

## МОДЕЛЬ РОЗПОВСЮДЖЕННЯ ІНФОРМАЦІЇ В СОЦІАЛЬНИХ МЕРЕЖАХ

А. М. Грайворонська<sup>1, а</sup>, Д. В. Ланде<sup>2</sup>, Б. О. Березін<sup>2</sup>

<sup>1</sup>Національний технічний університет України «Київський політехнічний інститут»

<sup>2</sup>Інститут проблем реєстрації інформації НАН України

### Анотація

В роботі приведені правила функціонування мультиагентної мережі, визначені статистичні закономірності, пов'язані з «життєвим циклом» агентів – інформаційних повідомлень. Виявлені статистичні закономірності, які відносяться до кількості лайків та репостів, розподіл яких, згідно з результатами моделювання, відповідає розподілу Вейбулла. Результати моделювання перевірені шляхом дослідження реальної мережі мікроблогів Twitter.

*Ключові слова:* інформаційний потік, розповсюдження інформації, мультиагентна модель, соціальна мережа

### Вступ

Інформація є необхідним ресурсом для розвитку сучасного суспільства. Події, що відбуваються в «реальному» світі відображаються через публікації в інформаційний простір, а з іншого боку аналізуючи опубліковані повідомлення, можна встановити події, що стали причиною публікації [1].

Аналізуючи динаміку інформаційного потоку можна встановити тематичні напрямки, які набувають або навпаки втрачають актуальність. Це зумовлює актуальність дослідження інформаційних потоків, виявлення закономірностей в них. Також інформаційні потоки впливають на відносини між різними соціальними групами та структуру суспільства, тому постає задача дослідження способів створення потоків то впливу на їх динаміку [2]. Це призводить до постановки задачі моделювання та встановлення закономірностей процесу розповсюдження інформації.

Більшість інформаційних джерел мають потокову структуру. Прикладами таких ресурсів в мережі Інтернет є стрічки новин, соціальні мережі, блоги, електронні листи. Публікація наукових статей також є класичним прикладом, який почали досліджувати значно раніше за електронні інформаційні джерела. В усіх цих випадках спосіб того, як саме інформація надходить з часом є суттєвою частиною її значення.

Вивчення закономірностей розповсюдження інформації в соціальних мережах дозволяє встановити принципи взаємодії людей. З іншого боку, при відомих закономірностях для реального тематичного інформаційного потоку, є можливість визначити його відповідність загальним принципам. При значній відмінності динаміки певного тематичного інформаційного потоку від стандарту можна ставити питання про неприродність даного потоку, що може бути наслідком дезінформації, нав'язування або заборони розповсюдження певної інформації [3].

При моделюванні поширення інформації у соціальних мережах часто беруть за основу граф відносин між користувачами [4]. Вершинам графу ставлять у відповідність певні параметри. Часто вводять характеристику «впливовості» вершини. В такому випадку моделювання розповсюдження інформації може відповідати побудові випадкового дерева. Основними параметрами при побудові такого дерева є ймовірності того, що користувач поширить існуюче інформаційне повідомлення. Такі моделі суттєво спираються на топологію вихідного графу, що потребує додаткових досліджень соціальних мереж. Модель, запропонована в даній роботі, не потребує інформації про граф зв'язків в соціальній мережі.

### 1. Побудова мультиагентної моделі розповсюдження інформації

Ключовим елементом при поширенні інформації є повідомлення (пост, публікація). При цьому повідомлення може викликати різні типи реакції, такі як позитивні/негативні коментарі або просто натиснення like/dislike, текст повідомлення може бути скопійований, розміщений на іншій сторінці за допомогою операції repost, також в одному повідомленні може бути розміщене посилання на інше повідомлення. Тому в якості агента в моделі будемо розглядати повідомлення, і еволюція агента буде пов'язана з подіями, які з ним відбуваються. В якості основної характеристики агента введемо енергію ( $E$ ), яка відображає актуальність повідомлення і степінь інтересу до нього. Природно, що старіння інформації або негативна реакція зменшує енергію повідомлення, а позитивна реакція або надання посилання на повідомлення збільшує його енергію.

Формалізуємо правила еволюції агента в моделі. Агент з'являється з початковою енергією  $E_0$  і з кожним дискретним відліком часу його енергія зменшується на 1. Будемо розглядати події, які є типовими для соціальних мереж: like, dislike, repost. Ці

<sup>а</sup>nastya\_graiv@ukr.net

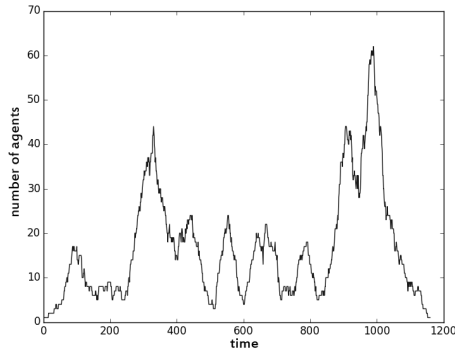


Рис. 1. Часовий ряд, що відповідає зміні кількості агентів, отриманий в результаті моделювання

