Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Навчально-науковий центр заочної форми навчання
(повна назва)
Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти перший (бакалаврський)

Розробка моделей безмасштабних мереж зі сталим показником гомофільності спільнот
(тема)

Виконав:
здобувач четвертого року навчання,
групи ІТШІз-22-1
Валерій ГРИНЕНКО
(власне ім'я, прізвище)
Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)
Тип програми освітньо-професійна
Освітня програма Штучний інтелект
(повна назва освітньої програми)
Керівник ас. Максим ПОЛІЇТ
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ШІ _____ Лариса ЧАЛА
(підпис) (власне ім'я, прізвище)

2026 р.

Харківський національний університет радіоелектроніки

Факультет Навчально-науковий центр заочної форми навчання

Кафедра Штучного інтелекту

Рівень вищої освіти перший (бакалаврський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-професійна

Освітня програма Штучний інтелект
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«_____» _____ 20 ___ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Гриненку Валерію Владиславовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Розробка моделей безмасштабних мереж зі сталим показником гомофільності спільнот

затверджена наказом університету від 18 травня 20 26 р. № 76Стз

2. Термін подання здобувачем роботи до екзаменаційної комісії 14 липня 20 26 р.

3. Вихідні дані до роботи Науково-технічні публікації, дані інтернет-джерел щодо безмасштабних мереж та їх моделювання

4. Перелік питань, що потрібно опрацювати в роботі
1) Огляд предметної галузі
2) Розробка моделі безмасштабних мереж з гомофільною структурою спільнот
3) Програмна реалізація розробленої моделі, проведення чисельного експерименту

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / термін виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	18.05.2026	виконано
2	Аналіз предметної галузі	25.05.2026	виконано
3	Постановка завдання та узгодження з керівником	27.05.2026	виконано
4	Дослідження видів, класів та моделей	31.05.2026	виконано
	безмасштабних мереж		
5	Дослідження поняття гомофільності мереж,	05.06.2026	виконано
	модулярності як міри гомофільності		
6	Програмна реалізація моделі безмасштабної	18.06.2026	виконано
	мережі з екстремальною гомофільністю спільнот		
7	Написання пояснювальної записки	29.06.2026	виконано
8	Попередній захист	10.07.2026	виконано
9	Захист перед ЕК	15.07.2026	

Дата видачі завдання 18 травня 2026 р.

Здобувач

(підпис)

Керівник роботи _____ ас. Максим ПОЛІЇТ

(підпис) (посада, власне ім’я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 68 с., 13 рис., 1 табл., 2 дод., 18 джерел.

БЕЗМАСШТАБНІ МЕРЕЖІ, ГЕНЕРАТИВНІ МОДЕЛІ, ГОМОФІЛЬНІСТЬ, МОДУЛЯРНІСТЬ, ПЕРЕВАЖНЕ ПРИЄДНАННЯ, СПІЛЬНОТИ ВУЗЛІВ.

Об’єктом досліджень є безмасштабні мережі.

Предметом досліджень кваліфікаційної роботи є моделі безмасштабних мереж з гомофільною структурою спільнот.

Метою роботи є моделювання структурованих безмасштабних мереж з екстремально високою ступінню гомофільності спільнот.

ABSTRACT

Bachelor’s thesis contains: 68 pp., 13 fig., 1 tabl., 2 ann., 18 references.

GENERATIVE NETWORK MODELS, HOMOPHILY, MODULARITY, NODE COMMUNITIES, PREFERENTIAL ATTACHMENT, SCALE-FREE NETWORKS.

The object of research is scale-free networks.

The subject of research are models of scale-free networks with a homophilic community structure.

The aim of the work is to model structured scale-free networks with an extremely high degree of community homophily.

ЗМІСТ

Вступ 7
1 Огляд предметної галузі 8
1.1 Складні мережі, їх властивості 8
1.2 Моделі складних мереж 13
1.3 Поняття безмасштабності мереж, моделі безмасштабних мереж 23
1.4 Постановка задач досліджень 29
2 Розробка моделі безмасштабних мереж з гомофільною структурою спільнот 31
2.1 Модулярність як міра гомофільності мереж 31
2.2 Розробка генеративної моделі безмасштабних мереж 36
2.3 Розробка генеративної моделі гомофільних мереж 39
2.4 Оцінка модулярності модельованої мережі 47
3 Програмна реалізація розробленої моделі, проведення чисельного експерименту 51
3.1 Інструменти програмної реалізації моделі 51
3.2 Моделювання гомофільних безмасштабних мереж 54
Висновки 58
Перелік джерел посилання 59
Додаток А Тексти програм 62
Додаток Б Відомість кваліфікаційної роботи 68

ВСТУП

У 90-х роках XX століття, коли інтернет тільки зароджувався, дослідники вже порушили питання, яким законам підпорядковується зростання інтернету і яка найбільш адекватна модель для опису властивостей цієї мережі. Серед перших тут були А.-Л. Барабаші та Р. Альберт. Вони знайшли ряд важливих емпіричних закономірностей у поведінці інтернету і на їх основі розробили модель, яку згодом по-різному формалізували багато авторів.

Теорія складних мереж – це комплексний науковий напрямок, що знаходиться на стику таких наук, як дискретна математика, теорія графів, теорія алгоритмів, нелінійна динаміка, теорія фазових переходів, теорія фракталів, перколяційні процеси, тощо. Тому для успішного моделювання складних мереж необхідні базові відомості з усіх цих галузей.

