УДК 519.87;004.94
DOI: 10.31673/2409-7292.2020.030613

Савченко В.А., Ахрамович В.М., Акулінічева М.В.

ОЦІНЮВАННЯ ПАРАМЕТРІВ БЕЗПЕКИ ПЕРСОНАЛЬНИХ ДАНИХ У СТЕПЕНЕВИХ СОЦІАЛЬНИХ МЕРЕЖАХ НА ОСНОВІ ЇХ ТОПОЛОГІЇ

У статті досліджується модель степеневої соціальної мережі, яка, на відміну від класичних підходів, дозволяє аналізувати динамічні процеси взаємодії окремих агентів всередині мережі, зокрема щодо поширення інформації соціального впливу. Розширення соціальних мереж, приєднання нових вузлів призводить до зростання навантаження на систему в цілому та негативно впливає на захист користувачів, в тому числі їх персональних даних. Традиційно параметри безпеки у соціальних мережах досліджуються за допомогою статистичних методів та узагальнених математичних залежностей і носять фрагментарний характер. Метою статті є розробка методики оцінювання параметрів безпеки персональних даних для мереж зі степеневим розподілом зв'язності вузлів на основі дослідження їх топологічних особливостей. Дослідження проведені на основі класичної моделі Барабаші-Альберта з застосуванням принципу переважного приєднання, який ставить ймовірність укладання нових зв'язків у залежність від кількості наявних зв'язків вузла. Більший степінь вузла означає більші можливості забирати собі нові зв'язки, що додаються в мережу. У якості основних параметрів безпеки розглядаються: степінь вузла, середня довжина шляху, ймовірність приєднання нових вузлів, коефіцієнт кластеризації, кореляція між степенями сусідніх вузлів. Показано, що збільшення степені вузла та довжини середнього шляху негативно впливає на захист персональних даних, оскільки при цьому збільшується ймовірність перехоплення інформації. Також, зі збільшенням коефіцієнта кластеризації зростає потік інформації, що приводить до зростання навантаження на систему захисту і негативно впливає на захист. Кореляція між степенями сусідніх вузлів впливає на перерозподіл потоків інформації і може, у залежності від степені вузлів, як негативно так і позитивно впливати на захист. Проведено моделювання для мереж різного масштабу та зроблено висновки щодо доцільності застосування методики.

Ключові слова: степенева соціальна мережа; персональні дані; принцип переважного приєднання; зв'язність мережі; степінь вузла; ймовірність приєднання вузла; коефіцієнт кластеризації; захист даних.

Вступ

У результаті поширення персональної інформації користувачів соціальними мережами у поєднанні з технологіями інформаційно-психологічного впливу на індивідуальну, колективну і масову свідомість, у суспільстві можуть мати місце прояви національної та релігійної ворожнечі, шантажу, булінгу, кіднепінгу, залякування, впливу на репутацію та довіру, думку користувачів, впливу на кар'єрне зростання, сімейні відносини та ін. Статистичні дані свідчать, що витік персональних даних через соціальні мережі збільшується з кожним роком. Відтак, захист персональних даних у соціальних мережах набуває все більшої актуальності.

Постановка проблеми

Традиційно, для опису соціальних мереж та оцінювання їх параметрів використовуються математичні залежності на основі графових моделей. Параметри безпеки у соціальних мережах досліджуються за допомогою статистичних методів і математичних підходів, які носять фрагментарний характер. Подальшим кроком дослідження соціальних мереж є застосування степеневих моделей, які дозволяють на основі отриманих математичних залежностей мати повну картину параметрів мережі та їх впливу на захист. Перевагами використання моделі мережі зі степеневим розподілом зв'язності вузлів є можливість збільшити рівень захищеності мережі шляхом вибору її раціональної структури, планувати захист інформації, в тому числі, безпеку персональних даних.

