

УДК 004.067

Д. В. Ланде¹, О. О. Дмитренко¹, А. О. Снарський^{1,2}

¹Інститут проблем реєстрації інформації НАН України
вул. М. Шпака, 2, 03113 Київ, Україна

²Національний технічний університет України «КПІ імені Ігоря Сікорського»
Проспект Перемоги, 37, 03056 Київ, Україна

Дослідження показника мережевої релаксації як характеристики вузлів мережі

Робота присвячена дослідженню нової характеристики вузлів мережевих структур — показнику релаксації. Установлено, що на відновлення окремих уже традиційних показників мережі, після збурення окремих вузлів, впливає її топологія. Показник релаксації було використано для дослідження структури мережі термів, що побудована для предметної області «інтернет-речей». Завдяки застосуванню показника релаксації вдалося визначити найбільш важливі компоненти мережі.

Ключові слова: складна мережа, показник релаксації, степінь вузла, HITS, PageRank, предметна область, мережа термів, мережа природних ієрархій термінів.

Вступ

Розвиток інформаційних ресурсів у мережі Інтернет спричинив ряд специфічних проблем, що пов'язані, в першу чергу, зі стрімким збільшенням обсягів даних у веб-просторі, зокрема і непотрібних, шумових. При цьому виявилось, що багато задач, що виникають під час роботи з мережевим інформаційним простором [1], мають багато чого спільного з математичними науками, що відкриває широкі можливості для застосування потужного математичного апарату [1], [2]. Враховуючи проблеми розмірності та динаміки інформаційних ресурсів у глобальних мережах, для дослідження інформаційних потоків застосовується знання з області дискретної математики, зокрема теорії графів і мереж, розпізнавання образів (класифікація, кластерний аналіз), лінгвістики, цифрової обробки сигналів, вейвлет- і фрактального аналізу.

Поряд із традиційною теорією графів активно розвивається область дискретної математики, що має назву теорія складних мереж (від англ. Complex Networks) [3], яка вивчає характеристики мереж, враховуючи не тільки їхню топологію, але й

© Д. В. Ланде, О. О. Дмитренко, А. О. Снарський

статистичні явища, розподіл вагових значень окремих вузлів і ребер, ефекти протікання та провідності в таких мережах струму, рідини, інформації і т.д. Основною причиною виникнення розвитку цієї області є властивості реальних сучасних мереж, включаючи веб-простір, Peer-To-Peer-мережі, соціальні, біологічні мережі, — практично кожна з яких може вважатися складною.

У прикладних дослідженнях зазвичай застосовують типові для мережевого аналізу характеристики вузлів мережі, найважливішими серед яких на цей час вважають степінь вузла та показники, що відповідають двом алгоритмам HITS та PageRank.

У рамках теорії складних мереж пропонуються нові підходи до вирішення обчислювально-складних задач, характерних для сучасних мережевих структур [3, 4]. Дослідження статистичних властивостей, які характеризують поведінку мереж, створення моделі мереж, прогнозування поведінки мереж при зміні структурних властивостей — актуальні завдання теорії складних мереж.

Метою дослідження є введення нової характеристики вузлів складних мереж, визначити її «фізичний зміст» і показати унікальність серед інших характеристик, а також навести приклади застосування, зокрема, в комп'ютерній лінгвістиці.

Алгоритм HITS

Алгоритм ранжування *HITS* (Hyperlink Induced Topic Search), що був запропонований і розроблений у 1996 році Дж. Клейнбергом (J. M. Kleinberg) [5] забезпечує вибір з інформаційного масиву кращих «авторів» (першоджерел, на які посилаються інші документи) та «посередників» (документів, які посилаються на ці першоджерела). Документ буде вважатися хорошим «автором», якщо на нього посилаються хороші «посередники». У свою чергу, хорошими «посередниками» вважаються ті, які містять посилання на цінні першоджерела.

