

АЛГОРИТМ РАНЖИРУВАННЯ ВУЗЛІВ КВАЗІЄРАРХІЧНИХ МЕРЕЖ СОЦІАЛЬНОГО ХАРАКТЕРУ

¹Інститут проблем реєстрації інформації НАН України,

²Інститут спеціального зв'язку і захисту інформації НТУУ «КПІ»

a4455395@gmail.com

Досліджено характер зв'язків в квазієрархічних мережах соціального характеру. Запропоновано алгоритм ранжирування вузлів таких мереж, що базується на модифікації відомого алгоритму HITS. Показано, що модифікований метод надає у ряді випадків результати, що відповідають реальним соціальним відношенням, а показники авторства вузлів – попередньо наданим їх соціальним ролям.

Ключові слова: мережі соціального характеру, соціальні актори, алгоритм ранжирування, метод HITS, соціальні ролі

Вступ

Під мережею соціального характеру будемо розуміти мережу, вузлами якої є соціальні актори (social actors), а зв'язками – контакти між ними, що супроводжуються обміном інформації. Будемо вважати мережу квазієрархічною, 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].

На практиці як квазієрархічну 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.

Мета роботи

Метою даної роботи є опис і обґрунтування алгоритму ранжирування вузлів квазієрархічних мереж соціального характеру.

Під ранжируванням мережі соціального характеру 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 (скриньки електронної пошти, абонентські телефонні номери, тощо), а ребрами направлені зв'язки між ними (рис. 1).

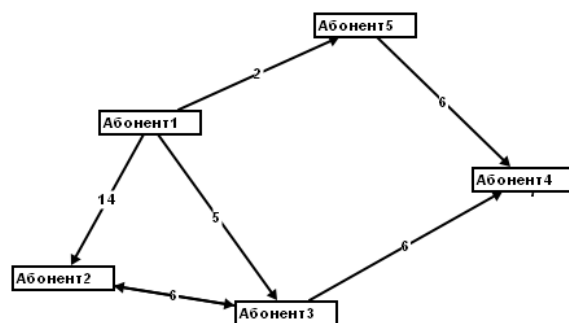


Рис. 1. Приклад мережі

Аналіз невеликих квазієрархічних мереж соціального характеру для визначення в них лідерів не представляє складності завдяки наочності їх представлення. Якщо трактувати мережу, зображену на рис. 1 як мережу зв'язків між керівництвом та підлеглими, то, враховуючи той факт, що керівник частіше відправляє повідомлення своїм підлеглим (відає доручення), чим навпаки, можна припустити, що керівником у даному випадку буде «Абонент1», оскільки він здійснює найбільшу кількість вихідних з'єднань з іншими абонентами.

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 (рис. 2).

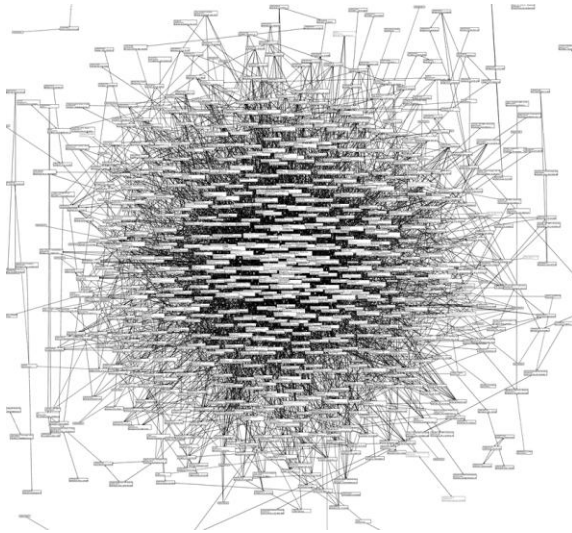


Рис. 2. Приклад складної квазієрархічної мережі соціального характеру [3]

Виклад основних результатів

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 зазначеного принципу пропонується використати модифікацію алгоритму HITS (Hyperlink-Induced Topic Search), запропонованого Дж. Клейнбергом для пошуку найбільш вагомих веб-сторінок відповідно запиту користувача на основі інформації, яка визначається множиною гіперпосилань [4, 5].

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 посиляється:

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

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 квазієрархічних 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 квазієрархічних 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 квазієрархічних 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 підлеглому;

- керівник зв'язується із підлеглим хаотично, тоді коли у нього виникає потреба. Тобто норма-вана дисперсія розподілу інтервалів часу зв'язку у нап-

рямку «керівник → підлеглий» приймає відносно великі значення;

– підлеглий звітує перед керівником щодо виконання відпрацьованого доручення з певною закономірною періодичністю. Розподіл інтервалів часу зв'яз-

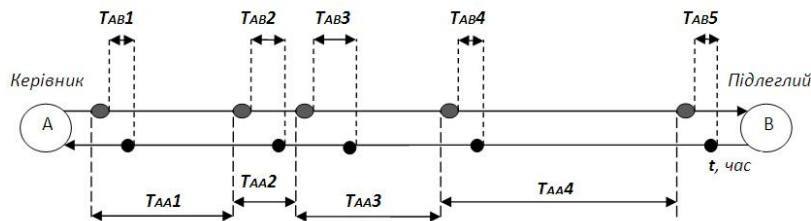


Рис. 3. Періодичність входних та вихідних контактів між керівником (А) та підлеглим (В)

На рис. 3 зображений приклад часового розподілу контактів між керівним вузлом (А) та підлеглим (В). Розподіл часових інтервалів між дорученнями керівника до підлеглого ($T_{AA1} \dots T_{AA4}$) не є сталим. Поміж тим, розподіл інтервалів часу між дорученнями керівника та звітуваннями підлеглого ($T_{AB1} \dots T_{AB5}$) є близьким до сталого. Дисперсія σ^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 [6].

