ПРОБЛЕМИ ДОСТОВІРНОГО ОЦІНЮВАННЯ ІНФОРМАЦІЙНОГО ВПЛИВУ В СОЦІАЛЬНИХ ІНТЕРНЕТ-СЕРВІСАХ

Риков Д.М., Федорченко В.М.
Харківський національний університет радіоелектроніки, Харків, Україна

Соціальні інтернет-сервіси стали середовищем швидкого поширення повідомлень, формування суспільних оцінок і координації взаємодій користувачів. Поряд із легітимним обміном інформацією в них можуть реалізовуватися координовані маніпулятивні кампанії, зокрема з використанням автоматизованих облікових записів. Для завдань кібербезпеки важливим є не лише встановлення факту поширення повідомлення, а й достовірне оцінювання потенційного інформаційного впливу, його динаміки та можливих наслідків.

Метою доповіді є визначення та наукове обґрунтування проблем, що виникають під час побудови моделей оцінки інформаційного впливу в соціальних інтернет-сервісах, а також окреслення напрямів дослідження для підвищення достовірності таких оцінок з урахуванням часових і структурно-поведінкових характеристик мережі.

Класичні моделі поширення впливу формалізують мережеві зв’язки та правила активації вузлів і створюють основу для аналізу потенційного охоплення аудиторії [1, 2]. Однак потенційне охоплення не є тотожним фактичному впливу, перегляд, републікація або реакція користувача самі по собі не доводять зміну його переконань чи поведінки. Тому першою проблемою є визначення змісту показника «вплив» і розмежування спостережуваних цифрових взаємодій, прогнозованої експозиції та реальних поведінкових ефектів.

Друга проблема пов’язана з неоднорідністю користувачів і зв’язків. Вузли соціального графа відрізняються активністю, тематичною спеціалізацією, довірою аудиторії та роллю у поширенні контенту. Модель, що однаково інтерпретує вплив кожного зв’язку, може спотворювати оцінки каскадів. Відомі підходи до навчання ймовірностей поширення за журналами дій користувачів [3], проте перенесення таких параметрів між платформами, темами та часовими проміжками потребує окремої перевірки.

Третя проблема полягає в динамічності мережевого середовища. Інтенсивність контактів, актуальність повідомлень та активність аудиторії змінюються з часом, інформаційні каскади можуть перериватися, повторно активуватися або конкурувати між собою. Внаслідок цього статичні характеристики графа та агреговані показники взаємодій недостатні для пояснення часової структури інформаційного впливу. Потребують дослідження способи врахування давності зв’язків, часових затримок, повторних контактів і конкуренції наративів без надмірного ускладнення моделі.

Четверта проблема стосується відокремлення природного поширення інформації від координованих маніпулятивних дій. Високі показники активності можуть бути наслідком суспільного інтересу, алгоритмічного рекомендування

або штучного підсилення. Дослідження поширення правдивих і неправдивих повідомлень демонструє відмінності в характеристиках інформаційних каскадів [4], проте не дає підстав автоматично визначати наміри окремих учасників лише за мережевими метриками. Отже, важливо поєднувати структурні, часові та поведінкові ознаки й окремо оцінювати ризик хибних звинувачень.

П'ята проблема - достовірність емпіричної перевірки моделей. Дані соціальних сервісів можуть бути неповними через обмеження доступу, приховані взаємодії, видалення дописів і зміни рекомендаційних алгоритмів. За таких умов необхідно запобігати витоку інформації між навчальною та тестовою вибірками, враховувати зміну розподілів даних і порівнювати результати з простими базовими моделями. Окремим завданням є узгодження метрик якості прогнозування, виявлення координованих кампаній та витрат обчислювальних ресурсів.

Для подальшого дослідження доцільно передбачити: формалізацію багатокомпонентного показника інформаційного впливу; дослідження гетерогенних часових графів і механізмів зміни сили зв'язку; розроблення критеріїв виявлення аномальної координації; побудову відтворюваного протоколу експериментальної оцінки з перевіркою стійкості моделей до неповних і змінних даних. Для вирішення визначених наукових завдань пропонується дослідити підхід до оцінювання інформаційного впливу, за якого топологія соціальної мережі, поведінкові характеристики різних груп користувачів та часова динаміка поширення повідомлень розглядаються як взаємопов'язані чинники. Основна дослідницька складність полягає в обґрунтуванні способу їх спільного врахування, що даватиме змогу розмежовувати спостережувану активність, потенційне охоплення аудиторії та оцінюваний інформаційний вплив без ототожнення цих показників.

Отже, центральною проблемою є побудова достовірної та інтерпретованої в часі оцінки інформаційного впливу за умов неоднорідності учасників, мінливості зв'язків і неповноти даних соціальних інтернет-сервісів. Подальша робота має охоплювати формалізацію відповідної моделі, визначення критеріїв її якості та незалежну експериментальну перевірку шляхом порівняння з відомими підходами.

Список літератури

1. Kempe D., Kleinberg J., Tardos É. Maximizing the spread of influence through a social network. Proceedings of KDD '03. 2003. P. 137–146. DOI: https://doi.org/10.1145/956750.956769.

2. Ram R., Rizoiu M.-A. Empirically measuring online social influence. *EPJ Data Science*. 2024. Vol. 13. Article 53. DOI: https://doi.org/10.1140/epjds/s13688-024-00492-z.

3. Goyal A., Bonchi F., Lakshmanan L. V. S. Learning influence probabilities in social networks. Proceedings of WSDM '10. 2010. P. 241–250. DOI: https://doi.org/10.1145/1718487.1718518.

4. Vosoughi S., Roy D., Aral S. The spread of true and false news online. Science. 2018. Vol. 359, No. 6380. P. 1146–1151. DOI: https://doi.org/10.1126/science.aap9559.