

**НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ ПРОБЛЕМ РЕЄСТРАЦІЇ ІНФОРМАЦІЇ НАН УКРАЇНИ**

**ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ
ТА БЕЗПЕКА**

**МАТЕРІАЛИ XXI МІЖНАРОДНОЇ
НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ**

ВИПУСК 21

Київ – 2021

*Рекомендовано до друку Вченою радою
Інституту проблем реєстрації інформації НАН України
(протокол № 12 від 28 грудня 2021 р.)*

Інформаційні технології та безпека. Матеріали XXI Міжнародної науково-практичної конференції ІТБ-2021. – Київ: Інжиніринг. – 174 с.
ISBN: 978-966-2344-84-4

До збірника увійшли матеріали доповідей, представлених на XXI Міжнародній науково-практичній конференції «Інформаційні технології та безпека» (ІТБ-2021, 9 грудня 2021 року, м. Київ, Україна).

У збірнику представлені матеріали, присвячені питанням безпеки та живучості критичних інфраструктур, комп'ютерного моделювання складних систем, технологій аналітики великих обсягів даних (Big Data), аналітичних систем на основі відкритих джерел інформації (OSINT), моделювання, аналізу та прогнозування процесів мережевої взаємодії, методів і засобів підтримки прийняття рішень.

Для фахівців в області інформаційних технологій, інформаційної і кібернетичної безпеки, а також для аспірантів і студентів старших курсів вищої школи відповідних спеціальностей.

Редакційна колегія:

О.Г. Додонов, д.т.н., професор; В.В. Голенков, д.т.н., професор; Д.В. Ланде, д.т.н., професор; В.В. Мохор, член-кор. НАН України, д.т.н., професор; В.В. Циганок, д.т.н., с.н.с.; О.С. Горбачик, к.т.н., с.н.с.; М.Г. Кузнецова, к.т.н., с.н.с.; О.В. Андрійчук, к.т.н.

ISBN 978-966-2344-84-4

© Інститут проблем реєстрації
інформації НАН України, 2021

© Колектив авторів, 2021

ПОКАЗНИК ЧАСУ РЕЛАКСАЦІЇ ЯК УНІКАЛЬНА ХАРАКТЕРИСТИКА ДЛЯ КЛАСТЕРИЗАЦІЇ МЕРЕЖ

А.О. Снарський^{1,2}[0000-0002-4468-4542] Д.В. Ланде^{1,2}[0000-0003-3945-1178]

О.О. Дмитренко^{1,2}[0000-0001-8501-5313]

- ¹ Інститут проблем реєстрації інформації НАН України, Київ, Україна
² Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна

asnarskii@gmail.com, dwlande@gmail.com, dmytrenko.o@gmail.com

В цій роботі досліджується числові характеристики вузлів мережевих структур – показник часу релаксації мережі та індивідуальний показник часу релаксації вузла, які характеризують стійкість складної мережі та, відповідно, кожного вузла окремо до зовнішніх збуджень. Обчислення показників часу релаксації здійснюється за допомогою уповільненого ітераційного алгоритму HITS. Показано, що показник часу релаксації є унікальними числовими характеристиками вузлів мережі, і їх можна використовувати для знаходження центроїдів кластерів та об'єднання вузлів у групи за цими показниками. Апробацію представлених характеристик показника часу релаксації та індивідуального показника часу релаксації було проведено на прикладі кластеризації випадкових мереж з чітко вираженими кластерами.

Вступ

Складні мережі широко поширені у природі. Наприклад, такі мережі, як Всесвітня мережа Інтернет, мережі співавторства [1] та інші є складними [2], [3]. Всі вони володіють нетривіальними топологічними властивостями, а отже, становлять великий інтерес для досліджень. Важливе місце також посідають соціальні мережі, де бінарні зв'язки між людьми у групі можна представити у вигляді мережі, де кожен об'єкт – це точка, а його зв'язок з іншим об'єктом – дуга. Розбиття вибірки об'єктів на підмножини, що не перетинаються, тобто розбиття на кластери й, наприклад, виявлення спільнот у соціальних мережах є актуальним завданням.

