

МОДЕЛІ ВПЛИВУ У СОЦІАЛЬНИХ МЕРЕЖАХ

Д. В. Чугай¹, С. А. Смирнов¹

¹ Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»,
НН Фізико-технічний інститут

Анотація

У цій статті йдеться про математичне представлення поширення впливу у соціальних мережах. Змістовно її можна поділити на 3 частини. У першій буде введено поняття індексів впливу та представлено їхні види. У другій частині статті йдеться про важливі алгоритми дослідження соціальних мереж: лувенський метод та PageRank. Третій розділ буде заключним і розкриє типи соціальних мереж, їхні особливості й математичний підхід для їхнього представлення.

Ключові слова: індекс впливу, модулярність, Лувенський метод, PageRank, dead ends, spider trap

Вступ

Сучасні соціальні мережі – складна структура, що містить у собі учасників та зв'язки між ними. Що більше, за останні два десятиліття вони стали потужним джерелом поширення впливу. Одним з красномовних підтверджень цього факту є кількість людей, що використовують соціальні мережі для пошуку інформації в умовах повномасштабної війни. Так, згідно з нещодавнім соціологічним опитуванням [1], проведеним Київським міжнародним інститутом соціології (КМІС) понад 3/4 українців використовують саме їх як джерело новин. І така тенденція спостерігається у багатьох інших країнах Заходу. Саме тому вивчення поширення впливу у соціальних мережах є досить важливим завданням, яке є і буде актуальним у найближчому майбутньому.

1. Коефіцієнти впливу

Впливом у соціальних мережах вважатимемо процес і результат зміни одним суб'єктом поведінки, уявлень, оцінок іншого. Коефіцієнтами впливу у соціальних мережах є такі числові характеристики груп, що відображають пропорції впливу у таких структурах [2].

1.1 Коефіцієнт Банцафа

$$\beta(i) = \frac{b_i}{\sum_j b_j},$$

де b_i – кількість коаліцій, у яких учасник є ключовим,
 $\sum_j b_j$ – сума кількості коаліцій з усіма ключовими учасниками.

1.2. Коефіцієнт Шеплі-Шубіка [3]

Цей коефіцієнт дуже схожий на індекс Банцафа, втім його відмінність полягає в тому, що тепер у голошуванні беруть участь послідовні коаліції [4] (тобто, тепер нам важливо знати у якому порядку учасники входять до коаліцій). До того ж слід зауважити, що до послідовних коаліцій входять усі учасники голошування (але в різному порядку). Сенса цього коефіцієнта впливу полягає в тому, щоб оцінити наскільки кожен учасник може бути ключовим.

$$\sigma(i) = \frac{b_i}{n!},$$

де b_i – кількість разів того, що учасник є ключовим,
 n – кількість учасників голосування.

1.3. Коефіцієнт Джонстона [5]

Для обчислення коефіцієнту Джонстона нам необхідно визначити:

- 1) Для кожного учасника p_i усі коаліції $C_1, C_2, \dots, C_k$, у яких p_i є ключовим.
- 2) Кількість ключових учасників $n_1, n_2, \dots, n_k$ у $C_1, C_2, \dots, C_k$.

Тоді можемо знайти загальний коефіцієнт Джонстона для учасника за такою формулою:

$$TJI(p_i) = \frac{1}{n_1} + \frac{1}{n_2} + \dots + \frac{1}{n_k}.$$

Коефіцієнт впливу Джонстона:

$$JI(p_i) = \frac{TJI(p_i)}{\sum_j TJI(p_j)}.$$

1.4. Коефіцієнт впливу Дигена-Пакела

Для знаходження коефіцієнту впливу Дигена-Пакела нам потрібно знайти:

- 1) Усі коаліції, що мінімально виграють $C_1, C_2, \dots, C_k$, у яких перебуває учасник p_i .
- 2) Кількість учасників $m_1, m_2, \dots, m_k$ у коаліціях $C_1, C_2, \dots, C_k$.

Тоді можемо знайти загальний коефіцієнт Дигена-Пакела для учасника p_i :

$$TDPI(p_i) = \frac{1}{m_1} + \frac{1}{m_2} + \dots + \frac{1}{m_k},$$

$$DPI(p_i) = \frac{TDPI(p_i)}{\sum_j TDPI(p_j)}.$$

1.5. Коефіцієнт Холлера-Пакела

$$HPI(p_i) = \frac{h_i}{\sum_j h_j},$$

де h_i — кількість коаліцій, що мінімально виграють, і до яких входить i -й учасник.

Недоліки наведених коефіцієнтів

Головним недоліком усіх описаних коефіцієнтів є те, що вони ігнорують відмінність різних учасників у інтенсивності надання переваги щодо створення коаліцій.

Цю хибу можна усунути, якщо ввести певні статистичні характеристики голосування. Одним із прикладів може послугувати введення функції інтенсивності $f(i, S)$ зв'язку учасника з коаліцією S .

$$\chi_i = \sum f(i, S).$$

2. Алгоритми дослідження соціальних мереж

2.1. Лувенський метод [6]

Алгоритм [7]:

- 1) Ініціалізація: маємо мережу, представлену графом з послідовністю вершин. Вважаємо, що кожен вузол належить до групи, що представлена цим же вузлом.
- 2) Рахуємо модулярність для усіх вершин.
- 3) Вершини, що увійшли в одну групу, об'єднуємо в один вузол. Зрештою отримуємо новий граф.
- 4) Повторюємо пункти 2 і 3 доти, поки не досягнемо максимальної модулярності.

Переваги

- Підходить для застосування на великих мережах оскільки забезпечує швидке виконання.
- На виході отримуємо високий ступінь модулярності.
- Забезпечує швидку збіжність.
- Підходить для зважених графів.
- Розпізнає ієрархію груп.

