П. Сагайдачного, Львів, Україна, <https://orcid.org/0000-0002-9961-1244>, shabaturayuriy@gmail.com

### Методи і засоби штучного інтелекту у виявленні мікроструктур в глобальних соціальних мережах

**Анотація.** У сучасних умовах глобальні соціальні мережі відіграють надзвичайно важливу роль, впливаючи як на індивідуальне спілкування, так і на формування професійних взаємодій. Складність цих мереж зумовлює необхідність застосування передових методів аналізу для виявлення взаємозв'язків між користувачами та їх об'єднаннями. У роботі пропонується підхід, який поєднує сучасні алгоритми кластеризації для детального дослідження внутрішніх мікроструктур соціальних мереж. Ми проаналізували методи Spectral Clustering [1] та Label Propagation [2] для оцінки взаємодії користувачів у соціальній мережі X, з акцентом на точність і ефективність виявлення спільнот. Дослідження показує, що ці методи значно підвищують точність ідентифікації груп та сприяють оптимізації процесу аналізу мережевих даних. Зроблено висновки про перспективність інтеграції цих підходів для подальших досліджень структури соціальних мереж.

**Ключові слова:** штучний інтелект, соціальні мережі, кластеризація, Spectral Clustering, Label Propagation, мікроструктури, великі дані

Соціальні мережі стали головним інструментом комунікації у сучасному суспільстві, сприяючи обміну інформацією, формуванню соціальних зв'язків та впливу на суспільну думку. Водночас, складна топологія цих мереж створює значні труднощі для розуміння і моделювання динаміки соціальної взаємодії. Великі обсяги даних, які генеруються користувачами щодня, вимагають використання передових методів аналізу для ефективного виявлення закономірностей та структурних особливостей мережі.

Використання сучасних алгоритмів штучного інтелекту дозволяє вирішувати завдання виявлення кластерів і прихованих структур у мережах з великим обсягом даних. Кластеризація, як один із ключових методів аналізу даних, дозволяє групувати об'єкти з подібними властивостями, що особливо важливо для соціальних мереж, де користувачі можуть утворювати різноманітні спільноти з унікальними характеристиками.

Багато досліджень у цій галузі зосереджуються на аналізі стійкості кластерів та їхньому впливі на інформаційні потоки. Зокрема, вивчення мікроструктур може допомогти зрозуміти механізми формування соціальних груп і впливових осіб (influencers) у мережі. Крім того, аналіз таких структур важливий для виявлення тенденцій у розповсюдженні інформації, що має значення як для бізнесу, так і для суспільних інститутів.

Алгоритми Louvain [3] і Girvan-Newman [4] є добре відомими підходами для виявлення спільнот у великих мережах. У нашому дослідженні акцент зроблено на використанні методів Spectral Clustering [1] і Label Propagation [2], які виявилися більш гнучкими і ефективними для аналізу динамічних і великих соціальних мереж. Ці алгоритми допомагають не лише виявляти сталі спільноти, але й відслідковувати короточасні зміни у структурі мереж, що робить їх важливими інструментами для аналізу сучасних соціальних взаємодій.

**Алгоритми кластеризації.** Spectral Clustering є методом кластеризації, що базується на спектральному аналізі матриці суміжності графа. Цей метод, запропонований Ng та співавторами у 2002 році [1], є ефективним для виявлення спільнот у графах із сильною неоднорідністю структури. Основна ідея полягає в тому, щоб використовувати інформацію про власні вектори та власні значення матриці Лапласа графа для поділу його на спільноти.

Однією з ключових переваг Spectral Clustering є його здатність знаходити приховані залежності між вузлами, які можуть бути неочевидними при використанні інших методів. Наприклад, в соціальних мережах, де зв'язки між користувачами можуть бути як тісними, так і слабкими, цей алгоритм дозволяє ідентифікувати підгрупи з високою внутрішньою когерентністю.

Метод також відрізняється універсальністю, оскільки може застосовуватися до різних типів мереж, включаючи орієнтовані та неорієнтовані графи. Це робить його особливо корисним у контексті аналізу соціальних мереж, де тип зв'язків між користувачами може варіюватися.