У теорії складних мереж виділяють три основних напрями: дослідження статистичних властивостей, які характеризують поведінку реальних мереж; створення моделей мереж; передбачення поведінки мереж у разі зміни структурних властивостей.

Тематика досліджень, які провадяться у поточній кваліфікаційній роботі, зосереджена на другому з перелічених напрямків, а саме на моделюванні складних мереж. Найважливішим класом таких моделей є генеративні моделі. Це такі, які дозволяють моделювати зростаючу мережу покроково; він маленького зародку, два вузла, одне ребро, до мереж розміром у сотні тисяч, чи мільйони.

Одними з важливих характеристик мереж є показник скейлінгу та показник еластичності (логарифмічної щільності).

Отже, метою роботи є моделювання зростаючих безмасштабних мереж з можливістю незалежної параметризації показників скейлінгу та еластичності мережі.

1 ОГЛЯД ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Складні мережі, їх властивості

Складні мережі (complex networks) це мережі реального світу, які описуються графовими моделями, але, на відміну від об'єктів традиційної теорії графів, мають характерні властивості:

– великий розмір. Якщо у звичайній теорії графів розглядаються графи з одиницями-десятками вузлів, то мережі реального світу можуть мати розмір в сотні мільйонів вузлів;

– динамічність. Складні мережі постійно змінюються, зазвичай, зростають. Додаються нові вузли, утворюються нові зв'язки. Інколи ці процеси проходять досить повільно, а інколи – в режимі реального часу, тобто зі швидкістю, співставною з інтенсивністю запитів щодо структури та властивостей мережі.

– стохастичність. Еволюція складних мереж, їх поточні розміри та структура не є детермінованими, а є наслідками реалізації випадкових процесів.

– наявність загальних спільних рис, зокрема, безмасштабність. Незважаючи на широке розмаїття мереж, які вважаються складними, вони мають між собою деякі спільні риси, які зумовлюють можливість класифікації, моделювання та узагальнення. Зокрема, відомо, що більшість (хоч і далеко не всі) мереж реального світу мають властивість інваріантності до масштабу (scale-free). Це означає, що мережа в цілому є статистично подібною за характеристиками своїм фрагментам. Які, в свою чергу, подібні своїм часткам і так далі.

Структуру взаємозв'язків між багатьма об'єктами природи, або суспільства, можна ефективно та повно описати у вигляді графової, або мережевої) моделі, де кожному об'єкту відповідає вершина, або вузол, а зв'язку між парою об'єктів – ребро, чи дуга. Так відносини між людьми у

групі, відносини між фірмами, комп'ютерні мережі, взаємодії між нейронами – це приклади мереж [2], [3].

Доказ знаменитої теореми про чотири фарби Аппелем і Хакеном (Kenneth Appel, Wolfgang Haken) в 1976 (чотирьох фарб достатньо, щоб розфарбувати будь-яку пласку карту так, щоб дві сусідні області були пофарбовані в різні кольори) – це найбільш відоме останнє досягнення в теорії графів.

Найвагоміші за останні 30 років результати у вивченні мережевих структур були отримані фізиками. Виявилося, що методи фізики, насамперед статистичної механіки, добре підходять до проблематики складних мереж. На відміну від математиків, фізики у своїх дослідженнях спираються на емпіричні дані про реальні мережі, такі як Інтернет, мережі друзів та знайомих або біологічні метаболічні мережі. Під час дослідження статистичних властивостей мереж, наприклад, законів розподілу вузлів за кількістю зв'язків, було виявлено велику кількість непередбачуваних властивостей реальних мереж, на які не звернули увагу ані математики, ані соціологи. Ці властивості стали стимулом для розробки нових теорій, моделей, виявлення нових фундаментальних властивостей мереж.

Топологічні властивості мереж, що розглядаються абстрактно від їхньої природи, але істотно визначають функціонування мереж, і становлять предмет дослідження теорії складних мереж.

Термін «складні мережі» виник на початку цього століття і відноситься до мереж з більш складною архітектурою, ніж класичні випадкові мережі із заданим числом вузлів та зв'язків або решіток у кристалах. Зазвичай у мережах, які вважаються «складними» є відносно невелика кількість вузлів із великою кількістю зв'язків – так звані хаби (від англійського hub – ядро, концентратор), які значною мірою визначають властивості мережі в цілому. При цьому виявилося, що більшість реальних мереж (біологічних, технічних, соціальних) є складними. [4]

1.1.1 Характеристики вузлів мереж

Для окремих вузлів виділяють такі параметри:

– вхідний ступінь вузла – кількість ребер графа, що входять до вузол;

– вихідний ступінь вузла – кількість ребер графа, що виходять із вузла;

– середня та максимальна відстані від цього вузла до інших;

– ексцентричність (eccentricity) – найбільша з мінімальних відстаней від цього вузла до інших;

– посередництво (betweenness), яке показує, скільки найкоротших шляхів проходить через цей вузол;

– центральність та інші параметри.

1.1.2 Розподіл ступенів вузлів

Розподіл вузлів за кількістю зв'язків $p(k)$ описує ймовірність того, що випадково обраний вузол мережі має ступінь k:

$$p(k) = \frac{N(k)}{N}, \tag{1.1}$$

де $N(k)$ – кількість вузлів степені k у мережі;

N – загальна кількість вузлів у мережі.