Недоліки

- У результаті роботи цього методу часом залишаються найбільші кластери, а важливі, проте маленькі кластери можуть бути поглинені іншими в процесі роботи алгоритму. Їх можна знайти на проміжних ітераціях [6].
- Для ієрархічних кластерів важко визначити оптимальну кількість груп.

2.2. PageRank

Алгоритм [7]:

- 1) Ініціалізація. Нехай маємо мережу з N сайтами. Прописуємо вектор, заповнюючи його елементи випадковими значеннями ймовірностей, наприклад:

$$\mathbf{r}^0 = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)^T.$$

- 2) Далі ітеративно рахуємо впливовість сайтів за формулою

$$r_j^{t+1} = \sum_{i \rightarrow j} \frac{r_i^t}{d_i}$$

або

$$\mathbf{r}^{t+1} = \mathbf{M} \cdot \mathbf{r}^t.$$

- 3) Задаємо точність ε для зупинки алгоритму. Критерій зупинки:

$$\sum_i |r_i^{t+1} - r_i^t| < \varepsilon.$$

Переваги

- простота використання.

Недоліки [7]

- Якщо сторінка з гарним контентом була створена нещодавно й не встигла набрати велику аудиторію, то PageRank не вкаже її серед впливових сайтів. Цей недолік можна прибрати, якщо постійно оновлювати значення PageRank.
- Сторінка посилається сама на себе (spider trap). Тобто виникають ситуації, коли алгоритм може зациклюватися. Рішення проблеми полягає в тому, що протягом кількох кроків алгоритм заходить на один і той самий сайт, а потім його викидає за межі цього зациклення на будь-який інший сайт.
- Є сторінки, з яких не виходять інші посилання (dead ends). Тобто сума коефіцієнтів зовнішніх посилань у стовпчику з таким сайтом у стохастичній матриці суміжності буде дорівнювати 0, що, власне, порушує її стохастичність.

3. Моделі соціальних мереж

3.1. Модель дифузії інновацій [8]

Моделі «дифузії інновацій» описують поширення нововведень у суспільстві з плином часу. Під терміном «нововведення» ми матимемо на увазі будь-який

новий для агентів продукт, що може набувати матеріальної чи нематеріальної форми (наприклад, думка, ідея, технологія).

3.2. Модель з порогами [9]

Ці моделі застосовують, коли необхідно спрогнозувати ухвалення рішення людиною з огляду на кількість інших осіб, що вже ухвалили певне рішення. Цю кількість ми можемо позначити певним порогом. Тобто якщо кількість осіб що ухвалили певне рішення більша за цей поріг, то й особа, для якої ми досліджуємо цей показник, ухвалить те саме рішення. Відповідно, якщо кількість осіб менша від порогу, то таке рішення певною особою ухвалене не буде.

Переваги

- Враховує фактори «соціальної кореляції»;
- Відображає фактори зовнішнього впливу;
- Показує «лідерів думок»;
- Враховує фактори «соціальної кореляції»;
- Відображає різні ступені агентів схильності до впливу;
- Існує поріг, що визначає чи зміниться рішення агента від впливу інших думок;
- Присутність власних міркувань агентів.

Недоліки

- Не бере до уваги непрямого впливу.

3.3. Стохастичні моделі в соціальних мережах [10]

Ці моделі описують схильність агентів ухвалювати рішення за допомогою ентропії розподілу ймовірності їхніх дій. На графах це означає, що ваги ребер матимуть стохастичний характер. Через це поведінка агентів є непередбачуваною.

Висновки

У цій статті було розкрито тему дослідження поширення впливу між різноманітними об'єднаннями учасників соціальних мереж: від двох осіб до цілої мережі. Методи, що були наведені для математичного опису таких структур, дозволяють піддавати їх

науковому дослідженню навіть у сучасному вигляді мереж. Втім, сучасні мережі невпинно модернізуються, що вимагає одночасного оновлення методів їхнього дослідження. З огляду на це у статті були наведені приклади проблем, які виникають через оновлення таких структур, а також запропоновані шляхи їх усунення за допомогою корекції алгоритмів та впровадження більш гнучких величин, що точніше можуть описати зв'язки між учасниками мережі. Однак, слід зауважити, що хоча ці методи можуть гарно впоратися з моделюванням впливів у мережах, але все ще залишається місце для їх покращення та модернізації.

Перелік використаних джерел

1. Опитування від ОПОРА. — 2022. — Режим доступу: https://oporaua.org/report/polit_ad/24068-mediaspozhyvannia-ukrayintsiv-v-umovakh-povnomasshtabnoyi-viini-opituvannia-opori.
2. Бінарні відношення, графи й колективні рішення / Алескеров Ф. Т., Хабіна Е.Л., Шварц Д.А. и Левит В. Ю. — Москва : Физматлит, 2012.
3. (Youtube) Mathispower4u. Weighted Voting: The Shapley-Shubik Power Index. — 2013.
4. Martin Jeremy L. Sequential Coalitions.
5. Garbig Nathan. The Johnston Power Index. — 2012.
6. Annalyn Ng, Kenneth Soo. Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. № 6. — ПИТЕР, 2019. — С. 84–92.
7. Leskovec Prof. Jure. CS224W: Machine Learning with Graphs | 2021.
8. Д.А. Губанов Д. А. Новіков А.Г. Чхартішвілі. Соціальні мережі: моделі інформаційного впливу, керування та протидії. — Физматлит, 2010. — С. 84–92.
9. Granovetter Mark. Threshold Models of Collective Behavior. — The University of Chicago Press, 1978. — P. 1420–1443.
10. Rezvanian Alireza, Meybodi Mohammad Reza. Stochastic graph as a model for social networks. — Elsevier, 2016. — P. 621–640.