Аналіз публікацій

Загальний огляд моделей соціальних мереж запропоновано у роботі Мазуренка В.В. та Штовби С.Д. [1]. Моделюванню соціальних мереж, в тому числі степеневих, присвячено низку робіт Барабаші А.Л. та Альберта Р., зокрема у [2] розглядається статистична механіка складних мереж; у [3] – толерантність до атак та помилок, а також масштабування у мережах. У роботі [4] – параметри оцінювання великих мереж, у [5] – масштабування у випадкових мережах, у [6] описано безмасштабні характеристики випадкових мереж відносно топології мережі Інтернет. Крім того, у роботі Опсала Т., Агнессенса Ф. та Скворця

© В.А. Савченко, В.М. Ахрамович, М.В. Акулінічева, 2020

Дж. [7] розглядається вплив централізації у зважених мережах та пошук найкоротших шляхів передачі інформації. Роботи Савченка В.А. та Ахрамовича В.М. [8, 9] описують параметри соціальних мереж з графічним представленням даних, а стаття Підлазова А. та Щетініна Д. [9] – модель зростання соціальної мережі. Більш детально математичні залежності для степеневих соціальних мереж розкриваються у статті Ахрамовича В.М. [11], при цьому у [12] досліджується центральність мереж та вузлів.

Однак вказані дослідження не розкривають всі параметри степеневих соціальних мереж в повному обсязі і носять фрагментарний характер, що негативно впливає на формування процесу захисту персональних даних.

Мета статті полягає у розробці методики оцінювання параметрів мереж зі степеневим розподілом зв'язності вузлів, які впливають на безпеку персональних даних, на основі дослідження топологічних особливостей мережі.

Обгрунтування параметрів безпеки персональних даних у степеневих соціальних мережах.

Топологія степеневих соціальних мереж. Мережі зі степеневим розподілом зв'язності вузлів називаються безмасштабними (scale–free). Саме цей розподіл особливо часто спостерігається в реально існуючих складних мережах (Рис. 1) [11].

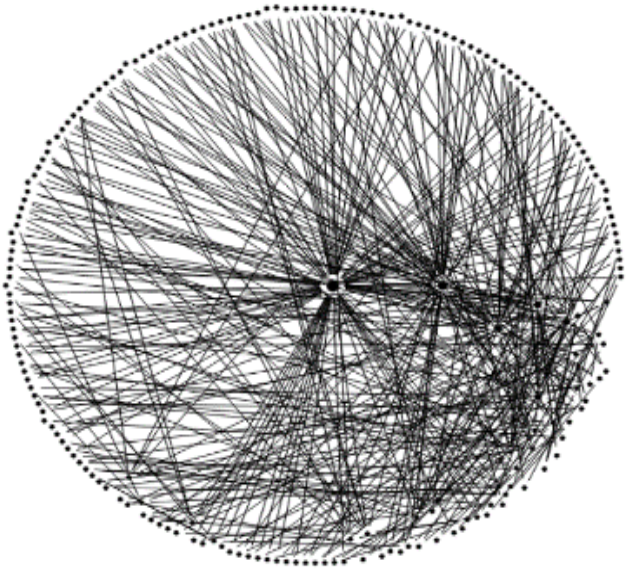


Рис. 1 - Степеневий граф

Дослідження [2, 3] свідчать, що 27% всіх з'єднань мають місце між всього 5% найбільших вузлів, 60% припадає на з'єднання інших 95% вузлів з 5% найбільших і тільки 13% – це з'єднання між вузлами, які не входять до лідируючих 5%. На Рис. 1 вузли з більшою валентністю розташовані в центрі, з меншою – на периферії.

Параметри мережі безпосередньо чи опосередковано впливають на захист інформації, в тому числі на безпеку персональних даних користувачів. При цьому, основними параметрами топології мережі з точки зору забезпечення безпеки персональних даних можуть бути: степінь вузла, середня довжина шляху, ймовірність приєднання нових вузлів, коефіцієнт кластеризації, кореляція між степенями сусідніх вузлів.

При розгляді моделей мереж виділяють два види степеневого розподілу [6, 11]. Перший з них пуасонівський розподіл

$$p_k = \frac{z^k e^{-z}}{k!}, \tag{1}$$

де z – це деяка константа, а k – степінь.

