

# Квазіієрархічні складові соціальних мереж

Ланде Д.В., д.т.н., зав. відділом ІПРІ НАН України,  
Нечаєв О.О., аспірант ІС ЗЗІ НТУУ «КПІ»

Мережа вважається квазіієрархічною, якщо вона у деякому сенсі близька до ієрархічної, тобто може бути отримана з ієрархічної доповненням її деякою відносно невеликою кількістю зв'язків, що порушують ієрархічність [1, 2]. Під соціальною мережею будемо розуміти мережу, вузлами якої є соціальні актори (social actors), а зв'язками – контакти між ними, що супроводжуються обміном інформації.

Як квазіієрархічну соціальну мережу можна розглядати мережу, вузлами якої є співробітники організації, комерційної структури, учасники злочинного угруповання, яким притаманна внутрішня ієрархія. Як зв'язки можуть розглядатися контакти соціальних акторів шляхом використання електронної пошти, засобів мобільного зв'язку, ІР-телефонії, тощо.

Таку мережу можливо представити у вигляді графа, в якому позначками вузлів виступають певні маркери соціальних акторів (скриньки електронної пошти, абонентські телефонні номери, тощо), а ребрами направлені зв'язки між ними.

Коли мова йде про складні мережі (Complex Networks), 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.

Ітеративний принцип ранжирування вузлів у великих мережах соціального характеру можливо сформулювати наступним чином: *«лідером у мережі є той, хто віддає найбільшу кількість доручень тим, хто, у свою чергу, теж віддають багато (але менше за перших) доручень, і т.д. ...»*

Для математичної реалізації ранжирування вузлів соціальних мереж пропонується використати модифікацію алгоритму HITS (Hyperlink-Induced Topic Search), запропонованого для веб-сторінок, що відповідають запиту користувача [3, 4]. Згідно з алгоритмом HITS, для кожного вузла мережі розраховується два показника:

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

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

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

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

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

Наведені формули задають швидко збіжний ітеративний алгоритм розрахунку показників портальності та авторства для всіх вузлів мережі.

Задача ранжирування квазіієрархічних мереж соціального характеру має деякі особливості, що унеможливають пряме застосування алгоритму HITS.

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.

Розподіл інтервалів часу між дорученнями керівника та звітуваннями підлеглого є близьким до сталого. Дисперсія  $\sigma^2$  для інтервалів часу, що застосовується для формальної оцінки у подібних випадках розраховується за формулою:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (T_i - \langle T \rangle)^2,$$

де  $n$  – кількість інтервалів часу  $T_i$ ,  $\langle T \rangle$  – середнє значення інтервалу часу.

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 [5].

Слід врахувати другу ключову відмінність між мережею документів, з'єднаних між собою гіперпосиланнями та складними квазіієрархічними соціальними мережами, а саме наявність в останніх кількісної характеристики (ваги) зв'язків, що необхідно враховувати при ранжируванні.

Іноді трапляються випадки, коли керівник вищого рангу надає доручення співробітнику значно меншого рівня ієрархії, тобто у відповідній мережі з'являються неієрархічні зв'язки. 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 його ваги) може сильно вплинути на значення порталності 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 модифікацій формул, на яких базується алгоритм HITS:

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

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

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

Модифікований метод, як свідчить практичний досвід, дає результати, що точніше відповідають реальним соціальним відношенням, а показники авторства вузлів – попередньо наданим їх соціальним ролям.

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. *Kleinberg J.* Authoritative sources in a hyperlinked environment // *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, PA, 1998. – P. 668–677.
4. *Ландэ Д.В., Снарский А.А., Безсуднов И.В.* Интернетика: Навигация в сложных сетях: модели и алгоритмы. – М.: Либроком (Editorial URSS), 2009. – 264 с.
5. *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).