події впливають на енергію агента наступним чином: like підвищує енергію на 1, dislike зменшує на 1, repost підвищує на 2. З іншого боку, ймовірність того, що якась із цих подій відбудеться, залежить від актуальності повідомлення, інтересу до інформації в ньому, що в термінах моделі виражається енергією. У зв'язку з цим визначимо ймовірності того, що з повідомленням з енергією відбулася певна подія, в такий спосіб:

$$p_{event}^{(E)} = p_{e0}\varphi(E),$$

де  $p_{l0}$ ,  $p_{d0}$ ,  $p_{r0}$  - параметри моделі, а  $\varphi$  - це деяка монотонно неспадна функція від поточної енергії агента зі значеннями в  $[0, 1]$ . При падінні енергії агента до 0, повідомлення «вмирає» і більше не розглядається. Моделювання динаміки всього тематичного інформаційного потоку починається з одного агента. Поява нового агента можлива двома способами. Перший полягає в копіюванні існуючого агента за допомогою операції repost. Також може з'явитися самонароджений агент, що відповідає публікації нового повідомлення. Таким чином, в кожен момент часу з певними ймовірностями, з кожним з агентів, може статися будь-яка з подій, і також з ймовірністю  $p_s$  може з'явитися новий агент.

## 2. Результати чисельного моделювання

Першим результатом чисельного моделювання є часовий ряд, який відповідає зміні кількості агентів в моделі. На (рис. 1, 2) представлений приклад такого часового ряду.

В змодельованій системі агентів також міститься інформація про зв'язки між агентами, а саме який з агентів є копією іншого або між якими агентами є посилення. Це дозволяє розглядати дерева поширення повідомлення, та аналізувати їх характеристики. Також кожен агент містить інформацію про кількість подій, які з ним відбулися, тому можна вивчати пов'язані з цим статистичні закономірності.

## 3. Теоретичні властивості одного агента

Розглянемо життєвий шлях одного агента. Агент з'являється з початковим значенням енергії  $E_0$  і далі його енергія змінюється в залежності від подій, які з ним відбуваються. Будемо вважати, що можливі

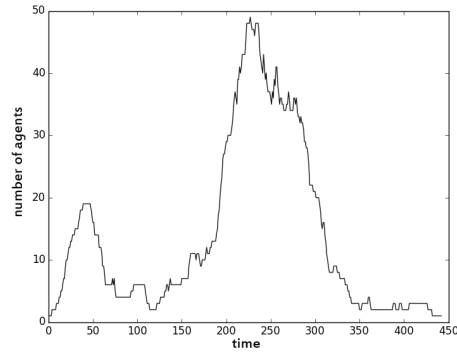


Рис. 2. Часовий ряд, що відповідає зміні кількості агентів, отриманий в результаті моделювання

дві події: like і repost. За одиницю часу може статися одна з цих подій, обидві одночасно або не відбутися жодного.

Позначимо  $\varepsilon_t$  значення енергії агента в момент часу  $t$ . Тоді значення енергії в наступний момент часу можна записати в такий спосіб  $\varepsilon_{t+1} = \varepsilon_t + \delta_t$ , де  $\delta_t$  є випадковою величиною зі значеннями в  $\{-1, 0, 1, 2\}$ . Можна вказати ймовірнісний розподіл  $\delta_t$  при відомій енергії  $\varepsilon_t$  позначення  $P_{\Delta}^{(E)} = P(\delta = \Delta | \varepsilon = E)$ .

Процес зміни енергії агента можна розглядати як цілочисельне випадкове блукання з перехідними ймовірностями

$$p_{ij} = \begin{cases} P_{j-i}^{(i)}, & \text{при } (j-i) \in \{-1, 0, 1, 2\}, i > 0 \\ 1, & \text{при } i = j = 0 \\ 0, & \text{інакше.} \end{cases}$$

Так як значення енергії в наступний момент часу залежить тільки від значення енергії в попередній момент часу, то стохастична послідовність  $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_t, \dots)$  є марковської ланцюгом з перехідними ймовірностями  $p_{ij}$ .

Дослідження властивостей випадкового блукання енергії дозволяє зробити певні висновки про вплив параметрів моделі на результати моделювання. Наприклад, умова  $2p_{r0} + p_{l0} < 1$  є достатньою умовою для того, щоб час життя агента був скінченним з ймовірністю 1.

## 4. Порівняння інформаційних потоків мережі Twitter з результатами моделювання

Для кожного агента можна визначити кількість лайків та репостів, які він отримав. На рис. 3 та рис. 4 показані щільності розподілів цих кількостей, та їх апроксимації розподілом Вейбулла.