© *В.А. Савченко, В.М. Ахрамович, М.В. Акулінічева, 2020*

Для будь-якого розподілу в (1) виконується властивість $\sum_k p_k = 1$. Фактично, даний розподіл має вигляд: $p_k = \frac{C_1^k C_2}{k!}$, де C_1 і C_2 деякі константи.

Ця функція спадає експоненціально швидко. При $k = 1000$ p_k буде дуже малою величиною (тобто ймовірність мати 1000 і більше контактів буде дуже низькою), що, на практиці, не відповідає дійсності.

Другий важливий закон – це розподіл за степеневим законом (power-law)

$$p_k = \frac{k^{-\alpha}}{\xi(\alpha)}, \tag{2}$$

де α – константа, ζ(α) дзета-функція Рімана, яка служить для того, щоб виконувалося рівняння $\sum_k p_k = 1$. Дзета-функція Рімана у (2) визначена за допомогою ряду $\xi(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$.

Степеневому закону підпорядковується емпіричний розподіл степенів у багатьох реальних мережах. Зокрема, для Інтернету α = 2.5, а для мережі Facebook α = 2.2.

Модель Барабаші-Альберта – одна з декількох запропонованих моделей з степеневим розподілом, які генерують безмасштабні мережі [5, 10]. Вона включає в себе дві важливі спільні концепції: 1) зростання мережі; 2) принцип переважного приєднання. Обидві концепції широко представлені в мережах реального світу. Зростання означає, що число вузлів мережі збільшується з часом.

Принцип переважного приєднання полягає в тому, що чим більше зв’язків має вузол, тим більш переважно для нього створення нових зв’язків. Вузли з найбільшим степенем мають більше можливостей забирати собі зв'язки, що додаються в мережу [12]. Інтуїтивно, принцип переважного приєднання може бути зрозумілий, якщо ми думаємо в термінах соціальних мереж, які об'єднують людей [7]. Тут зв’язок від А до В означає, що людина А “знає” або “знайома” з людиною В. Сильно пов'язані вузли представлені відомими людьми з великим числом зв’язків. Коли новачок потрапляє в співтовариство, для нього/неї більш переважно зв'язатися з одним з відомих людей, ніж з відносно невідомим. Подібним чином у всесвітній мережі сторінки зв’язуються з хабами, наприклад, з добре відомими сайтами, як Google або Вікіпедія, ніж зі сторінками, які мало кому відомі. Якщо вибирати для зв’язку нову сторінку випадковим чином, то ймовірність вибору певної сторінки буде пропорційна її степеню. Це пояснює принцип переважного приєднання.

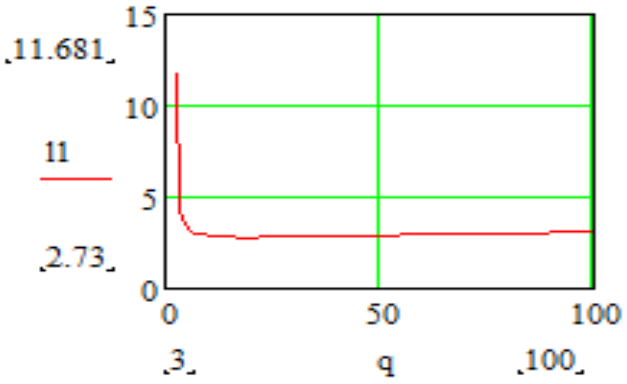
Середня довжина шляху. Довжина шляху в моделі Барабаші-Альберта збільшується в середньому, як логарифм розміру мережі (Рис. 2).

Збільшення довжини середнього шляху негативно впливає на захист, оскільки збільшується ймовірність, наприклад, перехоплення інформації. Точна форма має подвійну логарифмічну поправку [4] і виглядає, як

