

О.О. Нечасів

Інститут спеціального зв'язку та захисту інформації
Національного технічного університету України "КПІ"
вул. Московська, 45/1, 01011, Київ, Україна

Д.В. Ланде

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

Відновлюваність зв'язків у безмасштабних мережах

Розглянуто питання відновлювання зв'язків в мережах після їх вилучень шляхом деструктивного впливу. Наведені результати дослідження для мереж, отриманих за трьома алгоритмами: Барабаши-Альберт; перекидання зв'язків; реальній мережі понять та інформаційних зв'язків між ними у веб-публікаціях. Встановлено, що за допомогою розглянутого алгоритму у реальній мережі відновлюється приблизно 50% зв'язків; у штучних мережах їх кількість становить близько 25%.

Ключові слова: складні мережі, латентні зв'язки, уразливість, моделювання, відновлюваність, безмасштабні мережі

На сьогоднішній завдяки сучасним технологіям можливо не тільки досліджувати мережеві структури будь-якої природи, але й аналізувати їх, прогнозувати динаміку таких мереж в майбутньому та моделювати зміни їх структур при дії зовнішніх впливів. Існує є два найпростіших види деструктивного впливу на мережі ззовні: видалення зв'язків між двома вузлами, або видалення вузлів з усіма відповідними зв'язками, що вони містять.

На цей час виявлено багато з реальних мереж, що мають таку властивість, як безмасштабність [1]. Безмасштабною називають мережу із степеневим розподілом ступенів вузлів. Для такої мережі існує багато вузлів з маленькою кількістю зв'язків та невелика кількість так званих вузлів-концентраторів або хабів, які займають найбільш важливе положення у структурі мережі. Наприклад, в економіці практично у кожній галузі існує одна або декілька найбільш сильних компаній, що мають вплив на велику кількість компаній більш слабких. Так звана «пірамідальна структура» ієрархії вузлів в мережі як правило виражається у степеневим розподілом її вузлів.

Відомо, що цільність мережі не завжди залежить від вузлів-концентраторів. Завдяки відомим в концепції складних мереж (complex networks) [2] методам можливо виявити вузли-посередники, що мають меншу кількість зв'язків, але при видаленні яких, мережа швидко втрачає зв'язність. Експериментально доведено, що реальні безмасштабні мережі втрачають зв'язність при випадковому видаленні близько 15% вузлів [3]. У той же час, кожна мережа є системою динамічною, та вона реагує на деструктивні впливи відновленням старих або формуванням нових зв'язків, створенням нових вузлів. При видаленні вузла-концентратора, його роль бере на себе інший вузол, який раніше, можливо, займав другорядну роль у всій мережі.

Процес розвитку та відновлення мережі підпорядковується статистичним закономірностям, дослідження яких є актуальною проблемою в теорії складних мереж, оскільки знаючи методи моделювання таких процесів, можливо маніпулювати розвитком мережі, та спостерігати, забезпечувати такий стан, у якому кожний вузол буде відігравати призначену йому роль.

Одним з напрямів досліджень складних мереж є виявлення вузлів мережі, видалення яких призводить до її найшвидшого руйнування. Ці вузли, природно, найкритичніші в усій мережевій структурі. Для цього можна використовувати методи виявлення латентних зв'язків [4], які, наприклад, можуть базуватися на обчисленні матриці ймовірності зв'язків між вузлами.

Метод розрахунку нечіткої ймовірності зв'язків

Представимо мережу в якій всі існуючі зв'язки представлені ймовірностями їх існування, які можуть бути знайдені за певними змістовними ознаками, або отримані експертним шляхом. У цьому випадку мережа може бути представлена у вигляді так званої «матриці неявних зв'язків», елементами якої є ймовірність зв'язку між об'єктами.