Таким чином, у методі, що розглядається, після зваження ребер здійснюється безпосереднє ранжирування мережі, виявляються вузли-лідери рівного рангу.

На даному етапі слід врахувати другу ключову відмінність між мережею документів, з'єднаних між собою гіперпосиланнями та складними квазієрархічними

контактів «підлеглий → керівник» (звітування), які починаються з попереднього зв'язку «керівник → підлеглий» (відпрацювання доручення) має рівномірний розподіл.

ми мережами соціального характеру, 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 квазієрархічну. 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).

Для демонстрації різниці в використанні класичного та модифікованого алгоритмів розглянемо мережу з 19 вузлів, між якими завідомо розподілені певні соціальні ролі (рис. 4). Вузли в мережі мають між собою двосторонні контакти з визначеною вагою.

Для застосування алгоритмів слід попередньо оцінити кожний зв'язок за викладеною методикою та залишити лише ті, що відповідають відношенню «керівник → підлеглий».

На рис. 5 зображена однонаправлена мережа з відкинутими зв'язками «підлеглий → керівник». Також, візуалізовано ваги ребер (кількість доручень, які були відпрацьовані керівниками своїм підлеглим за певний проміжок часу).

Далі, для відфільтрованої однонаправленої мережі були розраховані показники авторства для кожного соціального актора із застосуванням класичного та модифікованого алгоритмів Клейнберга. Результати обчислень, відсортовані за

значенням показників авторства, приведені в табл. 1.

У табл. 1 спостерігається значна відмінність результатів ранжирування між різними варіаціями алгоритму HITS.

Висновки

В роботі введено поняття та досліджено характер зв'язків в квазі-ієрархічних мережах соціального характеру.

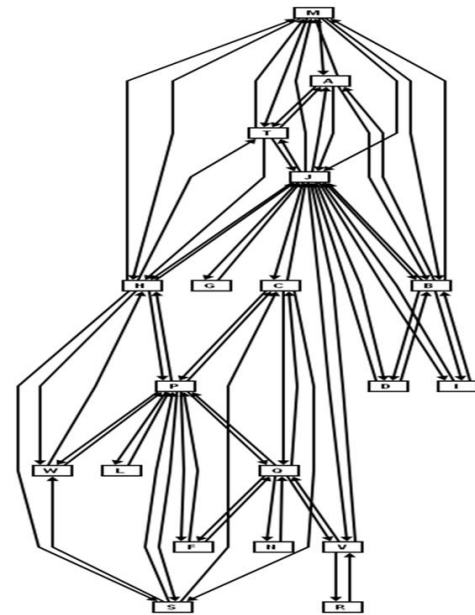


Рис. 4. Мережа контактів соціальних акторів з двосторонніми зв'язками

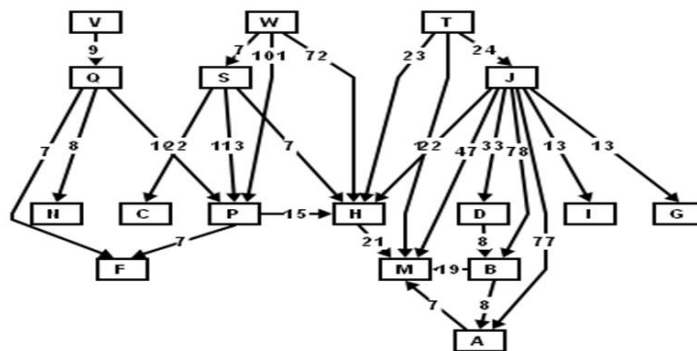


Рис. 5. Мережа з відфільтрованими зв'язками

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

Показано, що модифікований метод надає у ряді випадків результати, що від-

повідають реальним соціальним відношенням, а показники авторства вузлів – попередньо наданим їм соціальним ролям.

Табл. 1. Результати ранжування вузлів в мережі за класичним та модифікованим алгоритмами для вузлів з високим значенням показника авторства

	№	Вузол	Auth		№	Вузол	Auth
Класичний HTS	1	J	1	Модифікований HTS	1	T	1
	2	T	0,51606		2	J	0,49843
	3	B	0,50811		3	W	0,48652
	4	H	0,48723		4	B	0,48513
	5	A	0,31251		5	P	0,38136
	6	S	0,29636		6	S	0,3695
	7	M	0,27271		7	A	0,30829
	8	W	0,26119		8	C	0,28531
	9	P	0,23921		9	M	0,26766

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.

Статтю подано до редакції 15.03.2015