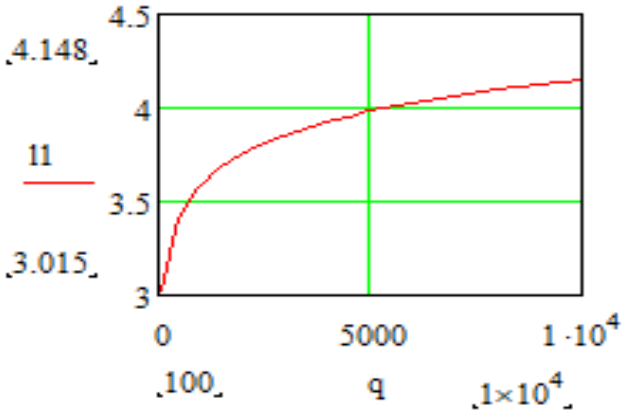
$$l \sim \frac{\ln N}{\ln \ln N}. \tag{3}$$

Модель Барабаші-Альберта має систематично коротший середній шлях (3), ніж випадковий граф. Пік кривої зліва на графіку (Рис. 2, а) пояснюється відсутністю початкового з’єднання в мережі, коли кількість вузлів дорівнює 0. По мірі збільшення кількості вузлів, які приєднуються до мережі, крива поступово стабілізується. Графіки на Рис. 2 підтверджують висновок Мілграма [4], що середня довжина шляху не перевищує шести.

© *В.А. Савченко, В.М. Ахрамович, М.В. Акулінічева, 2020*



а) q = 0...100



б) q = 100...10^4

Рис. 2 – Залежність середньої довжини шляху від кількості вузлів: l1– середня довжина шляху, q – кількість вузлів в мережі

Ймовірність приєднання нових вузлів. Додавання нових вузлів в мережу негативно впливає на захищеність мережі, оскільки збільшує навантаження на неї. Новий вузол може додатися у мережу у довільний момент часу. Кожен такий вузол з'єднується з існуючими вузлами з ймовірністю, пропорційною до числа зв'язків цих вузлів. Формально, ймовірність того, що новий вузол з'єднується з вузлом *i* дорівнює

$$p_i = \frac{k_i}{\sum_j k_j}, \qquad (4)$$

де k_i – степінь *i*-го вузла, а в знаменнику підсумовуються степені всіх існуючих вузлів.

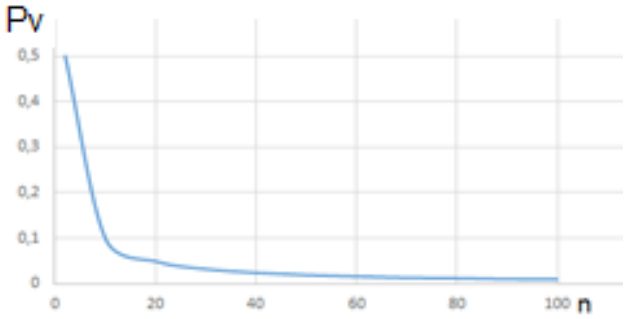
За (4) найбільш пов'язані вузли ("хаби"), як правило, накопичують найбільше зв'язків, тоді як вузли з незначним числом зв'язків навряд чи будуть обрані для приєднання нових вузлів. Нові вузли мають "перевагу" з'єднуватися з найбільш пов'язаними вузлами.

При випадковому приєднанні ймовірність того, що зв'язок буде створений з даної вершиною, не залежить від її характеристик і дорівнює $P_v = \frac{1}{n}$, де *n* – загальне число вершин графа у певний момент часу (Рис. 3).

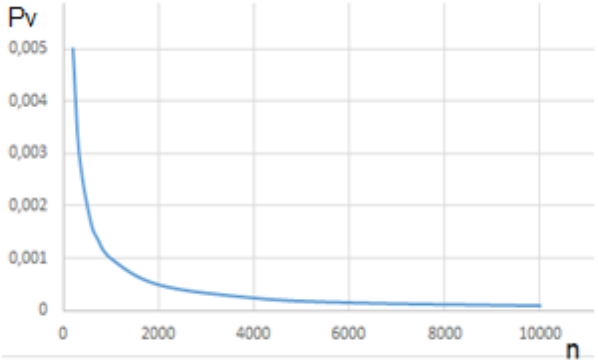
© *В.А. Савченко, В.М. Ахрамович, М.В. Акулінічева, 2020*

При переважному приєднанні ймовірність утворення зв'язку з вершиною i, яка вже має x_i зв'язків, може бути подана відношенням $x_i/S(t)$, де сумарна валентність $S(t)=\sum_{i=1}^{n(t)} x_i(t)$ дорівнює подвоєному числу зв'язків графа в момент часу t (Рис. 4).