Розподіл вузлів за кількістю зв'язків є найпростішою статистичною характеристикою довільної мережі. Важливо, що у багатьох випадках знання цієї характеристики достатньо для розуміння властивостей цієї мережі та процесів, що у ній відбуваються. У класичних випадкових графах розподіл вузлів за кількістю зв'язків (degree distribution) спадає досить швидко, оскільки для випадкових мереж (Ердеша-Реньї) цей розподіл підпорядковується закону Пуассона:

$$p(k) = e^{-q} \cdot \frac{q^k}{k!}, \tag{1.2}$$

де k – степінь вузла;

$q = N \cdot p.$

Багато реальних мереж – від метаболічних мереж у клітинах до мережі інтернет – мають розподіл вузлів, в якому щільність повільно спадає із зростанням k. В таких мережах хаби (вузли з великою кількістю зв'язків) становлять помітну частку від усіх вузлів, і саме вони визначають багато важливих властивостей цих мереж. Асимптотична залежність $p(k)$ є степеневою при високих значеннях k.

Розподіл вузлів за кількістю зв'язків у вигляді степеневого закону ще називається безмасштабним (scale free), а самі мережі отримали назву безмасштабних, або масштабно-інваріантних мереж (scale free networks), що означає відсутність типової кількості зв'язків у такій мережі.

1.1.3 Найкоротший шлях між вузлами

Відстань між вузлами визначається як кількість кроків, які необхідно зробити, щоб існуючими ребрами дістатися від одного вузла до іншого. Природно, вузли можуть бути з'єднані безпосередньо або опосередковано, через інші вузли.

Найкоротшим шляхом (SP, shortest path) між вузлами називається мінімальна відстань між ними. Для всієї мережі можна запровадити поняття середнього найкоротшого шляху, як середні за всіма парами вузлів мінімальні відстані між ними:

$$L^{sp} = \frac{2}{n(n+1)} \sum_{i,j} l_{ij}^{\min} \tag{1.3}$$

де n – кількість вузлів;

l_{ij}^{min} – найкоротша відстань між вузлами.

1.1.4 Посередництво

Посередництво (betweenness) – це параметр, який показує, скільки найкоротших шляхів проходить через вузол. Вузли з найбільшим посередництвом грають головну роль встановленні зв'язків між іншими вузлами у мережі. Посередництво b_m вузла m визначається за формулою:

$$b_m = \sum_{i \le j} \frac{B(i,m,j)}{B(i,*,j)}, \tag{1.4}$$

де $B(i,*,j)$ – загальна кількість найкоротших шляхів між вузлами i та j;

$B(i,m,j)$ – кількість таких найкоротших шляхів між i та j, які проходять через вузол m.

1.1.5 Коефіцієнт кластеризації

Уаттс та Строгатц запровадили такий параметр мереж, як коефіцієнт кластеризації. Цей коефіцієнт характеризує тенденцію до утворення груп взаємозалежних вузлів, так званих клік (clique). Для конкретного вузла коефіцієнт кластеризації показує, скільки найближчих сусідів даного вузла є найближчими сусідами друг для друга. Відношення реальної кількості зв'язків, які з'єднують найближчих сусідів даного вузла до максимально можливого (такого, при якому всі найближчі сусіди даного вузла були б пов'язані безпосередньо один з одним) називається коефіцієнтом кластеризації вузла. Природно, ця величина не перевищує одиниці.

Будемо називати два вузли сусідами, якщо існує зв'язок між ними. Для комплексних мереж характерно, що два вузли, сусідні до якогось вузла, часто також є сусідами між собою. Щоб охарактеризувати це і був запропонований коефіцієнт кластеризації C_i вузла i. Припустимо, що вузол має ступінь k_i це означає, що у нього є k_i сусідів, між якими може бути максимум $k_i(k_i - 1)$ зв'язків. Тоді

$$C_i = \frac{2n_i}{k_i(k_i - 1)} \tag{1.5}$$

де n_i – кількість зв'язків між сусідами вузла i.

Коефіцієнт кластеризації може бути співставлений не тільки окремому вузлу (1.5), а й всій мережі в цілому:

$$C = \frac{1}{n}\sum_{i=1}^{n} C_i \tag{1.6}$$

де n – кількість вузлів;

C_i – коефіцієнти кластеризації окремих вузлів.

Коефіцієнт кластеризації C в загальному випадку визначається як потроєние відношення кількості «трикутників» (три вершини, з'єднані трьома ребрами) у графі до кількості «виделок» (шляхів довжиною два).

1.2 Моделі складних мереж

З суто математичних позицій будь-яка модель має право існування. Однак для додатків – у тому числі додатків до транспортної проблематики – деякі з цих моделей цікавіші, деякі – менші. Відповідно нижче представлені три класи моделей, кожен із яких за десятиліття, минулі з його появи, зарекомендував себе плідним як у рамках «чистої» математики, так і у межах її різноманітних додатків, серед яких надійність транспортної мережі, моделювання зростання мережі інтернет, соціальних та біологічних мереж тощо. Далі виділено лише основні і важливі моменти теорії випадкових графів. [15]

Модель Ердеша-Реньї є найпростішою та дозволяє конструювати штучні мережі з великою кількістю одночасно існуючих у них функціональних структур.

Модель Барабаші-Альберт, навпаки, є найгіршою з погляду

конструювання якомога більшої кількості одночасно існуючих функціональних структур, проте найоптимальнішою з погляду схожості з реальними соціальними, комп'ютерними, біологічними мережами. Саме ця модель є базовою моделлю для безмасштабних мереж.