Основа методу, що розглядається, – обчислення ймовірностей не тільки між двома вузлами безпосередньо, але й урахування ймовірностей зв'язку через треті, четверті і т. д. вузли. Розглянемо для пояснення методу найбільш простий випадок, мережу, що складається з трьох вузлів – V_1, V_2, V_3 . Проста розрахунку ймовірності зв'язку між об'єктами при наявності певних ймовірностей їх непрямого з'єднання $p_{1,2}, p_{1,3}, p_{2,3}$:

$$p_{1,2} = 1 - (1 - p_{1,2})(1 - p_{1,3}p_{2,3}). \quad (1)$$

У загальному випадку якщо шляхи між вузлами i та j можливі через декілька вузлів, то загальна формула ймовірності зв'язку між ними буде мати такий вигляд:

$$p_{i,j} = 1 - (1 - p_{i,j}) \prod_{k \neq i,j} (1 - p_{i,k} p_{k,j}) \prod_{k \neq i \neq l \neq j} (1 - p_{i,k} p_{k,l} p_{l,j}) \dots \quad (2)$$

В результаті розрахунку такої формулою можливо знайти матрицю нечітких зв'язків через декілька вузлів, де вірогідність з'єднання може коливатись від нуля до одиниці. Ця матриця досить природно відображає стан в реальних мережах [4-5] і може застосовуватися при пошуку латентних, або відновленні старих зв'язків.

Здатність мережі до відновлення

Мережі мають різну структуру і відповідно мають різну здатність до відновлення. В реальних мережах за підходом, що пропонується, відновленню підлягають ймовірності зв'язків після їх вилучення. У подальшому будемо враховувати всі зв'язки в мережі, що можуть бути відновлені навіть з невеликими значеннями ймовірності, оскільки коли реальна мережа доведена до критичної точки (втрати зв'язності), для відновлення будуть використовуватись усі можливі ресурси.

Отже аналіз мережі на предмет її здатності до відновлення будемо розглядати як сукупність ймовірностей відновлення кожного окремого зв'язку мережі.

Відновлення мереж реальних та штучних

При прогнозування розвитку мережі, та її реакції на зовнішні деструктивні впливи, часто слід заздалегідь знати з певною точністю яка буде топологія мережі, як вона буде розвиватися. Для таких цілей використовується моделювання мереж. Існує декілька основних моделей побудови мереж, які мають різні властивості. Моделювання мереж з властивостями найбільш приближеними до реальних є актуальним напрямом дослідження в теорії складних мереж, оскільки виникнення зв'язків залежить від сукупності явних та скритих факторів, які на цей час складно описати одними лише математичними виразами.

Тому, з метою оцінки наближеності властивостей мереж штучних до мереж реальних, було проведено наступне дослідження:

- побудовано моделі безмасштабних мереж, властивості яких на даний час признані найбільш відповідними реальним властивостям, та створено мережу персон на основі реальних даних;
- побудовані мережі проаналізовано базовими методами, що застосовуються в теорії складних мереж; знайдено основні характеристики цих мереж;
- мережі проаналізовано щодо їхньої здатності до відновлення.

Мережа Барабаші-Альберт

У відповідності з принципом побудови мережі за Ласло Барабашем і Рекою Альберт мережа у початковий момент розглядається з кількістю вузлів m_0 із зв'язками між ними.

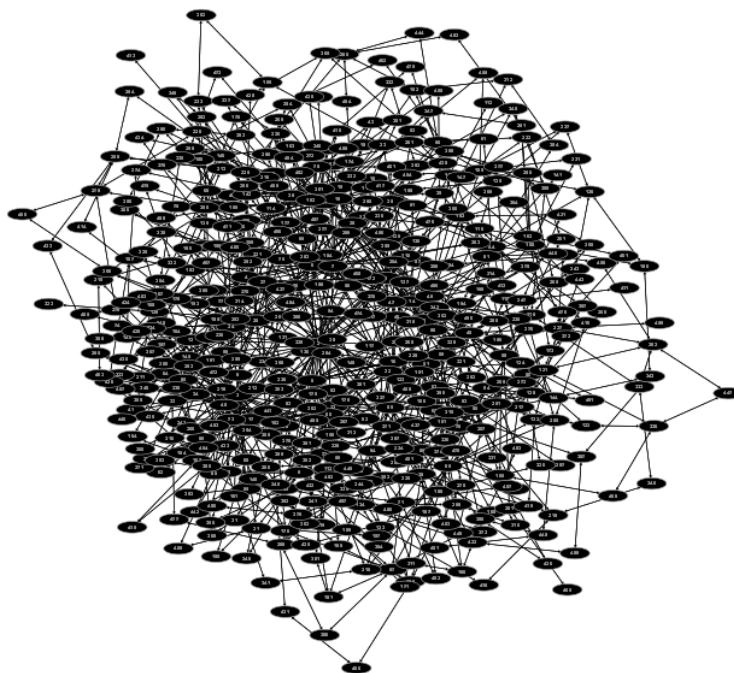


Рис. 1 – Штучна мережа побудована на основі методу Барабаші-Альберт

Було показано, що для виникнення безмасштабної мережі достатньо виконання двох вимог її розвитку [6] (рис. 1):