Степеневий розподіл в моделі Барабаші-Альберта є безмасштабним, точніше підпорядковується степеневому закону: $P(k) \sim k^{-3}$. Зі збільшенням валентності вузлів зростає кількість нових приєднань та збільшується трафік у мережі, що призводить до зростання навантаження на систему та негативно позначається на захисті, у т.ч. персональних даних.



а) n = 0...100



б) n = 100...10^4

Рис. 3 – Ймовірність створення нового зв'язку з довільною вершиною

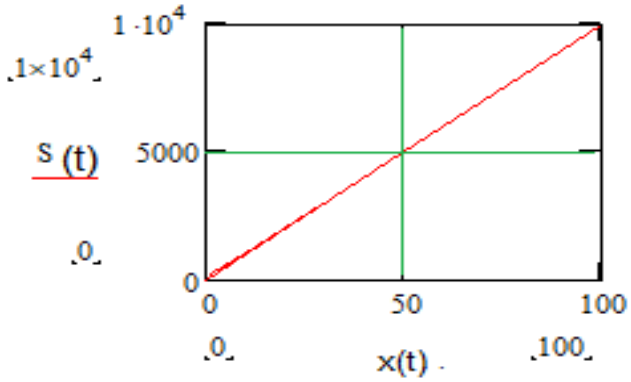


Рис. 4 – Сумарна валентність

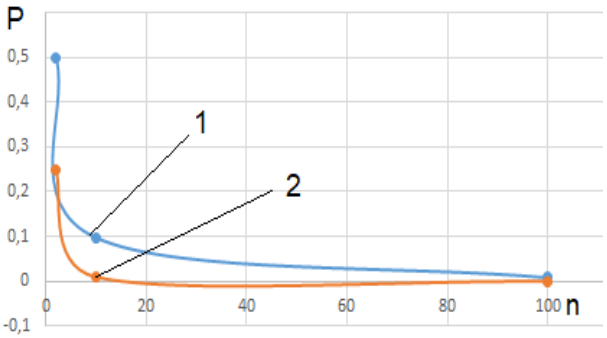
Алгоритм Барабаші-Альберта дозволяє генерувати мережі зі степеневим розподілом, однак ці мережі занадто формальні, в них не має змістовної складової, зокрема, їм не властивий феномен "клубу багатих", характерний для багатьох реальних мереж, наприклад, ASmallWorld, Beautifulpeople, HotEnough та ін. Такі мережі створювалися для збільшення

© *В.А. Савченко, В.М. Ахрамович, М.В. Акулінічева, 2020*

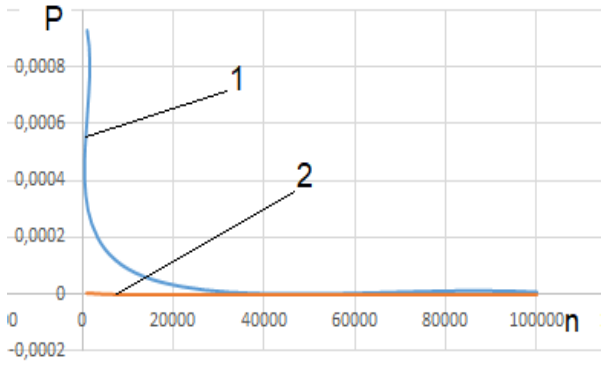
рівня безпеки персональних даних через неможливість попадання туди без рекомендації та наявності значних коштів на захист.

Коефіцієнт кластеризації. Коефіцієнт кластеризації визначається як відносна кількість вузлів *P*, які мають задану кількість зв'язків *n*. Істотно уявити, що переважна більшість користувачів соціальних мереж мають незначне коло зв'язків, у той час як обмежена кількість людей володіє достатньо значною їх кількістю.

