

УДК 004.724.4(045)

Ланде Д.В., д.т.н., Інститут проблем реєстрації інформації НАН України,
Нечаєв О.О., Інститут спеціального зв'язку і захисту інформації НТУУ «КПІ»

РАНЖИРУВАННЯ ВУЗЛІВ МЕРЕЖ СОЦІАЛЬНОГО ХАРАКТЕРУ

Як мережу соціального характеру можна розглядати мережу, вузлами якої є співробітники організації, комерційної структури, учасники злочинного угруповання тощо, яким притаманна внутрішня ієрархія. Як зв'язки можуть розглядатися контакти соціальних акторів шляхом використання електронної пошти, засобів мобільного зв'язку, IP-телефонії, тощо. Тобто мережею соціального характеру будемо розуміти мережу, вузлами якої є соціальні актори (social actors), а зв'язками – контакти між ними, що супроводжуються обміном інформації. На практиці такі мережі найчастіше квазіієрархічні, вони у деякому сенсі близькі до ієрархічних, тобто відрізняється від ієрархічної відносно невеликою кількістю додаткових зв'язків, що порушують ієрархічність [1, 2].

Метою роботи є опис і обґрунтування алгоритму ранжирування вузлів квазіієрархічних мереж соціального характеру. Під ранжируванням такої мережі будемо розуміти упорядкування соціальних акторів за деякою ознакою, що дозволяє визначати лідируючих за цією ознакою акторів.

Таку мережу можливо представити у вигляді графа, в якому позначками вузлів виступають певні маркери соціальних акторів, а ребрами направлені зв'язки між ними. Аналіз невеликих квазіієрархічних мереж соціального характеру для визначення в них лідерів не представляє складності завдяки наочності їх представлення. У той час, коли мова йде про більш складні мережі, де кількість вузлів може становити більше 1000 (великі організації тощо), майже неможливо без спеціальних алгоритмів робити припущення щодо соціальних акторів, які виконують роль керівників різного рангу.

Виклад основних результатів

Принцип ранжирування вузлів у великих мережах соціального характеру можливо сформулювати такими словами: *«лідером у мережі є той, хто віддає найбільшій кількості доручень тим, хто, у свою чергу, теж віддають багато (але менше за перших) доручень, і т.д.»*

Для математичної реалізації зазначеного принципу пропонується використати модифікацію алгоритму HITS (Hyperlink-Induced Topic Search), що базується на основі інформації, яка визначається множиною гіперпосилань [4, 5].

Згідно з алгоритмом HITS, для кожного вузла мережі розраховується два показника:

- показник порталності (hub);
- показник авторства (auth).

Показник порталності $hub(A_i)$ вузла A_i мережі, що складається із n вузлів дорівнює сумі значень авторства вузлів, на які він посилаються:

$$hub(A_i) = \sum_{A_j \rightarrow A_i} auth(A_j),$$

а показник авторства $auth(A_i)$ дорівнює сумі значень порталності вузлів, які посилаються на нього:

$$auth(A_i) = \sum_{A_j \rightarrow A_i} hub(A_j),$$

Задача ранжирування квазіієрархічних мереж соціального характеру має деякі особливості, що унеможливають пряме застосування алгоритму HITS. Слід відмітити, що в мережах соціального характеру майже відсутні випадки виключно однонаправлених зв'язків між вузлами. Навіть коли керівник віддає доручення підлеглим, останні мають звітувати про виконання [6]. З іншого боку, якщо при розгляді квазіієрархічних мереж спостерігається чіткий (ще й вагомий) односторонній зв'язок, то навіть не виникає питання відносно того, який з пари документів має більший ранг.

У соціальних квазіієрархічних мережах розглядаються два загальних варіанта відносин для пари вузлів:

- керівник-підлеглий;
- вузли рівного рангу.

Вага того факту, що соціальний актор A є керівником соціального актора B пропонується розглядати як сумарну вагу факторів, що визначаються наступними гіпотезами:

- керівник частіше зв'язується із підлеглим, тобто вага вихідного зв'язку мережі від керівника має перевищувати вагу вхідного зв'язку, що йде від підлеглого;
- керівник зв'язується із підлеглим хаотично, тоді коли у нього виникає потреба. Тобто нормована дисперсія розподілу інтервалів часу зв'язку у напрямку «керівник \rightarrow підлеглий» приймає відносно великі значення;
- підлеглий звітує перед керівником щодо виконання відпрацьованого доручення з певною закономірною періодичністю. Розподіл інтервалів часу зв'язку контактів «підлеглий \rightarrow керівник» (звітування), які починаються з попереднього зв'язку «керівник \rightarrow підлеглий» (відпрацювання доручення) має рівномірний розподіл.