Модель Ватса-Строгатца є, за сутністю, перехідною ланкою між регулярною кільцевою решіткою та випадковим графом моделі Ердеша-Реньї. Вона дозволяє конструювати мережі типу «світ тісний» (small world).

1.2.1 Модель Ердеша-Реньї

Модель Ердеша – Реньї є історично першою з моделей випадкового графа, в якому ймовірність появи ребра між двома вершинами не залежить від інших ребер та є постійною. Згідно з цією моделлю, граф конструюється за допомогою з'єднання пар вершин випадковим чином. Кожне ребро включається до графа з ймовірністю $0 < p < 1$ незалежно для кожної пари існуючих вершин. Зі зростанням p від 0 до 1 граф стає все більш насиченим ребрами. При $p = 0$ маємо порожній граф (граф без ребер), при $p = 1$ граф є повнозв'язним. При фіксованій кількості вершин у графі значення p є єдиним вхідним параметром моделі Ердеша – Реньї.

Простота моделі Ердеша-Реньї дозволяє отримувати різні аналітичні оцінки. Однак варто відзначити, що модель в середньому не відтворює такі типові властивості реальних мереж, як степеневий розподіл степенів вершин, величину коефіцієнта кластеризації та довжини маршруту.

1.2.2 Модель Уаттса-Строгатца

У період з кінця XX сторіччя до теперішнього часу вийшло багато книг, що містять сучасні уявлення про мережі, якими можна певною мірою пояснити феномен популярності тематики соціальних мереж. Автором двох із них є Дункан Уаттс. У 1999 році він опублікував книгу «Малі світи, динаміка мереж від порядку до випадковості» (Small Worlds: The Dynamics

of Networks between Order and Randomness), написану за матеріалами власної дисертації. Дослідження наводилися в Santa Fe Institute, відомому ще як Інститут вивчення складних систем. У 2004 році з'явилася розширена версія тієї ж книги під назвою «Шість рукостискань, наука про зв'язаний світ» (Six Degrees: the science of a connected age). У цих роботах автор розглядає феномени системних зв'язків між людьми з позиції психолога та соціолога. Уаттс був одним із перших, хто з часів перших експериментів Мільграма спробував побудувати математичну модель малого світу і поширити її на дослідження, які безпосередньо не пов'язані з соціальними мережами. Виявляється, що підхід, який ґрунтується на принципах малого світу, добре підходить для досліджень, пов'язаних з поширенням епідеміологічних захворювань, катастроф в енергосистемах та інших системах, що мають мережеву природу.

Дункан Уаттс та Стівен Строгатц виявили феномен, характерний для багатьох реальних мереж, названий ефектом «тісного світу».

Модель «тісного світу» полягає в тому, що перебираючи коло своїх найближчих знайомих, потім людей, які знають наших найближчих знайомих (але не знають нас безпосередньо), і т.д. легко переконатися в наступному: достатньо простежити за невеликим числом ланцюжків таких знайомств, щоб зрозуміти, що будь-хто опосередковано знайомий з будь-яким членом суспільства. У цьому сенсі наш світ є тісним, звідки і походить назва цієї моделі. [6]

Мережеві структури, що відповідають властивостям малих світів, мають такі типові властивості: мала середня довжина шляху між вузлами у порівнянні з діаметром мережі та великий коефіцієнт кластеризації.

При дослідженні цього феномена ними було запропоновано процедуру побудови моделі моделі, якій властивий цей феномен.

На рисунку 1.1 представлена ілюстрація модели «світ тісний».

На рисунку можна побачити трансформацію мережі з регулярної кільцевої решітки в модель «світ тісний», а потім у випадковий граф.