На теперішній час для моделі Барабаші-Альберта не існує аналітичних значень коефіцієнта кластеризації. Разом з тим, коефіцієнти кластеризації, отримані емпіричним шляхом [12], в загальному випадку для переважного приєднання є значно вищими, ніж для випадкових мереж. Коефіцієнт кластеризації залежить від розміру мережі і для великих соціальних мереж (Facebook, Instagram) наближено відповідає степеневому закону $P \sim N^{-0.75}$. Для мереж меншого масштабу типовим є степеневий розподіл вершин [2, 3, 12] (Рис. 5) за валентністю з щільністю ймовірності виду $P \sim n^{-(1+\alpha)}$ (Рис. 5).



а) n = 0...100



а) n = 100...10^5

Рис. 5 – Степеневий розподіл вершин за валентністю з α = 0.1 (крива 1) та α = 2.0 (крива 2)

Для соціальних мереж показник розподілу $1 \prec \alpha \prec 2$, хоча для мереж іншої природи він може бути і більшим. Ця поведінка все ж відрізняється від поведінки малих мереж, де кластеризація не залежить від розміру мережі. У разі ієрархічних мереж, кластеризація як функція степеня вузла також підпорядковується степеневому закону: $C(k) = k^{-1}$.

Для безмасштабних мереж значення коефіцієнта кластеризації значно вище, ніж для випадкових мереж такого ж розміру [3]. Зі збільшенням коефіцієнта кластеризації зростає потік інформації, що приводить до зростання навантаження на систему захисту і негативно впливає на захист.

Кореляція між степенями сусідніх вузлів. Значна кількість структурних і динамічних властивостей мережі визначається за допомогою оцінки кореляції між ступенями сусідніх вузлів. Така кореляція може бути виражена через сукупний розподіл $P(k,k')$, тобто як

© *В.А. Савченко, В.М. Ахрамович, М.В. Акулінічева, 2020*

ймовірність того, що довільно вибране ребро з'єднує вузол ступеня k з вузлом ступеня k'. Середню ступінь найближчих сусідів вузлів із заданим ступенем k обчислюють за формулою

$$S(k) = \sum_{k} k' P(k' / k). \quad (5)$$

Інтерес у (5) представляє зміна величини $S(k)$ при зростанні степеня розглянутої вершини k (Рис. 6). Якщо середній степінь найближчих сусідів зростає, то це означає, що вершини з високим степенем (хаби) пов'язані переважно між собою, і, навпаки, вершини з низьким степенем зв'язуються переважно з вершинами низького ступеня – так зване асортативне змішування [11].

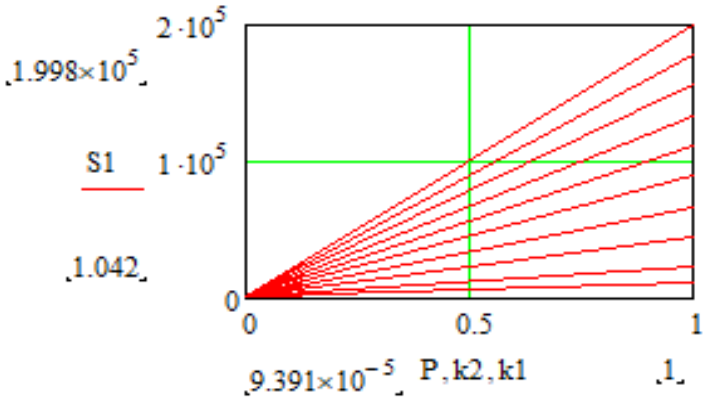


Рис. 6 – Залежність середнього степеня від кореляції між степенями сусідніх вузлів: $k_1 = (2,4,22)$, $k_2 = (100,200,1800)$

Отже, як видно з Рис. 6, кореляція між степенями сусідніх вузлів лінійно впливає на середню степінь найближчих сусідів вузлів із заданим степенем k. Також, кореляція відіграє важливу роль в перерозподілі потоків інформації і може, як негативно так і позитивно впливати на захист. У випадку вирівнювання степенів вузлів спостерігається, як правило, позитивний вплив, оскільки це означає, що мережа базується на загальних принципах і застосовує подібні засоби захисту. В іншому випадку, коли спостерігається достатньо слабка кореляція між степенями сусідніх вузлів – негативно, оскільки це свідчить про наявність різнорідної категорії учасників мережі, серед яких достатньо складно організувати ефективний та захист персональних даних.