Для кожного документа d_i ($i=1, \dots, N$, де N — кількість документів) рекурсивно обчислюється його важливість як «автора» $a(d_j)$ і як «посередника» $h(d_j)$ відповідно до формул:

$$a(d_j) = \sum_{i=1, i \neq j}^N h(d_i), \quad h(d_j) = \sum_{i=1, i \neq j}^N a(d_i),$$

де $a(d_j)$ — це загальна кількість вхідних посилань документа d_j , а $h(d_j)$ — загальна кількість вихідних посилань документа d_j .

Після обчислення $a(d_j)$ та $h(d_j)$, здійснюється процес нормування:

$$a(d_j) = \frac{a(d_j)}{\sum_{i=1}^N a(d_i)}, \quad h(d_j) = \frac{h(d_j)}{\sum_{i=1}^N h(d_i)}.$$

Далі процес повторюється доти, поки значення $a(d_i)$ та $h(d_i)$ не стануть стійкими.

Як підсумок, відбувається ранжування гіпертекстових документів відповідно до $a(d_i)$ та $h(d_i)$.

Алгоритм PageRank

PageRank (Пейдж-ранк) — один із алгоритмів оцінки важливості та ранжування веб-сторінок за гіперпосиланнями — був створений у Стенфордському університеті Ларрі Пейджем і Сергієм Бріном у 1996 році в рамках науково-дослідного проекту про новий вид інформаційно-пошукової системи [6] та вперше використаний у Google.

Значення PageRank для веб-сторінки A обчислюється за такими правилами: нехай T_1, \dots, T_n — сторінки, що посилаються (цитують) сторінку A . Алгоритм також використовує коефіцієнт демпінгу d , значення якого знаходяться в проміжку між 0 та 1, та зазвичай має значення 0,85. Функція $C(T)$ дорівнює кількості посилань, що виходять зі сторінки T . Тоді значення PageRank сторінки A , $PR(A)$, дорівнює

$$PR(A) = (1 - d) + d \sum_{i=1}^n \frac{PR(T_i)}{C(T_i)},$$

де n — кількість веб-сторінок.

Не зважаючи на відмінності HITS і PageRank, у цих алгоритмах спільним є те, що «авторитетність» (вага) вузла, що відповідає веб-сторінці чи документу як першоджерелу, залежить від ваги інших вузлів, а «авторитетність» «посередника» залежить від того, наскільки «авторитетними» є вузли, на які він посилається.

Дослідження показника мережевої релаксації

У роботі представлено дослідження ще однієї характеристики вузлів мережі — показника мережевої релаксації. Зміст запропонованої характеристики полягає в наступному. Після досягнення стійких значень вузлів, унаслідок застосування ітераційного алгоритму HITS або PageRank, певному вузлу надається збурення. Виникає питання: скільки алгоритмічних ітерацій потрібно здійснити, щоб система врівноважилася — тобто значення всіх вузлів мережі стали стійкими.

У даному дослідженні як збурення було взято величину, що рівна середньому значенню відповідного показника для всіх вузлів.

Для ітераційного алгоритму HITS формально це можна представити як

$$h(d_j) = h_{init}(d_j) + \sum_{i=1}^N h(d_i),$$

де N — кількість вершин у мережі; $h_{init}(d_j)$ — значення «авторитетності» вузла d_j як «посередника» перед збуренням.

Необхідно визначити ітерацію k , після якої значення вузлів стануть стійкими, тобто для

$$\forall_j : |h_{k+1}(d_j) - h_k(d_j)| < \tau \text{ та } |a_{k+1}(d_j) - a_k(d_j)| < \tau,$$

де τ — задане порогове значення.

У випадку застосування ітераційного алгоритму PageRank

$$PR(A) = PR_{init}(A) + \sum_{i=1}^N PR(T_i),$$

де N — кількість вершин у мережі; $PR_{init}(A)$ — вагове значення вузла перед збуренням.

Необхідно визначити ітерацію k , після якої значення вузлів стануть стійкими для

$$\forall_A : |PR_{k+1}(A) - PR_k(A)| < \tau.$$

Отримане значення k для кожного збуреного вузла мережі і є показником релаксації мережі.