Існує безліч методів кластеризації графів складних мереж [4, 5], які відрізняються ідеєю об'єднання схожих вузлів. Оскільки різні моделі використовують різні алгоритми, то розрізняють різні моделі кластеризації, зокрема такі як моделі зв'язності, центроїдні, статистичні, групові, нейронні моделі та інші. Наприклад, існують моделі, які ґрунтуються на основі зв'язку, де об'єкти, які знаходяться у просторі ближче, є більш подібними (спорідненими), а також існують моделі, що засновані на знаходженні центроїда – кластери представлені у вигляді центрального вектора.

В цій роботі ми пропонуємо використовувати нову характеристику для кластеризації складних мереж – показник часу релаксації [6].

Показник часу релаксації

Під час дослідження складних мереж було помічено, що значення мережевих характеристик вузлів після зовнішніх збурень, в результаті перерахунку відновлюються до своїх початкових рівноважних значень за деякий індивідуальний для кожного вузла час. Порівнюючи показник часу релаксації із самою мережевою характеристикою, значення якої піддавалося збуренню, простежується нечітка залежність, яка показує, що показник часу релаксації є унікальною та неподібною до інших числовою характеристикою вузлів мережі [6].

У випадку дослідження складних мереж показником часу релаксації називається кількість ітераційних кроків τ відповідного алгоритму, які необхідні, щоб з деякою наперед заданою точністю μ – умовою збіжності (зазвичай $\mu=10^{-4}$) досягти початкових рівноважних числових значень певної характеристики після її збурення.

У випадку, коли досліджується відновлення всієї мережі, то показником часу релаксації мережі після збурення певного m -го вузла $\tau^{(m)}$ називають кількість ітераційних кроків, які необхідні, щоб відновилося значення кожного вузла (відновились вся мережа), або $\max_k(\tau_k^{(m)})$ ($k = 1, \dots, n$, де n – кількість вузлів мережі). А кількість ітерацій $\tau_m^{(m)}$ (або просто τ_m), які необхідні для відновлення саме того вузла, числове значення характеристики якого було збурене, називатиметься індивідуальним показником часу релаксації вузла. Тобто перший показник характеризує вузол в термінах стійкості всієї мережі після наданого йому зовнішнього збурення, останній – характеризує індивідуальну стійкість вузла після його збурення.

В якості числової характеристики, числове значення якої піддається збуренню, в цій роботі використовувалась характеристика, що відповідає ітераційному алгоритму ранжування HITS (Hyperlink Induced Topic Search). Цей алгоритм був запропонований та розроблений в 1998 році Дж. Клейнбергом [7] для вибору із масиву документів кращих «авторів» (першоджерел, на які посилаються інші документи) та «посередників» (документів, які посилаються на ці першоджерела). Для кожного документа j обчислюється його важливість як «автора» $a(j)$ (з англ. Authority) і як «посередника» $h(j)$ (з англ. Hub) відповідно до формул:

$$a(j) = \sum_{i \rightarrow j} h(i), \quad h(j) = \sum_{j \rightarrow i} a(i), \quad (1)$$

Оскільки вузли складної мережі мають велику кількість взаємозв'язків, то ітераційний процес перерахунку значень вузлів після їх збурення зазвичай є швидким, і достатньо всього декілька ітераційних кроків – часу релаксації, аби досягти початкових рівноважних значень характеристики HITS. Тож після збурення певного вузла j перерахунку числових значень,

для того, щоб уповільнити процес збіжності, у даній роботі пропонується здійснити уповільнення алгоритмів. Тож після надання збурення одному із вузлів мережі застосовується відповідний ітераційний алгоритм HITS з уповільненням:

$$h_0(j) \rightarrow h_1(j) \rightarrow h_2(j) \rightarrow \dots \rightarrow h_{\text{рівноважне}}(i) \quad (2)$$

$$\widehat{A}h_1(i) = h_2(i), \widehat{A}h_n(i) = h_{n+1}(i) \quad (3)$$

$$h_{n+1}(i) \leftarrow \beta h_{n+1}(i), \quad (4)$$

де $0 < \beta < 1$ – коефіцієнт уповільнення, \widehat{A} – оператор алгоритму HITS.

Приклади кластеризації мереж

Апробацію представлених характеристик показника часу релаксації та індивідуального показника часу релаксації було проведено на прикладі кластеризації випадкових мереж різної розмірності з чітко вираженими кластерами. Зокрема, було досліджено випадково згенеровану матрицю розмірності 30×30 з 3-ма кластерами та матрицю 100×100 з 4-ма кластерами.