Висновки.

Витік персональних даних у соціальних мережах щороку збільшується через що зазначена проблема стає дедалі більш актуальною. На сьогоднішній день захист інформації користувачів полягає не лише у застосуванні певних технічних рішень, а й у побудові безпечної топології мережі. Дослідження соціальних мереж на основі степеневих моделей є подальшим розширенням класичних підходів, які дозволять на основі отриманих математичних залежностей мати повну картину параметрів мережі та їх впливу на захист персональних даних користувачів.

Застосування принципу переважного приєднання, який визначає, що вузли з найбільшим степенем мають більше можливостей забирати собі зв'язки, що додаються в мережу, дозволяє створити адекватну модель для оцінки топологічних характеристик захисту мережі. Значна кількість структурних і динамічних властивостей мережі визначається за допомогою оцінки кореляції між ступенями сусідніх вузлів.

© *В.А. Савченко, В.М. Ахрамович, М.В. Акулінічева, 2020*

Оцінювання валентності вузлів мережі, ймовірності приєднання нових вузлів, середньої довжини шляху та коефіцієнта кластеризації дає можливість в цілому прогнозувати зростання навантаження мережу та її систему захисту для застосування у подальшому організаційних та технічних заходів безпеки.

Напрямом подальших досліджень є комплексне дослідження параметрів соціальних мереж в аспекті їх впливу на безпеку персональних даних користувачів та побудова загальної моделі безпеки соціальних мереж.

Перелік посилань

1. В. В. Мазуренко, С. Д. Штовба, "Огляд моделей аналізу соціальних мереж", Вісник Вінницького політехнічного інституту, № 2, сс. 62-74, 2015.
2. R. Albert, H. Jeong, A. Barabasi, "Attack and error tolerance of complex networks", Nature, Vol. 406, pp. 378-382, 2000.
3. A. L. Barabasi, R. Albert, "Emergence of scaling in random networks", Science journal, Vol. 286, no. 5439, pp. 509-512, 1999.
4. A. L. Barabasi, Linked: The New Science of Networks, N.Y.: Perseus Book Group, 2002.
5. A. L. Barabasi, R. Albert, "Emergence of scaling in random networks", Science, Vol. 286, pp. 509-512, 1999.
6. A. L. Barabasi, R. Albert, H. Jeong, "Scale-free characteristics of random networks: The topology of the world–wide web", Physica A. Vol. 281, pp. 69-77, 2000.
7. T. Opsahl, F. Agneessens, J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths", Social Networks Journal, Vol. 32(3), pp. 245-251, 2010.
8. P. Shchypanskyi, V. Savchenko, V. Akhramovych, T. Muzshanova, S. Lehominova, V. Chegrenets, "The Model of Secure Social Networks Activity Based on Graph Theory", International Journal of Innovative Technology and Exploring Engineering, Vol. 9, № 4, pp. 1803-1810, 2020.
9. V. Savchenko, V. Akhramovych, A. Tushych, I. Sribna, I. Vlasov, "Analysis of Social Network Parameters and the Likelihood of its Construction", International Journal of Emerging Trends in Engineering Research, Vol. 8, No. 2, pp. 271-276, 2020.
10. A. Pidlazov, D. Shchetinina, "Social network growth model", Preprints M. Keldysh`s IAM, № 95, pp. 16, 2013.
11. В. М. Ахрамович, "Степеневі соціальні мережі", Colloquium journal (Warszawa, Polska), №5 (57), pp. 27-29, 2020.
12. В. М. Ахрамович, О. С. Стефурак, "Центральність соціальних мереж". Colloquium–journal (Warszawa, Polska), № 6(58), pp. 20-22, 2020.

Надійшла: 10.06.2020 Рецензент: д.т.н., професор Вишнівський В.В.

© В.А. Савченко, В.М. Ахрамович, М.В. Акулінічева, 2020