Наприклад, розглянемо мережу, що описується наступною матрицею суміжності:

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

У результаті застосування ітераційного алгоритму HITS для кожного вузла було обчислено показник релаксації мережі. На рис. 1 зображено стовбчасту діаграму, де представлені значення степеня та показника релаксації, отриманого для алгоритму HITS, для кожного вузла мережі, що відсортовані за зростанням степенів вузлів і нормовані на $[0,1]$.

На рис. 2 зображено стовбчасту діаграму, де представлені значення степеня та показника релаксації, отриманого в результаті застосування алгоритму PageRank, для кожного вузла мережі, що відсортовані за зростанням степенів вузлів і нормовані на $[0,1]$.

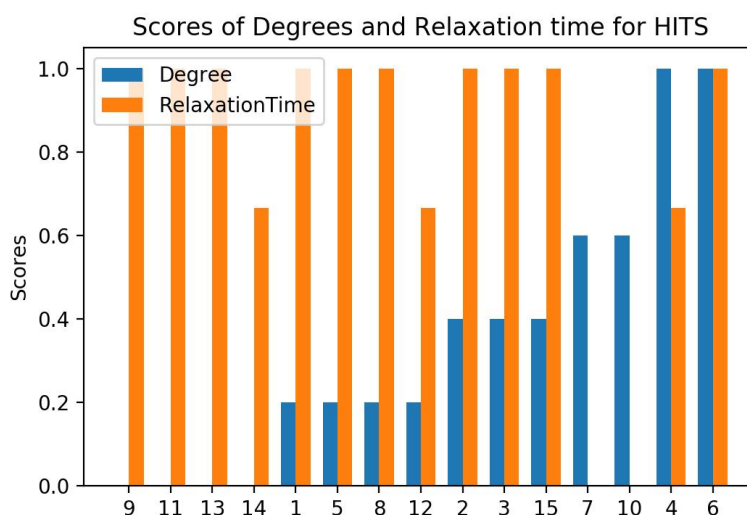


Рис. 1. Порівняльна стовбчаста діаграма степеня та показника релаксації для кожного вузла мережі для алгоритму HITS

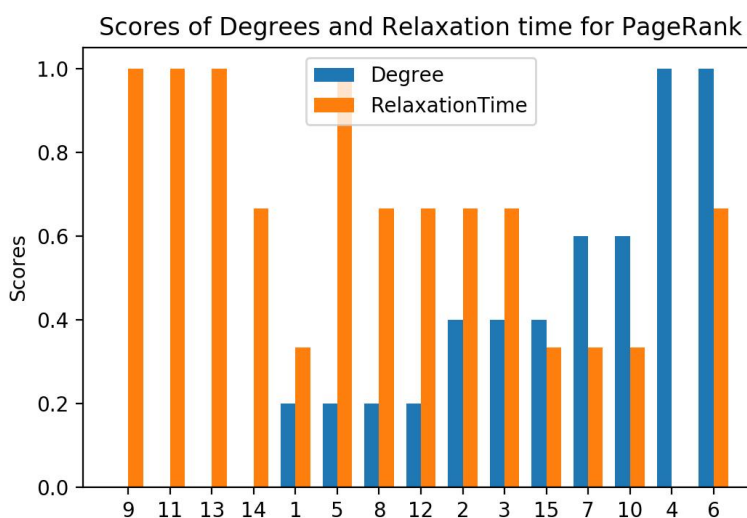


Рис. 2. Порівняльна стовбчаста діаграма степеня та показника релаксації для кожного вузла мережі для алгоритму PageRank

Можна помітити, що у випадку збурення вузлів, що мають низький степінь, кількість ітерацій, необхідних для стабілізації системи (релаксації значень всіх вузлів мережі), є більшою ніж у випадку збурення вузлів з більшим степенем. Це означає, що на вузол, який пов'язаний з невеликою кількістю інших вузлів у мережі, здійснюється менший вплив. А отже, значення вузла з меншим степенем стабілізується повільніше. Протилежну динаміку можна спостерігати для вузлів з високим степенем.