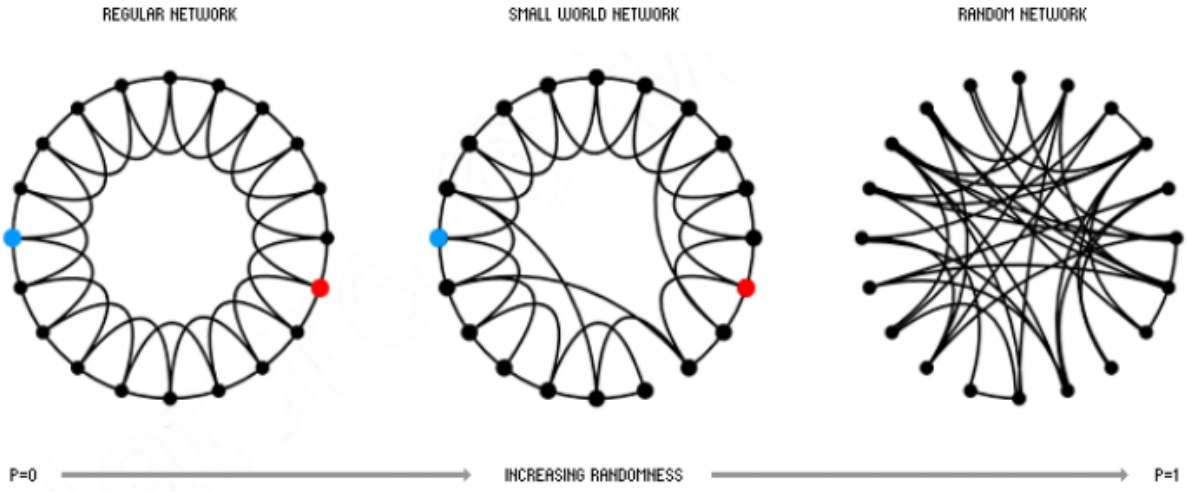


Рисунок 1.1 – Модель «світ тісний» Уаттса-Строгатца

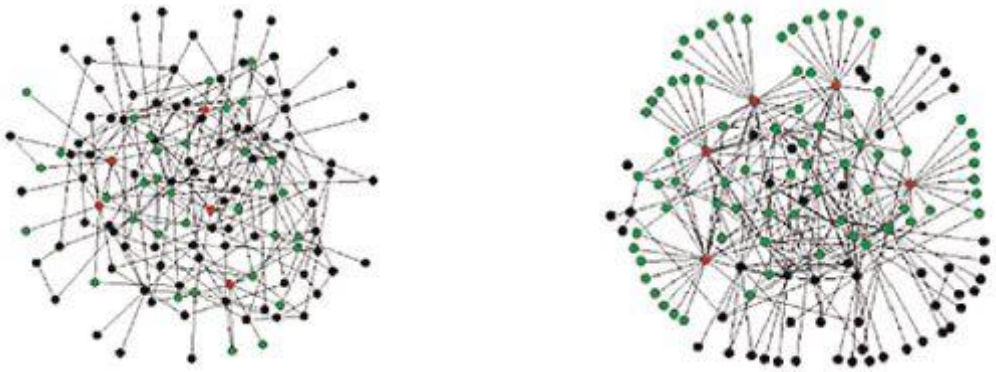
Комп'ютерна модель «малого світу» була розроблена Уаттсом та Строгатсом декількома роками пізніше. Модель Уаттса-Строгатса є модель генерації випадкового графа, що має високий коефіцієнт кластеризації вершин і відносно невелику середню довжину шляху.

1.2.3 Модель Барабаши-Альберт

Однією з найцікавішою робот у галузі складних мереж визнано статтю Альберта Ласло Барабаші «Зв'язаність, нова наука про мережі». За своєю освітою та тодішньою основною спеціальністю Барабаші – фізик. Отже, він підійшов до вивчення соціальних мереж з природничих позицій. Барабаші настільки впевнений у фундаментальній важливості мереж і пояснює специфікою мереж таку безліч явищ, що його порівнюють з Піфагором, який говорив «все є число». Так саме Барабаші сказав, що «все є мережа». Його підхід отримав висвітлення у багатьох виданнях, починаючи від наукових журналів Nature, Science, Science News та популярних American Scientist, Discovery, National Geographic та New Scientist до газет New York Times, USA Today, Washington та Business Week. Цей дослідник давав інтерв'ю багатьом значним телевізійним і радіоканалам.

Барабаші побудував свою систему поглядів, використовуючи нові

підходи до теорії мереж і зокрема розроблений ним математичний апарат безмасштабних мереж (scale-free network) (рисунок 1.2). Новизна поглядів Барабаші полягає у тому, що до нього соціальні мережі вважалися абсолютно випадковими, а він показав, що ці мережі мають складну внутрішню структуру. Зокрема, він запропонував генеративну модель для таких мереж та пояснив, чому у мережах відбувається поділ вузлів на «бідну більшість» таких, що мають невелику кількість зв'язків та «багату меншість» вузлів-хабів.



а) б)

Рисунок 1.2 – Приклади мереж: а) з експоненційним розподілом степенів вузлів; б) безмасштабна (з степеневим розподілом)

Зокрема, Барабаші показав, що у тому випадку, якщо деяка мережева система еволюціонує без впливу зовнішніх регуляторів, то кількість зв'язків, якими обростають вузли, не є випадковою. Кількість зв'язків у окремо взятого вузла розподіляється не за законом Пуассона, а за степеневим законом (рисунок 1.3). Звідси випливає, що в більшості реальних мереж основна частина вузлів має обмежену кількість зв'язків, а окремі вузли-концентратори (Барабаші називає їх hub) мають аномально велику кількість.

Можна говорити про наявність ієрархії вузлів у мережах, яка складається під впливом тих, чи інших факторів. Ієрархічність у структурі мереж є поясненням деяким емпіричним законам, таким, наприклад, як

«правило Парето 20/80», а також успіхів таких речей, як мережевий маркетинг, особливості поширення епідеміологічних захворювань та багато інших феноменів, які ми спостерігаємо в житті. На основі теорії Барабаші можна навіть пояснити євангельську тезу «Хто має, тому дано буде; а хто не має, у того забереться і те, що він думає мати».

Рисунок 1.3 – Типовий розподіл вузлів масштабно-інваріантної мережі за кількістю зв'язків

Модель графа Барабаші-Альберт (граф БА) є алгоритмом покрокової генерації випадкових безмасштабних мереж з використанням правила переважного приєднання (рисунок 1.4).

Правило переважного приєднання стверджує, що чим більший ступінь зв'язності має вершина, тим вища ймовірність приєднання до неї нових вершин.

Мережа БА вирощується з невеликого графа-зародку, у якого ступінь зв'язності кожної вершини має бути не менше одиниці.

Кожна нова вершина приєднується до вже існуючих вершин з ймовірністю, яка пропорційна до степеню (тобто кількості зв'язків) цієї існучої вершини. Отже, ймовірність (p_i) того, що між новою вершиною та

існуючою вершиною i утвориться зв'язок, дорівнює

$$p_i = \frac{k_i}{\sum_{j=1}^{n} k_j}, \tag{1.7}$$

де k_i – степінь вузла i; n – поточний розмір мережі (кількість існуючих вузлів).