1. *Зростання*. Починаючи з невеликої кількості вузлів, на кожному часовому кроці додається один новий вузол із зв'язками, які з'єднують цей новий вузол з різними вузлами, що вже існували m ($m < m_0$);

2. *Переважає приєднання*. Коли обираються вузли, до яких приєднується новий вузол, передбачається, що ймовірність Π , з якою новий вузол буде з'єднуватися з вже існуючим вузлом i , залежить від кількості зв'язків q_i , якими цей вузол зв'язаний з іншими вузлами, так що $\Pi(q_i) = q_i / \sum_j q_j$.

Модель перекидання зв'язків

Принцип побудови моделі мережі методом перекидання відноситься до мережі з фіксованою кількістю вузлів і полягає у наступному:

Якщо розподіл випадкової величини p_k задано, то процедура формування безмасштабної мережі зводиться до наступних операцій [7]:

- генератором псевдовипадкових чисел формується послідовність чисел, яка відповідає степеневому розподілу. Обирається N чисел k_i , де $i = 1, \dots, N$;
- кожній вершині i мережі присвоюється k_i «заготовок» (кінців) для майбутніх ребер;
- із степеневі послідовності випадково виділяються пари «заготовок». Вони з'єднуються в тому випадку, якщо нове ребро не призведе до появи ребер-циклів (петель) або мультиребер;
- попередній крок виконується доти, поки залишаються «заготовки», що не використані.

На рис. 2 наведено приклад мережі, побудованої за даним алгоритмом. Перевагою даного алгоритму є його універсальність, так як з його допомогою можливо побудувати мережі з будь-яким розподілом степенів вузлів.

Мережа персон з новинного контенту мережі Інтернет

Мережа прізвищ персон формується на базі змістовного аналізу масиву електронних публікацій, які було отримано системою контент-моніторингу Infostream [8]. Ребра мережі було обрано наступним чином: якщо у веб-публікації в одному реченні з'являлося два прізвища (вузла), то вважалось що ці персони зв'язані інформаційним зв'язком, тобто до ваги ребра між ними додавалася одиниця. Побудовану мережу зв'язків відображено на рис. 3.