Дослідження показника релаксації для мережі термів

Дослідження також були проведені на основі мережі термів, яка була побудована для корпусу текстових документів, що тематично пов'язані з тематикою «інтернет-речей». Для створення мережі слів у даному дослідженні використову-

вався алгоритм формування направленої мережі зі слів і словосполучень — алгоритм формування мереж природніх ієрархій термінів. Як зазначено у роботі [9], алгоритм формування мереж природніх ієрархій термінів можна представити у вигляді послідовних етапів, які охоплюють попередню обробку отриманого корпусу текстових документів, виділення ключових слів і словосполучень, що є інформаційно-важливими в межах розглянутої предметної області, побудова компактифікованого графа горизонтальної видимості (Compactified Horizontal Visibility Graph — CHVG) [10], [11], перерахунок сортування вагових значень виділених термів за обраним ваговим критерієм і вибір з них найбільш вагомих. Кінцевим етапом є формування мережі природніх ієрархій термінів (з'єднання вузлів зв'язками «входження») та її відображення.

Для формування корпусу текстових документів було використано вільну доступну пошукову систему, яка індексує повний текст наукових публікацій — Google Scholar (<https://scholar.google.com>), звідки було вивантажено анотації перших 877-ми статей за запитом «Internet of things».

На етапі обробки текстових документів проведено процес попереднього лексичного аналізу — розбиття тексту на елементарні одиниці (токени або леми), вилучено стоп-слова, які не мають ніякого смислового навантаження, здійснено процес стематизації — скорочення слова до основи шляхом відкидання допоміжних частин (таких як закінчення чи суфікс) і подальше зважування та виокремлення термів. Як вагові значення термів для формування часового ряду як функції, яка ставить у відповідність слову число, в даному дослідженні використовується статистичний показник важливості терма — глобальний TF (Global Term Frequency, GTF), що дорівнює відношенню загальної кількості появи терма у всіх документах корпусу до загальної кількості термів у документах корпусу.

Використання цього показника дає змогу уникнути ситуації, що виникає під час роботи з текстовим корпусом заздалегідь визначеної тематики, коли інформаційно-важливий терм зустрічається майже у кожному документі корпусу і має низький ваговий показник TF.

Для послідовності термів і їхніх вагових значень будується компактифікований граф горизонтальної видимості (CHVG). Загалом, мережа слів із використанням алгоритму горизонтальної видимості будується в три етапи. На першому етапі на горизонтальній осі відмічається ряд вузлів, кожен з яких відповідає словам у тому порядку, в якому вони з'являються в тексті, а по вертикальній осі відкладаються вагові значення — числові оцінки. На другому етапі будується граф горизонтальної видимості. Третій етап полягає в тому, що отримана на попередніх етапах мережа компактифікується. В результаті буде отримано нову мережу слів — компактифікований граф горизонтальної видимості (CHVG).

Наступним кроком є перерахунок вагових значень, що відповідають термам у CHVG. Ця процедура дозволяє врахувати в подальшому також ті терми, які мають велике значення для загальної тематики текстового корпусу [11]. Під час виконання досліджень перерахунок ваг здійснюється з використанням алгоритму NITS [5, 12], завдяки якому визначається «авторство» чи «посередництво» для кожного вузла CHVG. Вибір форми вагового значення («авторство» або «посередництво») не має значення, оскільки граф є ненаправленим. Після цього всі терми

Таблиця 1. Списки найбільш вагомих термів (слів, біграм і триграм) для «інтернет-речей»

№	Слова	Біграми	Триграми
1	internet	internet_thing	internet_of_thing
2	iot	thing_iot	wireless_sensor_network
3	thing	futur_internet	secur_and_privaci
4	network	wireless_sensor	abstract_the_internet
5	technolog	smart_citi	base_on_internet
6	applic	base_internet	sensor_and_actuat
7	commun	secur_privaci	billion_of_devic
8	servic	connect_internet	comput_and_commun
9	devic	cloud_comput	privaci_and_secur
10	smart	sensor_network	thing_iot_applic
11	object	smart_object	comput_and_internet
12	inform	intern_confer	big_data_analyt
13	comput	physic_object	inform_and_commun
14	secur	fog_comput	sensor_network_wsn
15	data	iot_devic	number_of_devic
16	sensor	technolog_internet	radio_frequenc_identif
17	develop	suppli_chain	servic_and_applic
18	connect	sensor_actuat	number_of_smart
19	challeng	big_data	applic_of_iot
20	system	abstract_internet	thing_and_cloud

