

# Перспективи використання гетерогенних даних для аналізу соціальних мереж

Катерина Молодецька

Кафедра КТiМС  
Поліський національний університет  
Житомир, Україна  
kateryna.molodetska@polissiauniver.edu.ua

Наталія Дурова

Кафедра КТiМС  
Поліський національний університет  
Житомир, Україна  
ndurova@polissiauniver.edu.ua

**Abstract.** *Social networks are a popular tool for social communications, influencing not only the processes of communication and information exchange but also decision-making by users. However, the use of social networks is associated with accumulating a large amount of heterogeneous data. Using such data in social network research allows for obtaining more detailed information about virtual communities' structure and development dynamics. However, the application of such data is associated with problems in their integration, processing, and interpretation. Therefore, developing new methods and approaches to analysing heterogeneous data will create prerequisites for the sustainable development of the information society in the context of globalisation.*

**Ключові слова:** соціальні мережі, гетерогенні, користувачі, дослідження, інтеграція, дані.

## ВСТУП

Внаслідок швидкого розвитку інформаційних технологій та всеосяжної діджиталізації суспільних процесів, соціальні мережі стали одним із провідних інструментів комунікації користувачів у мережі інтернет. Така популярність пояснюється великою кількістю соціальних сервісів, які представлені у мережі та дозволяють не тільки здійснювати комунікацію, але і задовольняють їх інформаційні потреби, функціонують як майданчики для комерції, рекомендують нові контакти та групи, надають інструменти для стрімінгу й створення дописів про своє життя та багато іншого. Здійснення таких видів активностей користувачів пов'язане із

завантаженням до сервісу та подальше накопичення великої кількості різномірних даних про користувачів - від текстової інформації до аудіо- та відео- матеріалів, які прийнято називати гетерогенними [1-3].

Під гетерогенними будемо розуміти комбінацію різних типів даних, що класифікуються за низкою критеріїв та ознак [2]. Гетерогенні дані можуть ефективно використовуватися для аналізу соціальних мереж, зокрема для вирішення задач аналізу віртуальних спільнот з метою встановлення закономірностей їх функціонування. Таким чином, існує об'єктивне протиріччя між значним зростанням обсягів гетерогенних даних у соціальних мережах, їх інтеграція з різних джерел, та відсутністю уніфікованих методів та загальних підходів до їх класифікації, що актуалізує обраний напрям досліджень.

## АНАЛІЗ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Критичний аналіз досліджень і публікацій у галузі оброблення гетерогенних даних [3, 4] показав, що їх використання для аналізу соціальних мереж ускладнюється через наступні причини:

- гетерогенність даних охоплює велику кількість аспектів, таких як різні джерела даних, формати даних, їх семантика, інтерпретація та інше, що робить ускладнює процес дослідження;

- швидкий розвиток технологій і зростання кількості джерел даних створює нові виклики для інтеграції гетерогенних даних, що потребує постійного оновлення методів та технологій, яких наразі недостатньо;

- недостатній розвиток стандартизації даних і відсутність загальноприйнятих методів та підходів до інтеграції гетерогенних даних також ускладнюють проведення досліджень.

Таким чином, на шляху вирішення проблеми дослідження особливостей взаємодії користувачів у соціальних мережах для моделювання і прогнозування суспільних процесів, особливого значення набуває наукове завдання розроблення нових методів та підходів до аналізу гетерогенних даних, що дозволить сформувати передумови для сталого розвитку інформаційного суспільства в умовах глобалізації.

## ГЕТЕРОГЕННІ ДАНІ ТА ЇХ ВИКОРИСТАННЯ

В результаті аналізу різних типів гетерогенних даних запропоновано об'єднати їх у такі групи [1, 3, 4]:

- текстові дані, до яких належать публікації, коментарі, повідомлення, відгуки, хештеги тощо. Інструментами для аналізу таких даних є текстова аналітика і оброблення природної мови;

- графові дані, що несуть інформацію про особливості взаємодії між користувачами - друзями, підписниками, їх реакції на дописи у формі вподобайок, репостів, тегів користувачів. Такі дані формалізуються як графи з використанням спеціалізованих інструментів;

- дані геолокації, які є доступними в окремих соціальних мережах, наприклад Twitter або Instagram. Вони дозволяють користувачам додавати геотеги до своїх дописів і можуть бути використані для вивчення географічних моделей, а також верифікації інформації;

- мультимедійні дані - включають в себе зображення, відео та аудіо, які користувачі додають до своїх публікацій.

- дані про поведінку користувачів у сервісах містять інформацію про їх активність - час публікацій, частота активності, взаємодія з іншими користувачами тощо;

- метадані, які описують інші дані про активність у соціальних мережах.

Аналіз гетерогенних даних у соціальних мережах відкриває нові можливості для розуміння поведінки користувачів, а також

встановлення фактів поширення деструктивного контенту. Особливе значення має використання таких даних для підвищення ефективності персоналізації контенту. Аналіз гетерогенних даних є дієвим інструментом ідентифікації тенденцій та прогнозування подій, які відбуваються у суспільстві, що є важливим для функціонування бізнесу, реалізації кампаній маркетингу, діяльності громадських організацій та дослідників.

## ВИСНОВКИ

Використання гетерогенних даних для аналізу функціонування віртуальних спільнот відкриває нові можливості для дослідження та розуміння широкого спектру соціальних, економічних, культурних та інших аспектів людської діяльності. Застосування гетерогенних даних дозволяє отримати глибокі інсайти та виявляти закономірності функціонування спільнот користувачів для покращення їх взаємодії, забезпечення безпеки та приватності, розвитку віртуальних спільнот та інформаційного суспільства в цілому.

## ЛІТЕРАТУРА

- [1] Tommasel, Antonela, and Daniela Godoy. "Multi-View Community Detection With Heterogeneous Information From Social Media Data". *Neurocomputing*, vol 289, 2018, pp. 195-219. Elsevier BV, doi:10.1016/j.neucom.2018.02.023. Accessed 10 May 2023.
- [2] Dutt, B. Social media wellbeing: Perceived wellbeing amidst social media use in Norway. *Social Sciences & Humanities Open*, vol.7, 2023.
- [3] Гришук Р. В., Мамарев В. М., & Молодецька-Гринчук К. В. Класифікація профілів інформаційної безпеки акторів в соціальних інтернет-сервісах (на прикладі мікроблогу Twitter). *Інформаційні технології та комп'ютерна інженерія*, 2017, 12–19.
- [4] Seoane Santos, M., Abreu, P. H., Fernández, A., Luengo, J., & Santos, J. "The impact of heterogeneous distance functions on missing data imputation and classification performance." *Engineering Applications of Artificial Intelligence*, May 2022.