Використання Spectral Clustering дозволяє не тільки виявити спільноти, але й оцінити їхню внутрішню когерентність, що є критичним для аналізу великих мереж. Однією з ключових переваг цього методу є його здатність виявляти складні, багаторівневі структури у мережах, які можуть бути не видимі при використанні інших методів.

Алгоритм Label Propagation, розроблений Рагаваном та співавторами у 2007 році [2], є швидким та ефективним підходом до кластеризації, який автоматично визначає кількість кластерів. Його основна ідея полягає в тому, що вузли графа поступово приймають "мітки" від своїх сусідів, поки не буде досягнуто стабільного стану.

Метод не потребує попереднього знання кількості спільнот, що робить його дуже гнучким та зручним для аналізу великих динамічних мереж. Це особливо важливо у випадках, коли структура мережі може змінюватися з часом, а кількість спільнот не є заздалегідь відомою.

Label Propagation є особливо корисним для великих динамічних мереж, де користувачі можуть змінювати свою активність або зв'язки з іншими користувачами. Цей метод дозволяє відстежувати, як змінюються кластери з часом, і дає можливість проводити аналіз в реальному часі. Наприклад, у соціальних мережах, де користувачі можуть швидко утворювати нові групи або виходити зі старих, Label Propagation забезпечує ефективний механізм для постійного оновлення структури мережі.

### **Опис дослідження**

#### **Попередня обробка даних**

Для проведення дослідження використовувалися дані соціальної мережі X, що складаються з 10 мільйонів користувачів та 50 мільйонів зв'язків між ними. Основними метриками для аналізу були активність користувачів, кількість взаємодій, час їх

проведення, типи взаємодій (особисті або групові). Дослідження передбачало аналіз як окремих користувачів, так і великих груп, що формуються у процесі соціальної взаємодії.

Процес попередньої обробки даних включав кілька етапів:

1. Збір даних: Дані були зібрані за допомогою API соціальної мережі X, що дозволило отримати інформацію про користувачів та їх взаємодії.
2. Очистка даних: Було проведено видалення дублікатів, некоректних записів та неактуальних зв'язків. Це включало фільтрацію користувачів з низькою активністю та видалення зв'язків, які не відповідають критеріям дослідження.
3. Преобразування даних: На основі зібраних даних була побудована матриця суміжності, яка відображає взаємодії між користувачами. Для цього використовувалися різні вагові коефіцієнти, що відображають силу зв'язків.
4. Нормалізація: Дані були нормалізовані для забезпечення однакової ваги різних типів взаємодій, що дозволило уникнути упередженості в аналізі.

Формування графа мережі

Після попередньої обробки даних було сформовано граф мережі, де вузлами є користувачі, а ребрами – взаємодії між ними. Матриця суміжності відображає кількість та типи взаємодій між користувачами. Особливу увагу було приділено аналізу орієнтованих та неорієнтованих зв'язків, що дозволило глибше зрозуміти структуру мережі.

Застосування алгоритмів

На першому етапі дослідження ми застосували алгоритм Spectral Clustering для виявлення великих кластерів у мережі [1]. Цей алгоритм дозволив ідентифікувати кілька ключових груп, які характеризуються високим рівнем взаємодії між їх учасниками. Виявлені групи мали сильні внутрішні зв'язки та були відокремлені від інших частин мережі, що свідчить про формування автономних соціальних спільнот.

Наступним етапом було застосування алгоритму Label Propagation для визначення динамічних змін у поведінці користувачів [2]. Цей метод дозволив відстежити, як користувачі групуються на основі локальних зв'язків, а також виявити короткострокові тенденції у формуванні нових груп. Зокрема, було спостережено утворення тимчасових спільнот навколо актуальних подій або тем, що обговорюються в мережі.

### Результати

Алгоритми Spectral Clustering та Label Propagation продемонстрували різні, але взаємодоповнюючі результати у процесі аналізу соціальної мережі X.