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Отримані результати моделювання порівнювались з результатами дослідження життєвого циклу повідомлень новин в мережі мікроблогів Twitter, яке було виконане авторами. Аналізувались характеристики обраних повідомлень [5]. В результаті отримано розподіли лайків і репостів, які відповідають розподілу Вейбулла, як і в моделі (рис. 5, 6).

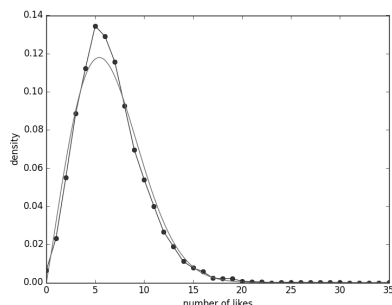


Рис. 3. Щільність розподілу кількості лайків, отриманих агентом.

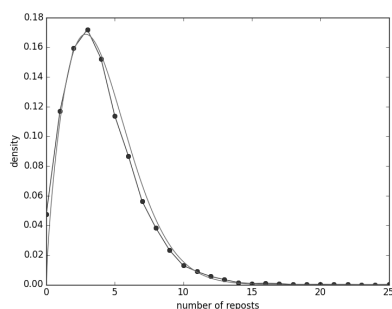


Рис. 4. Щільність розподілу кількості репостів для агента.

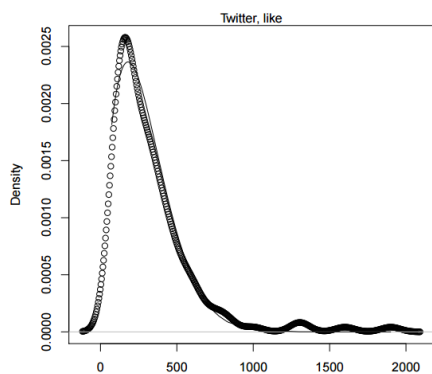


Рис. 5. Щільність розподілу кількості лайків із реальної мережі.

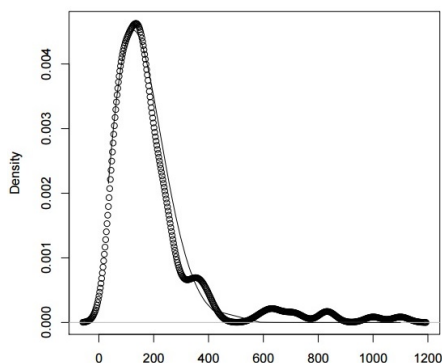


Рис. 6. Щільність розподілу кількості репостів із реальної мережі.

Параметр розподілу Вейбулла  $k$ , що є коефіцієнтом форми, подібний для випадку змодельованих даних та реальних. При моделюванні отримано  $\lambda = 1.7, \lambda = 2.1$  (для лайків/репостів), при дослідженні реальних даних  $\lambda = 1.9, \lambda = 1.6$ .

## Висновки

В роботі розглянута мультиагентна модель, в якій агенти асоційовані з документами і їх життєвий цикл співпадає з життєвим циклом документа в інформаційному просторі. Динаміка тематичного інформаційного потоку, отримується як результат моделювання еволюції системи агентів.

Показана відповідність зміни енергії агента випадковому блуканню з перехідними ймовірностями, які залежать від поточної енергії агента. Встановлена обмеженість часу життя агента при певних параметрах моделі.

Визначено ймовірнісний розподіл кількості лайків та репостів для агентів. Отриманий розподіл з високою точністю апроксимується розподілом Вейбулла, що відповідає статистичним даним соціальної мережі Twitter.

Знайдені закономірності можуть бути використані при визначенні аномалій в життєвому циклі окремих повідомлень, штучно створених інформаційних компаній.

## Перелік використаних джерел

1. Kleinberg J. Temporal dynamics of on-line information streams. — Data Stream Management: Processing High Speed Data Streams — Springer, 2006.
2. Додонов А. Г., Ландэ Д. В., Прищепя В. В., Путятин В. Г. Конкурентная разведка в компьютерных сетях. — К.: ИПРИ НАН Украины, 2013. — 248 с. — с. 34–38.
3. Додонов А. Г., Ландэ Д. В., Додонов В. А. Распознавание информационных операций: мультиагентный подход. — Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2016): материалы VI междунар. науч.-техн. конф. (Минск 18-20 февраля 2016 года). — Минск: БГУИР, 2016. — С. 253-256.
4. Dashun Wang, Zhen Wen, Hanghang Tong Information spreading in context. — Proceedings of the 20th international conference on World wide web. — 2011. — pp. 735-744.
5. Li R., Lei K. H., Khadiwala R., Chang K. C. A Twitter-based Event Detection and Analysis System. — Data Engineering (ICDE), 2012 IEEE 28th International Conference. — P. 1273-1276.