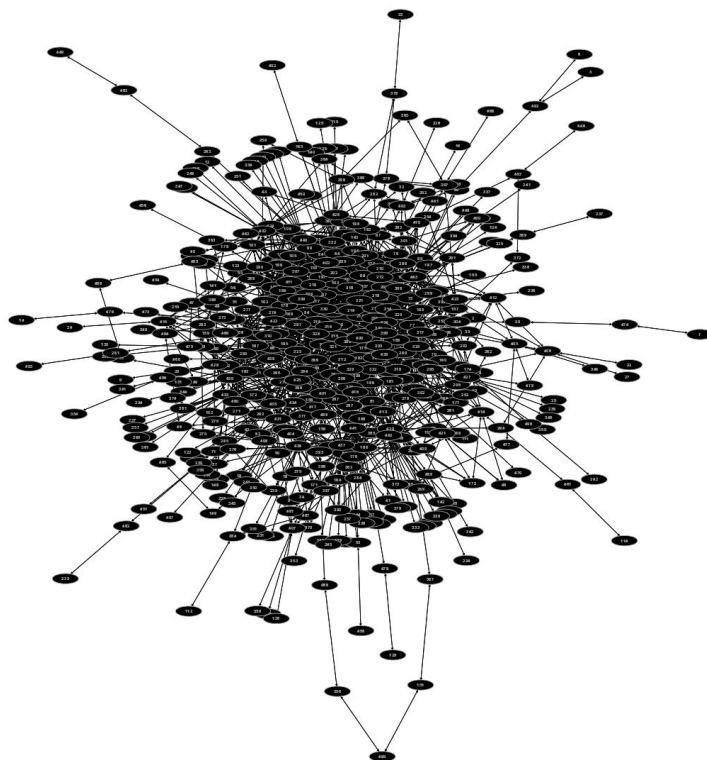


Рис. 2 – Штучна модель з заданим степеневим розподілом

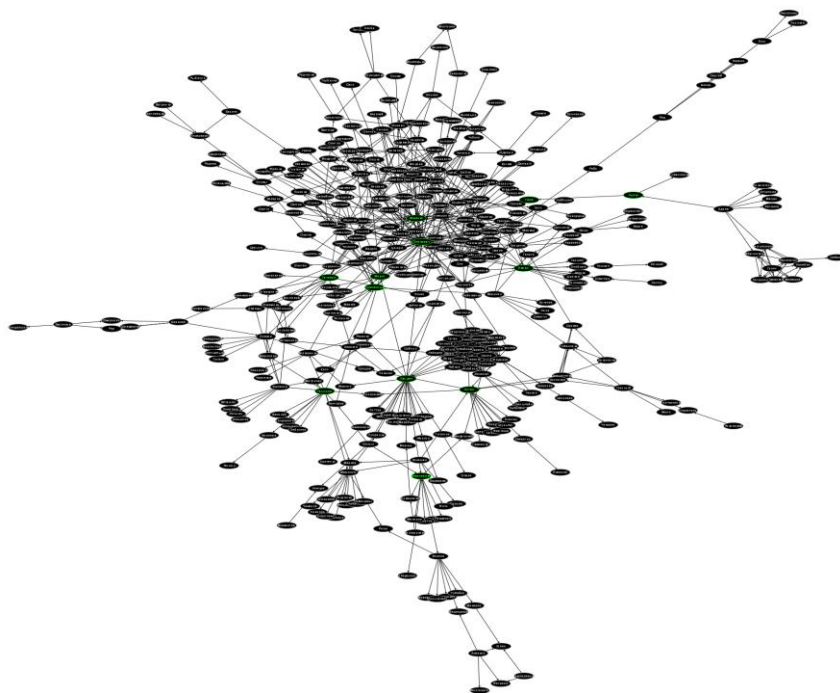


Рис. 3 – Мережа зв'язків персон (реальна мережа), побудована за допомогою системи контент-моніторингу InfoStream.

Дослідження відновлення зв'язків в мережах

Отже, для дослідження були використані три мережі: реальна мережа персон і мережі, побудовані у відповідності з моделями: із заданим степеневим розподілом та

Барабаші-Альберт. Кількість вузлів для всіх мереж однакова, дорівнює 500. На рис. 4 показані розподіли ступенів вузлів кожної з побудованих мереж.

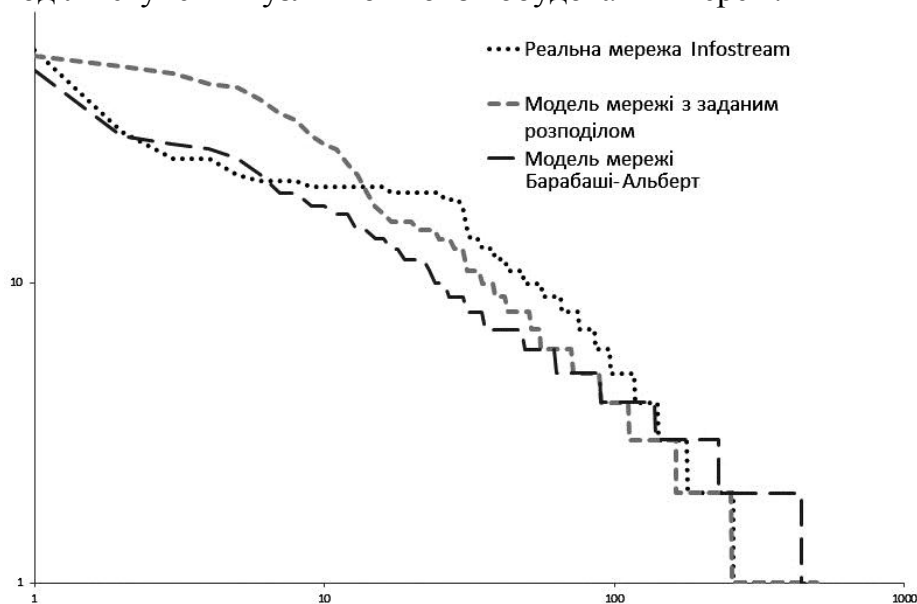


Рис. 4 – Розподіли ступенів мереж, що досліджуються

З рис. 4 видно що розподіли досить подібні, отже можливо провести порівняльний аналіз на предмет відновлення зв'язків та зробити висновок щодо доцільності використання розглянутих моделей при моделюванні відновлювальних процесів.