Крім того, на практиці розподіл часових інтервалів між дорученнями керівника до підлеглого не є сталим, у той час, як розподіл інтервалів часу між дорученнями керівника та звітуваннями підлеглого є близьким до сталого.

Необхідно зазначити, що розглядається лише узагальнений та найбільш розповсюджений шаблон контактів підлеглого та керівника. У реальній роботі цей шаблон може бути доповнений у часткових випадках іншими специфічними характеристиками, які збільшать ефективність вагової оцінки контактів пари вузлів [6].

Таким чином, у методі, що розглядається, після зваження ребер здійснюється безпосереднє ранжирування мережі, виявляються вузли-лідери різного рангу. Також слід врахувати другу ключову відмінність між мережею документів, з'єднаних між собою гіперпосиланнями та мережами соціального характеру, а саме наявність в останніх ваги зв'язків, що необхідно враховувати.

Іноді трапляються випадки, коли керівник вищого рангу надає доручення співробітнику значно меншого рівня ієрархії, тобто у відповідній мережі з'являються неієрархічні зв'язки, але таких випадків, як правило, не багато, тому

мережу й розглядаємо як квазіієрархічну. Поміж тим, використання методу HITS в чистому вигляді для ранжирування вузлів в таких мережах може не надати необхідного результату, оскільки навіть слабкий та незначний зв'язок (якщо не враховувати ваги) може сильно вплинути на значення портальності та авторства вузлів [7].

З метою адаптації методу для виконання поставленої задачі пропонується при розрахунку кожного показника вузлів мережі враховувати вагу ребер. При цьому, у зв'язку з тим, що нерівномірність розподілу значень ваги, що може вплинути на достовірність результату, їх значення мають бути зменшені шляхом множення на деяку монотонно зростаючу функцію, менш круту за лінійну. В рамках методу, що пропонується, як таку функцію обрано логарифм, що приводить до модифікації наведених вище формул:

$$hub(A_i) = \sum_{A_j \rightarrow A_i} auth(A_j) \cdot \log_2 E_{ij},$$

$$auth(A_i) = \sum_{A_j \rightarrow A_i} hub(A_j) \cdot \log_2 E_{ji},$$

де E_{ij} (E_{ji}) – вага зв'язків між вузлами A_i та A_j (відповідно, між вузлами A_j та A_i).

Висновки

В роботі введено поняття квазіієрархічних мереж соціального характеру та досліджено характер зв'язків в таких мережах.

Запропоновано алгоритм ранжирування вузлів таких мереж, що базується на модифікації алгоритму HITS.

При практичному застосуванні моделі авторами показано, що модифікований метод надає у ряді випадків результати, що відповідають реальним соціальним відношенням, а показники авторства вузлів – їх соціальним ролям.

Список літератури

1. *Przewozniczek M., Walkowiak K.* Quasi-hierarchical evolution algorithm for flow assignment in survivable connection-oriented networks // *Int. J. Appl. Math. Comput. Sci.*, 2006, – **16**. – № 4. – P. 487–502.
2. *Ravasz E., Barabasi A.L.* Hierarchical organization in complex networks // *Physical Review e-67*, 2003. – P 026112 (1-7).
3. *Langville A.M., Meyer C.D.* *Google's PageRank and Beyond: The Science of Search Engine Rankings.* – Princeton NJ, USA: Princeton University Press, 2006. – 224 p.
4. *Kleinberg J.* Authoritative sources in a hyperlinked environment // *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, PA, 1998. – P. 668–677.
5. *Ландэ Д.В., Снарский А.А., Безсуднов И.В.* Интернетика: Навигация в сложных сетях: модели и алгоритмы. – М.: Либроком (Editorial URSS), 2009. – 264 с.
6. *Liu Y.Y., Jean-Jacques Slotine J.J., Barabasi A.L.* Control centrality and hierarchical structure in complex networks // *PLOS ONE*, 2012. – **7**. – № 9. – P. – e44459 (1-7).
7. *Bargh J.A., Chen M., Burrows L.* Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action // *Journal of Personality and Social Psychology*, 1996. – **71**. – № 2. – P. 230-244.