Таблиця 2. Степінь і показник релаксації для кожного вузла мережі для алгоритму HITS

Терм	Степінь вузла	Показник релаксації	Терм	Степінь вузла	Показник релаксації
challeng	0	0	data	0.2	0.4
develop	0	0	inform_and_commun	0.2	0
intern_confer	0	0	iot_devic	0.2	0
radio_frequenc_identif	0	0	object	0.2	0.4
suppli_chain	0	0	privaci_and_secur	0.2	0
system	0	0	secur_and_privaci	0.2	0
billion_of_devic	0.1	0	sensor_actuat	0.2	0
cloud_comput	0.1	0	sensor_and_actuat	0.2	0
connect	0.1	0	servic_and_applic	0.2	0
fog_comput	0.1	0	smart_object	0.2	0
futur_internet	0.1	0	technolog_internet	0.2	0
inform	0.1	0.4	wireless_sensor	0.2	0.4
number_of_devic	0.1	0	applic	0.3	0.8
number_of_smart	0.1	0	devic	0.3	0.4
physic_object	0.1	0	internet_of_thing	0.3	0
servic	0.1	0.4	internet_thing	0.3	0.2
smart_citi	0.1	0	network	0.3	0.8
technolog	0.1	0	secur	0.3	0.6
thing_and_cloud	0.1	0	secur_privaci	0.3	0.6
abstract_internet	0.2	0	sensor_network_wsn	0.3	0
abstract_the_internet	0.2	0	smart	0.3	0.6
applic_of_iot	0.2	0	thing_iot	0.3	0.6
base_internet	0.2	0	comput	0.4	0.6
base_on_internet	0.2	0	iot	0.4	0.8
big_data	0.2	0.4	sensor_network	0.4	0.6
big_data_analyt	0.2	0	thing_iot_applic	0.4	0
commun	0.2	0.4	wireless_sensor_network	0.4	0
comput_and_commun	0.2	0	thing	0.5	0.8
comput_and_internet	0.2	0	sensor	0.6	1
connect_internet	0.2	0	internet	1	0

У табл. 3 представлено значення степеня та показника релаксації, отримано-го для алгоритму PageRank, для кожного вузла мережі, відсортовані за зростанням степенів вузлів і нормовані на [0,1].

Таблиця 3. Степінь і показник релаксації для кожного вузла мережі для алгоритму PageRank

Терм	Степінь вузла	Показник релаксації	Терм	Степінь вузла	Показник релаксації
challeng	0	0	data	0.2	1
develop	0	0	inform_and_commun	0.2	0
intern_confer	0	0	iot_devic	0.2	0
radio_frequenc_identif	0	0	object	0.2	0.5
suppli_chain	0	0	privaci_and_secur	0.2	0
system	0	0	secur_and_privaci	0.2	0
billion_of_devic	0.1	0	sensor_actuat	0.2	0.5
cloud_comput	0.1	0	sensor_and_actuat	0.2	0
connect	0.1	0.5	servic_and_applic	0.2	0
fog_comput	0.1	0	smart_object	0.2	0
futur_internet	0.1	0	technolog_internet	0.2	0
inform	0.1	0.5	wireless_sensor	0.2	0.5
number_of_devic	0.1	0	applic	0.3	0.5
number_of_smart	0.1	0	devic	0.3	0.5
physic_object	0.1	0	internet_of_thing	0.3	0
servic	0.1	0.5	internet_thing	0.3	0.5
smart_citi	0.1	0	network	0.3	0.5
technolog	0.1	0.5	secur	0.3	0.5
thing_and_cloud	0.1	0	secur_privaci	0.3	0.5
abstract_internet	0.2	0.5	sensor_network_wsn	0.3	0
abstract_the_internet	0.2	0	smart	0.3	0.5
applic_of_iot	0.2	0	thing_iot	0.3	0.5
base_internet	0.2	0.5	comput	0.4	0.5
base_on_internet	0.2	0	iot	0.4	0.5
big_data	0.2	0.5	sensor_network	0.4	0.5
big_data_analyt	0.2	0	thing_iot_applic	0.4	0
commun	0.2	0.5	wireless_sensor_network	0.4	0
comput_and_commun	0.2	0	thing	0.5	1
comput_and_internet	0.2	0	sensor	0.6	1
connect_internet	0.2	0	internet	1	0.5