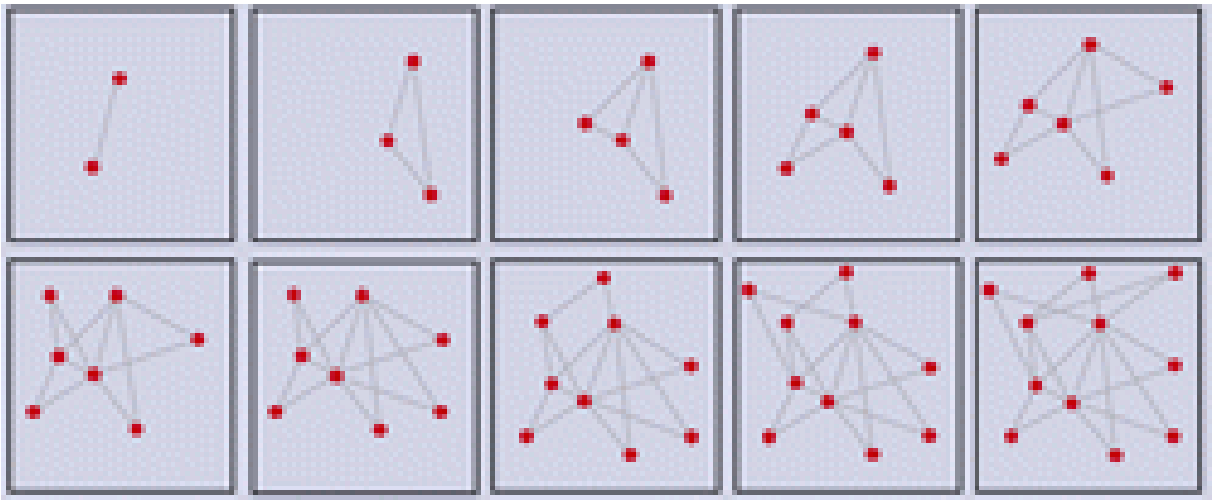


Рисунок 1.4 – Еволюція безмасштабної мережі

1.2.4 Безмасштабність мереж

Безмасштабна мережа (scale-free network) – це така мережа, в якій ступені вершин розподілені за степеневим законом або таким, який асимптотично наближається до степеневого. Це означає, що частка вузлів мережі, які мають k зв'язків для великих значень k становить

$$p(k) \approx c \cdot k^{-\gamma}, \tag{1.8}$$

де c – нормалізуючий параметр;

γ – показник скейлінга мережі, який, зазвичай, знаходиться у діапазоні $\gamma \in (2; 3)$, але може бути й більше трьох.

Безмасштабні мережі ще називають інваріантними до масштабу, або масштабно-інваріантними мережами.

Емпірично було встановлено, що багато природних мереж – соціальні, комунікаційні, біологічні, графи цитувань, посилань у WWW, та інші – добре моделюються безмасштабними графами.

У 1956 року Дерек Джон де Солла Прайс показав, що у мережах цитувань між науковими публікаціями кількість посилань на публікацію має розподіл із важкими хвостами, тобто розподіл, схожий на степеневий закон. У своїх роботах він не використовував термін «безмасштабні мережі». Цей термін з'явився набагато пізніше, наприкінці 90-х років, коли Альберт-Ласло Барабаші у своїх дослідженнях топології всесвітньої павутини виявив вузли-концентратори, що мають велику кількість зв'язків. Згодом він виявив подібні вузли в інших мережах, після чого дав знайденому класу мереж назву «безмасштабні мережі».

Безмасштабна мережа є зростаючою структурою, яка складається з вузлів, що зв'язуються особливим чином. Її побудова починається з єдиного поодинокого вузла. Потім на кожному кроці еволюції в структуру додається по одному новому вузлу, який зв'язується з одним з вузлів, які вже є в мережі. Так мережа постійно зростає.

Головна відмінна риса безмасштабних мереж – існування вузлів-концентраторів (hubs), степені яких є дуже великі у порівнянні з степенями переважної більшості вузлів.

Безмасштабні мережі малочутливі до пошкоджень. Виявляється, що великі концентратори здебільшого оточені меншими концентраторами, а ті, своєю чергою ще меншими тощо. Якщо один із концентраторів буде втрачено, то майже всі зв'язки в мережі збережуться за рахунок існування концентраторів меншого розміру.

Ключова особливість масштабно-інваріантної мережі полягає в особливому алгоритмі з'єднання вузлів. Новий вузол тим швидше приєднається до якогось існуючого, чим більше інших вузлів до нього вже

приєдналося. Нехай, наприклад, є мережа з трьома вузлами (рисунок 1.5). Ймовірність приєднання нового вузла до одного з них визначається пропорцією 1/4:1/2:1/4, тому що відповідні вузли мають зв'язки в пропорції 1:2:1. Таким чином, з ймовірністю 1/2 новий вузол приєднається до другого вузла, і це найімовірніший розвиток подій. Взагалі, вузол тим швидше накопичує нові зв'язки, чим їх у нього більше, бо ймовірніше, що нові вузли приєднуватимуться саме до нього. Тобто, «багаті» вузли стають ще багатшими. Навпаки, «бідні» вузли схильні і залишатися «бідними».

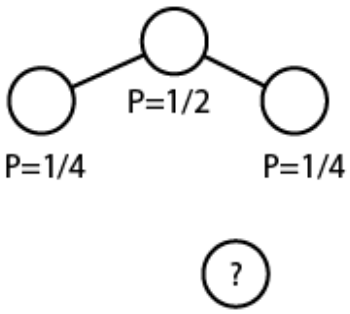


Рисунок 1.5 – Переважне приєднання вузлів

В результаті розвивається мережа характерного виду, яку називають масштабно-інваріантною (рисунок 1.6).