Дослідження мережі щодо відновлення зв'язків виконувалось за таким алгоритмом:

- 1) Видалявся i -й зв'язок між вузлами;
- 2) За формулою (2) обчислювалася вірогідність i -го зв'язку через треті, четверті і т.п. вузли (тобто, у деякому розумінні – відновлювання);
- 3) Зберігання даних щодо ймовірності відновлення розглянутого зв'язку та циклічний перехід до першого шагу, поки не перебрано всіх зв'язків.

Для кожної із розглянутих мереж за описаним вище алгоритмом були побудовані лінії ймовірностей відновлення кожного зв'язку відповідної мережі, які представлені на рис. 5.



Рис. 5 – Ймовірності відновлення зв'язків різних мереж

На рис. 5 показані три лінії, що відображають аналіз одного з видів розглянутих мереж. Значення кожної з ліній – це сукупність ймовірностей відновлення кожного зв'язку, відсортована в порядку зростання. Як можна бачити, доля відновлення зв'язків у різних моделях різна. Якщо у реальній мережі відновлюється близько 50% зв'язків, то у штучних мережах їх число близько 25%.

Загальну стійкість мережі до атак можливо розглядати як середнє арифметичне від ймовірностей відновлення кожного окремого зв'язку мережі, тобто:

$$V = \frac{\sum_{i=1}^n p_i}{n} \quad (3)$$

де p_i – ймовірність відновлення i -го зв'язку, n – загальна кількість зв'язків.

Звідси підставивши формулу (2) у, отримаємо загальну формулу розрахунку сукупності ймовірностей відновлюваності мережі:

$$V = \frac{\sum \left(1 - (1 - p_{i,j}) \prod_{k \neq i,j} (1 - p_{i,k} p_{k,j}) \prod_{k \neq i \neq l \neq j} (1 - p_{i,k} p_{k,l} p_{l,j}) \dots \right)}{n} \quad (4)$$

Розрахувавши ймовірності відновлення всіх окремих вузлів, отримаємо:

- 1) графік ймовірностей, що відображаються загальну сукупність вузлів здатних та не здатних до відновлення (рис. 5);
- 2) кількість членів груп зв'язків з однаковою ймовірністю відновлення;
- 3) стійкість мережі до атак як середнє арифметичне всіх ймовірностей відновлення зв'язків.

Даний метод дозволяє аналізувати мережі на їх відновлюваність, результати чого можливо застосовувати у порівнянні мереж різних типів, що допоможе при прогнозуванні розвитку мереж, аналізі їх стійкості до атак, допоможе зробити більш точний вибір моделей для їх максимального наближення до реальності.

Цей метод розширяє можливості оцінки якості моделей мереж, може використовуватись при аналізі та оцінці мережевих структур в таких галузях як маркетинг, конкурентна розвідка, економіка, інформаційно-мережеві технології.

Література

1. Newman M.E.J., Watts D.J. Scaling and percolation in the small-world network model // Phys. Rev. E, 1999. – № 60. – P. 7332-7342.
2. Newman M.E.J. The structure and function of complex networks // SIAM Review, 2003. – 45. – P. 167-256.
3. Clauset A., Moore C., Newman M. E. G. Hierarchical structure and the prediction of missing links in networks // Nature. – 2000. – № 453. – P. 98–101.
4. Додонов О.Г., Ланде Д.В. Імовірнісна модель виявлення латентних зв'язків у мережах понять// Реєстрація, зберігання і обробка даних, 2011. – № 2, – 13. – С. 38-46.
5. Додонов А.Г., Ландэ Д.В. Факторы сетевой мобилизации // Материалы XII Международной научно-практической конференции «Информационная безопасность». – Ч. 1. – Таганрог: Изд-во ТТИ ЮФИ, 2012. – С. 19-27.

6. Albert R., Barabasi A.-L. Statistical mechanics of complex networks, 2002.– Rev. Mod. Phys. – **74**. – P. 47; Preprint Arxiv: cond-mat/0106096.
7. Ландэ Д.В. Моделирование контентных сетей // Проблемы інформатизації та управління: Збірник наукових праць: Випуск 1(37). – К.: НАУ, 2012. – С. 78-84.
8. Григорьев А.Н., Ландэ Д.В., Бороденков С.А., Мазуркевич Р.В., Пацьора В.Н. InfoStream. Мониторинг новостей из Интернет: технология, система, сервис: научно-методическое пособие. – Киев: Старт-98, 2007. – 40 с.