На рис. 4 зображено стовбчасті діаграми, де представлені значення степеня та показника релаксації, отриманого для алгоритму HITS та PageRank, для кожного вузла мережі, відсортовані за зростанням степенів вузлів та нормовані на [0,1].

Найбільш цікавими із семантичної точки зору в представленій мережі термів виявились вузли з високим степенем, серед яких можна виділити такі відповідні їм терми: «internet», «sensor», «thing».

Аналізуючи результати, отримані внаслідок проведення досліджень (табл. 2 та табл. 3), можна зробити висновок, що вузли, яким відповідають терми «challeng», «develop», «system», «intern_confer», «suppli_chain» та «radio_frequenc_identif», та які не пов'язані з іншими вузлами мережі (їхній степінь рівний нулю) мають нульовий показник релаксації у випадку застосування алгоритмів HITS та PageRank. Це є наслідком того, що збурення, яке надається таким вузлам, не розповсюджується на іншу частину мережі й інші вузли внаслідок цього не збурюються.

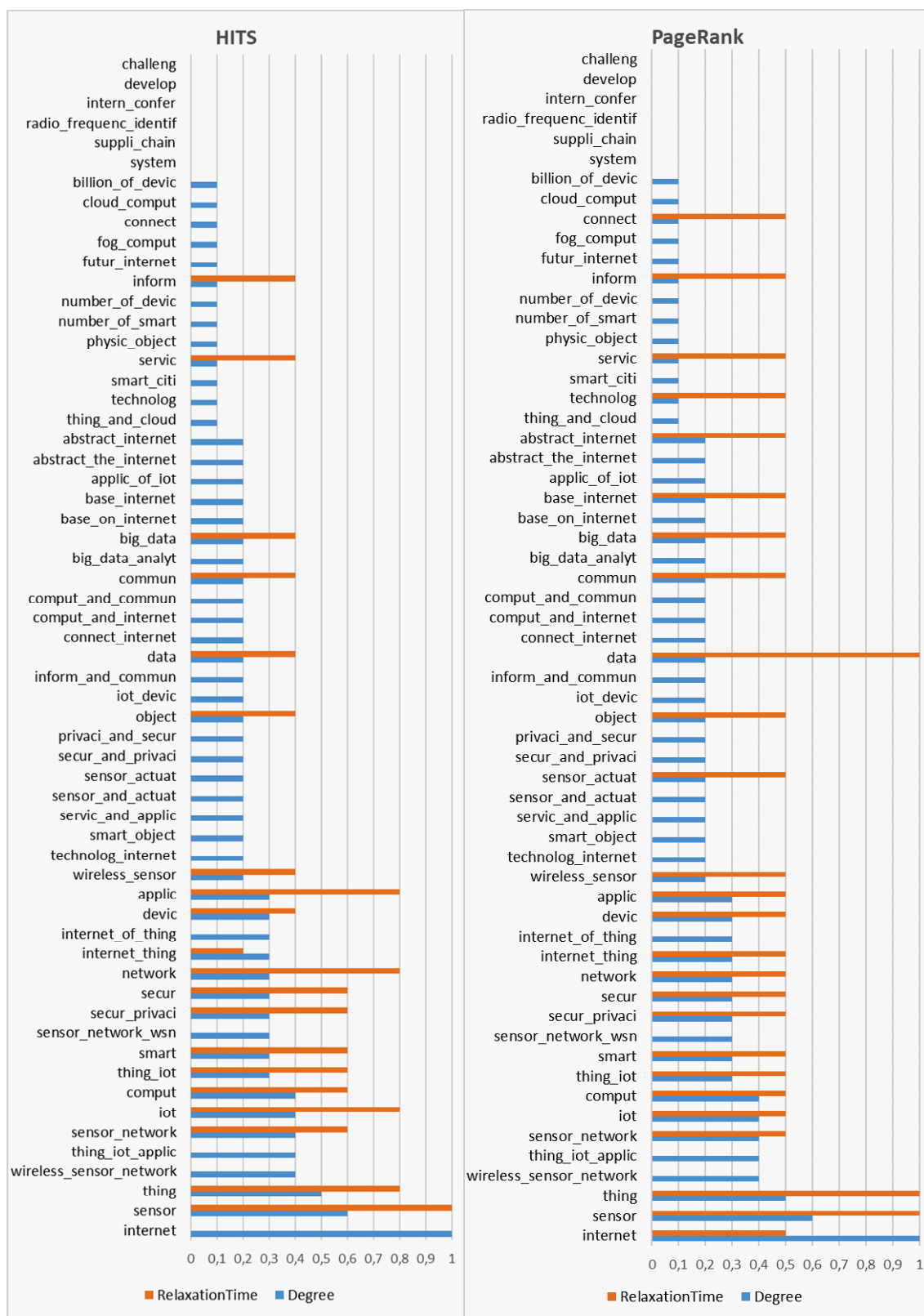


Рис. 4. Столбчасті діаграми значення степеня та показника релаксації, що отримані для алгоритму HITS та PageRank для кожного вузла мережі

Також можна помітити, що майже всі вузли зі степенем 0,1 та 0,2 мають також нульовий показник релаксації у випадку застосування алгоритму HITS. У даному випадку, наявність лише одного вхідного або вихідного з'єднання з іншими вузлами мережі призводить до швидкого відновлення всієї системи після збурення. Терми «inform», «servic», «commun», «data», «object», «big_data» та «wireless_sensor» мають посередній показник релаксації, причиною якого є посилення на вузли, які не мають вихідних посилення, а навпаки — принаймні одне вхідне посилення. Це призводить до відновлення значень таких вузлів і, як наслідок, всієї мережі вже на початкових ітераціях алгоритму.