Слід зазначити, що зображення масштабно-інваріантної мережі, наведене на рисунку 1.6, є умовним, тому що насправді вузли не мають якогось просторового розташування, воно для вузлів не задано. Можна було б зобразити мережу будь-яким заплутаним чином.

Очевидною особливістю цієї структури є наявність великої кількості вузлів, які мають дуже мало зв'язків (один, чи два), та невеликої множини вузлів, які мають дуже багато зв'язків (десятки, сотні та більше). Статистично це виражається в тому, що ранговий розподіл вузлів за кількістю зв'язків для досить великої мережі зазвичай має яскраво виражений степеневий характер (рисунок 1.7).

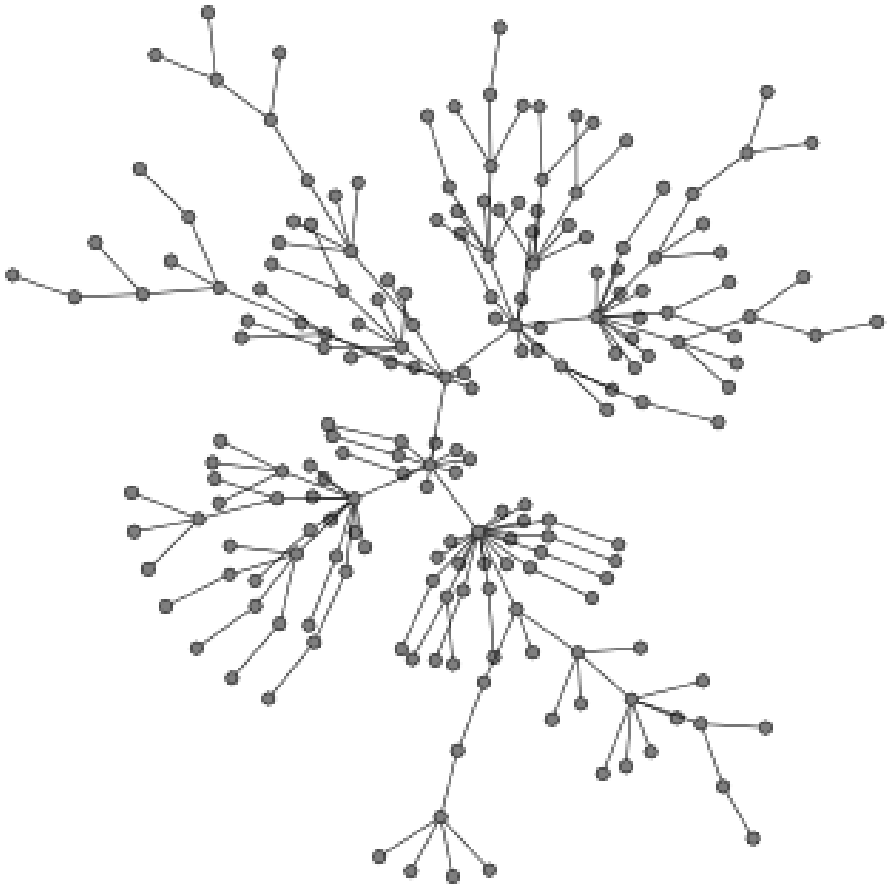


Рисунок 1.6 – Масштабно-інваріантна мережа – дерево Барабаші-Альберт

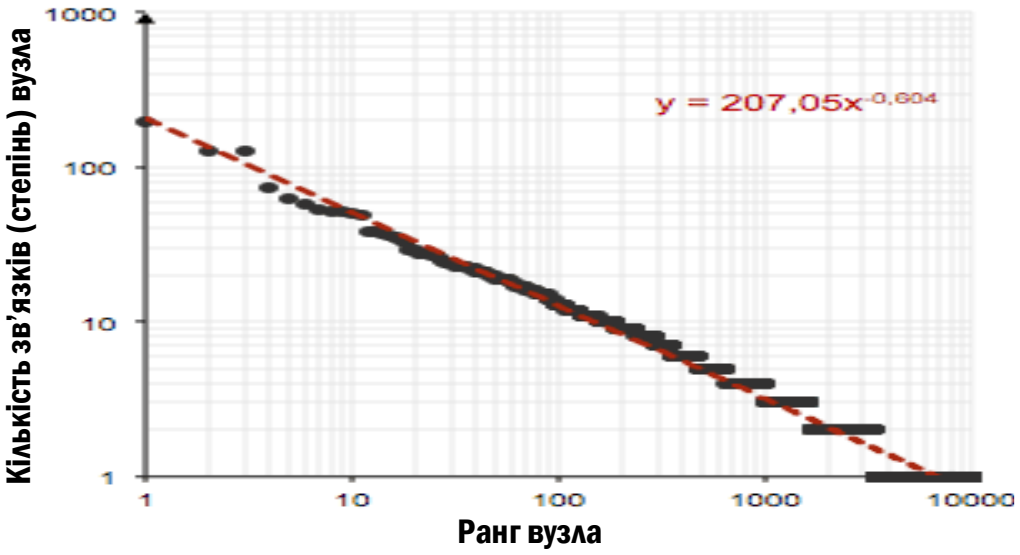


Рисунок 1.7 – Приклад рангового розподілу вузлів у масштабно-інваріантній мережі

Важливим є те, що подібний степеневий розподіл виникає вже на ранніх стадіях еволюції моделі, коли вузлів ще не більше 100, і зберігається майже незмінним на всьому доступному для огляду за допомогою звичайного комп'ютера горизонті часу.