Spectral Clustering:

- Визначення ключових кластерів: Spectral Clustering дозволив виявити кілька великих, добре зв'язаних спільнот, які функціонують як автономні структури всередині глобальної мережі [1]. Ці кластери характеризуються стабільними зв'язками між учасниками і меншою схильністю до зовнішнього впливу. Наприклад, було ідентифіковано кластери, що представляють професійні спільноти, хобі-групи та географічні спільноти.
- Внутрішня когерентність: аналіз внутрішньої когерентності виявив, що спільноти, виявлені за допомогою Spectral Clustering, мають високий рівень взаємодій між учасниками, що підтверджує ефективність алгоритму у виявленні міцних соціальних зв'язків.

Label Propagation:

- Виявлення динамічних груп: Label Propagation показав, що багато користувачів є частиною невеликих динамічних груп, які постійно змінюються [2]. Це свідчить

про високу варіативність у структурі взаємодій та потенційно дозволяє відстежувати зміни в реальному часі. Наприклад, було спостережено утворення та розпад тимчасових груп навколо популярних подій, таких як спортивні змагання або культурні фестивалі.

- Аналіз короткострокових тенденцій: динаміка зміни міток дозволила простежити за тим, як групуються користувачі на основі локальних зв'язків, а також виявити короткострокові тенденції у формуванні нових груп. Це корисно для виявлення актуальних тем обговорення та швидкої реакції на зміни у мережі.

Додаткові результати:

- Визначення ключових кластерів: Spectral Clustering дозволив виявити кілька великих, добре зв'язаних спільнот, які функціонують як автономні структури всередині глобальної мережі [1]. Такі кластери зазвичай характеризуються стабільними зв'язками між учасниками і менш схильні до зовнішнього впливу. Наприклад, один з кластерів міг бути сформований навколо певної професійної галузі, де користувачі активно обмінюються спеціалізованою інформацією та досвідом.
- Виявлення динамічних груп: Label Propagation показав, що багато користувачів є частиною невеликих динамічних груп, які постійно змінюються [2]. Це вказує на високу варіативність у структурі взаємодій та потенційно дозволяє відстежувати зміни в реальному часі. Наприклад, нові групи можуть формуватися навколо актуальних тем або подій, що дозволяє соціальній мережі оперативно реагувати на зміни у сприйнятті та інтересах користувачів.

### **Висновки**

Застосування алгоритмів Spectral Clustering та Label Propagation дозволяє отримати нові знання про структуру соціальних мереж. Методи кластеризації є ефективними для виявлення як великих стійких кластерів, так і динамічних мікроструктур, що можуть бути важливими для аналізу поширення інформації та стійкості мережі.

Дослідження показує, що поєднання цих методів дозволяє одночасно аналізувати як довготривалі, так і короткострокові зміни в мережевих структурах. Spectral Clustering забезпечує глибоке розуміння стабільних спільнот, тоді як Label Propagation дозволяє відстежувати динаміку змін та оперативно реагувати на нові тенденції.

У подальших дослідженнях планується використати додаткові алгоритми, такі як Infomap та Walktrap [3], для детальнішого аналізу мережевих структур та порівняння ефективності різних методів. Особлива увага буде приділена інтеграції методів машинного навчання для покращення точності виявлення спільнот та адаптації алгоритмів до змінних умов соціальних мереж [4].

Крім того, планується розширення дослідження на інші соціальні мережі та аналіз мультимедійних даних, що дозволить отримати більш всебічне розуміння глобальних соціальних структур. Використання таких комплексних підходів сприятиме розвитку більш ефективних інструментів для аналізу соціальних мереж, що в свою чергу може мати значний вплив на різні сфери, від маркетингу до управління суспільними процесами.

Список використаної літератури

1. Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14.
2. Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithms to detect community structures in large-scale networks. *Physical Review E*, 76(3), 036106.

3. Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5), 75-174.
4. Gao, J., Feng, W., & Tang, J. (2022). Enhancing spectral clustering for community detection with deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 33(4), 1501-1513.
5. Wang, H., & Zhang, Y. (2024). Integrating spectral clustering and label propagation for robust community detection in heterogeneous networks. *IEEE Transactions on Computational Social Systems*, 11(2), 345-359.