Для кожної мережі, що описується відповідною матрицею, після збурення значення мережевої характеристики $h(j)$ завдяки уповільненому алгоритму HITS було обчислено показник часу релаксації мережі та індивідуальний показник часу релаксації для кожного вузла. Для кожної мережі коефіцієнт уповільнення β алгоритму HITS та умова збіжності μ числових значень $h(j)$ та $a(j)$ до передзбурених значень обиралася індивідуально.

Також для того щоб уникнути великих числових значень, які інколи виникають після застосування уповільнених алгоритмів, всі результуючі значення показників часу релаксації були нормовані в інтервалі $[0,1]$.

На рис. 1 та 2 наведено візуальне представлення мереж. На кожному з рисунків вузли розфарбовані за класом модулярності (рис. 1(a) та 2(a)) та за показником часу релаксації мережі та індивідуальним показником часу релаксації вузла (для представлених прикладів нормовані числові значення цих характеристик співпадають та представлені у вигляді міток вузлів – рис. 1(b) та 2(b)). Показник часу релаксації для мережі, що представлена на рис. 1, був отриманий в результаті уповільнення алгоритму HITS з коефіцієнтом уповільнення $\beta = 0.9$ та умовою збіжності $\mu = 10^{-6}$. Для мережі, що представлена на рис. 2 – $\beta = 0.9$ та $\mu = 10^{-4}$.

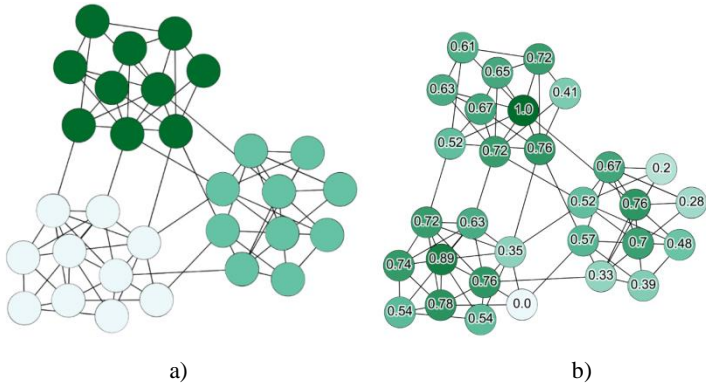


Рис. 1. Приклад мережі, що відповідає випадково згенерованій матриці розмірності 30×30 з 3-ма кластерами, де групи вузлів об'єднані (розфарбовані) за а) класом модулярності та б) показником часу релаксації (числові значення представлені у вигляді міток вузлів)

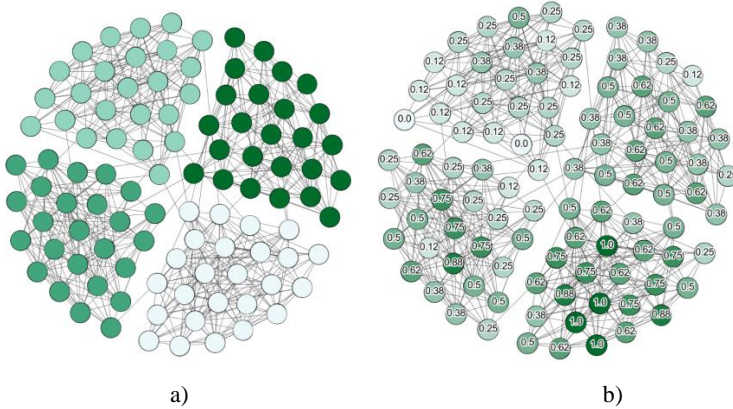


Рис. 2. Приклад мережі, що відповідає випадково згенерованій матриці розмірності 100×100 з 4-ма кластерами, де групи вузлів об'єднані (розфарбовані) за а) класом модулярності та б) показником часу релаксації (числові значення представлені у вигляді міток вузлів)

Можна помітити, що групи вузлів, об'єднані за класом модулярності, відрізняються від груп, отриманих за допомогою об'єднання вузлів за

показником часу релаксації. Це означає, що показник часу релаксації мережі та індивідуальний показник часу релаксації вузла є унікальними числовими характеристиками вузлів мережі, і їх можна використовувати для знаходження центроїдів кластерів та об'єднання вузлів у групи за цими показниками.