У випадку PageRank переважна більшість вузлів зі степенем 0,1 та 0,2 мають також нульовий показник релаксації. Середній показник релаксації, окрім термів «inform», «servic», «commun», «object», «big_data» та «wireless_sensor» (як у випадку HITS), мають також «connect», «technology», «abstract_internet», «base_internet» та «sensor_actuat». Причиною цього є наявність вихідних посилення на вузли, що є першоджерелами «важливих» вузлів. Аномально високим є показник релаксації після збурення вузла «data» у випадку застосування алгоритму PageRank. Довготривале відновлення значень системи обумовлене додатково ізольованістю вузлів «data»-«big_data»-«big_data_analyt», тобто наявністю окремого невеликого класу.

Також середнім за своїм числовим значенням є показник релаксації після збурення вузлів зі степенем 0,3 та 0,4 для HITS та PageRank. Серед них «internet_of_thing», «thing_iiot_applic», «sensor_network_wsn» та «wireless_sensor_network» мають нульовий показник релаксації, оскільки не мають вихідних посилення на вузли, і, навпаки, вузли, що на них посилюються, є «авторитетними».

Під час збурення вузлів мережі, які мають степінь вищий або рівний середньому (0,5 та 0,6), у випадку застосування алгоритмів HITS та PageRank можна спостерігати відносно тривале відновлення значень мережі. Така ситуація виникає тому, що вузли «thing» та «sensor» мають велику кількість вихідних посилення, та, навпаки, характеризуються відсутністю вхідних посилення. Тобто розглянуті вузли є хорошими «розповсюдниками» наданого їм збурення.

Щодо вузла «internet», то після надання йому збурення у випадку застосування алгоритму HITS спостерігається дуже швидка релаксація значень мережі, в той час, коли для PageRank характерна інша поведінка. Це обумовлено тим, що початкове збурення отримувало значення «авторитетності» вузла як «посередника» $h_{init}(d_j)$.

Висновки

Запропоновано нову характеристику вузлів мережі — показник релаксації. Показано, що у випадку збурення вузлів, що мають низький степінь, кількість ітерацій, які необхідні для стабілізації системи (релаксації значень усіх вузлів мережі), є більшою ніж у випадку збурення вузлів з більшим степенем. На основі алгоритму природніх ієрархій термінів для предметної області «інтернет-речей» побудовано направлену мережу термів. Проаналізувавши результати досліджень, що проведені на основі побудованої мережі, можна зробити висновок, що на відновлення значень системи, після збурення окремих вузлів, впливає топологія мережі,

а саме: відсутність або наявність вихідних посилань у збурених вузлів і їхню кількість; наявність вихідних посилань у збурених вузлів на вузли, які є тільки «авторами»; наявність вихідних посилань на вузли, що є першоджерелами «важливих» вузлів; наявність «авторитетних» вузлів, які посилаються на збурений вузол; входження збуреного вузла в окремих невеликий кластер і т.д. Враховуючи це, для розглянутої предметної області «інтернет-речей» цікавими елементами, з точки зору дослідження, виявилися такі поняття, які відповідають вузлам побудованої мережі термів, а саме: «internet» та «data».

Отже, запропонована в ході досліджень числова характеристика вузлів мережі може бути використана під час дослідження та аналізу структури мережі, даючи змогу виявити найбільш важливі структурні елементи. Також результати досліджень можуть бути використані під час побудови персональних пошукових інтерфейсів користувачів інформаційно-пошукових систем, що, в свою чергу, дозволить спростити процес пошуку необхідної інформації.

1. Ландэ Д.В., Снарский А.А., Безсуднов И.В. Интернетика: Навигация в сложных сетях: модели и алгоритмы. Москва: Editorial URSS, 2009.
2. Додонов А.Г., Ландэ Д.В., Путятин В.Г. Компьютерные сети и аналитические исследования. Киев: ИПРИ НАН Украины, 2014. 486 с. ISBN 978-966-02-7422-8.
3. Newman M.E.J. The structure and function of complex networks. *SIAM Review*. 2003. Vol. 45. P. 167–256. doi: 10.1137/S003614450342480.
4. Dorogovtsev S.N., Mendes J.F.F. Evolution of networks: from biological networks to the Internet and WWW. *Oxford University Press*. 2003, P. 180–182.
5. Kleinberg J.M. Authoritative sources in a hyperlink environment. In Processing of ACM-SIAM Symposium on Discrete Algorithms. 1998. **46**(5). P. 604–632.
6. Page L., Brin S., Motwani R., and Winograd T. PageRank: Bringing order to the web. *Stanford Digital Libraries Working Paper*. 1997. 72.
7. Page L., and Brin S. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 30: 1998. P. 107–117. ISSN 0169-7552. doi:10.1016/S0169-7552(98)00110-X.
8. Leskovec J., Anand R., and Jeffrey D.U. Mining of massive datasets. *Cambridge university press*. 2014.
9. Lande D., Snarskii A., Yagunova E. Network of Natural Hierarchies of Terms of News Messages on Events «Euromaydan». CEUR Workshop Proceedings. Selected Papers of XVI All-Russian Scientific Conference «Digital libraries: Advanced Methods and Technologies, Digital Collections». 2013. Vol. 1297. P. 66–74.
10. Lacasa L., Luque B., Ballesteros F., Luque J., and Nuño J.C. From time series to complex networks: the visibility graph. *Proc. Natl. Acad. Sci. USA*. **105**. 2008. P. 4972–4975. doi:10.1073/pnas.0709247105
11. Lande D.V., Snarskii A.A., Yagunova E.V., and Pronoza E. The Use of Horizontal Visibility Graphs to Identify the Words that Define the Informational Structure of a Text. In: Proceedings of the 12th Mexican International Conference on Artificial Intelligence. 2013. P. 209–215. doi:10.1109/MICAI.2013.33
12. Langville A.N., and Meyer C.D. Google's PageRank and beyond: the science of searchengine rankings. *Princeton university press*. 2011.

Надійшла до редакції 07.03.2019