Висновки

В цій роботі було розглянуто нові числові характеристики вузлів мережеских структур – показник часу релаксації мережі та індивідуальний показник часу релаксації вузла. Апробацію показника часу релаксації було проведено на прикладі кластеризації випадкових мереж з чітко вираженими кластерами. Було встановлено, що показник часу релаксації мережі та індивідуальний показник часу релаксації вузла є унікальними числовими характеристиками вузлів мережі, і їх можна використовувати для знаходження центроїдів кластерів та об'єднання вузлів у групи за цими показниками.

Літературні джерела

1. Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T.: Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3-4), 590-614 (2002).
2. Newman, M. E. J.: The structure and function of complex networks. *SIAM Review*, vol. 45. pp. 167–256. (2003). doi:10.1137/S003614450342480
3. Strogatz, S. H.: Exploring complex networks. *nature*, 410(6825), 268-276 (2001).
4. Lancichinetti, A., & Fortunato, S. Community detection algorithms: a comparative analysis. *Physical review E*, 80(5), 056117 (2009).
5. Estivill-Castro, V.: Why so many clustering algorithms: a position paper. *ACM SIGKDD explorations newsletter*, 4(1), 65-75 (2002).
6. Lande D., Snarskii A., Dmytrenko O., Subach I.: Relaxation time in complex network. *ARES '20: Proceedings of the 15th International Conference on Availability, Reliability and Security* August 2020. Article No.: 99 1-6. (2020) doi: <https://doi.org/10.1145/3407023.3409231>
7. Kleinberg, J. M.: Authoritative sources in a hyperlinked environment. In *Processing of ACM-SIAM Symposium on Discrete Algorithms*, 46(5), pp. 604–632 (1998).

ЗМІСТ

<i>О.Г. Додонов, О.С.Горбачик, М.Г.Кузнєцова</i> Автоматизовані системи організаційного управління об'єктів критичних інфраструктур: безпека і функціональна стійкість.	3
<i>А.О. Снарський, Д.В. Ланде, О.О. Дмитренко</i> Показник часу релаксації як унікальна характеристика для кластеризації мереж.	9
<i>Oleksandr Koval, Valeriy Kuzminykh, Iryna Husyeva, Xu Beibei, Zhu Shiwei</i> Improving the efficiency of typical scenarios of analytical activities.	14
<i>Т. П. Рудник, О. Р. Чертов</i> Метод визначення політичного рейтингу за допомогою соціальних мереж.	23
<i>А.М., Соколов, Д.В. Ланде</i> Розподілені інтелектуальні агенти добування контенту із соціальних мереж.	26
<i>А.В. Никифоров, А.Г. Додонов, В.Г. Путятин</i> Принятие решений организационного управления на основе онтологии деятельности.	30
<i>Юлія Рогущина, Анатолій Гладун</i> Використання мереологічного підходу для формування структури семантичних зв'язків типу «частина-ціле» між сторінками семантизованого Wiki-ресурсу.	37
<i>Georgy Vedmedenko, Iryna Stopochkina, Oleksii Novikov, Mykola Ilin</i> Cascading effects simulation for cyber attacks on the power supply networks.	44
<i>Михайло Коломицев, Світлана Носок</i> Динамічне маскування даних зі збереженням формату.	50
<i>Наталія Кузнєцова, Петро Бідюк</i> Аналіз і прогнозування відмовостійкості засобів зберігання інформації.	55
<i>Юлія Рогущина Віталіївна</i> Розробка засобів аналізу контенту семантизованих Wiki-ресурсів.	61
<i>Nataliia Kuznietsova, Amoroso Remi</i> An approach to green financial credit risks modeling	65
<i>Анатолій Гладун, Родріго Мартінез-Бежар</i> Розробка структурованого інформаційного поля об'єкта із заданими властивостями для побудови його семантичної моделі	70
<i>Гальчинський Л.Ю., Грайворонський М.В., Носок С.О.</i> Оцінювання методів машинного навчання для виявлення DoS/DDoS атак в IoT	75