Степеневий розподіл є властивим для параметрів, які не мають характерного масштабу, тобто для масштабно-інваріантних. Для мереж таким параметром є кількість зв'язків у вузлів. Степеневий розподіл їх кількості означає, що цей параметр немає якогось «типового» чи «характерного» значення.

Модель масштабно-інваріантної мережі дуже проста, і це робить її привабливою як пояснення походження багатьох існуючих у реальному світі степеневих розподілів. Зокрема, як правило, з її допомогою пояснюють степеневий розподіл популярності веб-сайтів, статистику завантаження інтернет-вузлів тощо.

Нехай кожен вузол є конкретним географічним місцем, а зв'язки між ними – це люди. Тоді появу нового вузла можна порівняти з ситуацією народження нової людини в певній географічній точці, а зв'язок, який новий вузол простягне до одного з вже наявних вузлів – це відображення вибору місця проживання, який зрештою зробить ця людина. Правило «багатий стає багатшим» в даному випадку відображає перевагу людей вибирати для життя ті місця, де вже мешкає багато людей, тому що в цьому випадку відкриваються максимальні можливості для реалізації своїх здібностей, збагачення тощо.

1.3 Поняття безмасштабності мереж, моделі безмасштабних мереж

У 1999 році фізик з університету Нотр Дам (США) Л. Барабаші (Laszlo Barabasi) разом зі своєю аспіранткою Р. Альберт (Reka Albert) вивчали властивості реальних мереж з дещо іншого погляду. Якщо Строгатц і Уаттс у своєму дослідженні мереж виходили з феномену тісного світу, то Барабаші

та Альберт вирішили дослідити закон розподілу вузлів деяких реальних мереж за кількістю зв'язків. Результат виявився несподіваним. Замість розподілу числа зв'язків за законом Пуассона, який має строгий максимум близько середнього значення, для багатьох реальних мереж, наприклад, таких як метаболічні мережі та білкові взаємодії в клітинах, структура авіаційних повідомлень, структура інтернету тощо, такого середнього значення не існує.

Таким чином, у багатьох реальних мережах невелика кількість вузлів містить дуже велику кількість зв'язків, а величезна кількість вузлів містить лише кілька зв'язків. Такі мережі одержали назву безмасштабних мереж (scale free networks). У своїх роботах Барабаші, а також Х. Джеонг описали ті статистики інтернету, які лягли в основу науки про зростання цієї мережі – науки, що має глибокі додатки як у власне інтернет-проблематиці, так і в численних близьких дисциплінах.

Насправді більшість реальних мереж (соціальні, біологічні, транспортні та ін.) мають схожу «топологію». Мережа інтернет – це так званий веб-граф, вершинами якого є деякі конкретні структурні одиниці в інтернеті: сторінки, сайти, хості, власники тощо. Найчастіше ними є саме сайти. Дугами з'єднуються ті вершини, між якими є посилання. При цьому розумно проводити стільки дуг між двома вершинами, скільки посилань між відповідними сайтами. Отже, зв'язки (дуги) природно вважати спрямованими. Таким чином, веб-граф є орієнтованим і може мати кратні ребра, петлі і навіть кратні петлі (посилання цілком можуть йти з однієї сторінки даного сайту на іншу його сторінку). Тобто такий граф є псевдомультіорграфом. Відразу зрозуміло, що для такого «звіра» модель Ердеша – Реньї навряд чи підходить.

Основних тез у дослідженні Барабаші та Альберт є, фактично, три. По-перше, веб-граф є дуже розрідженим графом. Він має на n вершинах приблизно $k \cdot n$ дуг, де $k \geq 1$ – це деяка не дуже велика (у порівнянні з n) константа, в той час як максимально можлива кількість дуг у звичайному

орієнтованому графі без петель та кратних дуг становить $n \cdot (n-1)$.

Однак – і це по-друге, – діаметр веб-графа є дуже невеликим. У 1999 року він мав величину 6. Це добре всім відома властивість будь-якої соціальної мережі, яку прийнято в повсякденному мовленні характеризувати виразом «світ тісний». Наприклад, кажуть про те, що будь-які дві людини у світі «знайомі через 5–6 рукостискань». Так само і сайти: переходячи за посиланнями, можна перейти з будь-якого сайту на будь-якій іншій за 5–7 натискань. Звичайно, тут є важливе застереження: новостворені сайти можуть не бути пов'язані з зовнішнім по відношенню до них світом. Отже, веб-граф є дуже специфічним графом: він розріджений, але, одночасно з цим, він тісний.

По-третє, у веб-графа дуже характерний розподіл ступенів вершин: степеневий. Цей факт ріднить інтернет з багатьма реальними мережами – біологічними, соціальними, транспортними. Усі вони підкоряються степеневому закону.

Зважаючи на перераховані спостереження, не залишається жодних сумнівів у тому, що модель Ердеша – Реньї не застосовна для опису зростання інтернету та подібних мереж. Якщо підбором ймовірності p ще можна домогтися розрідженості і тісноти (хоча і не з тими параметрами), то степеневий розподіл зовсім не може з'явитись як наслідок схеми Бернуллі, в рамках якої з'являються ребра звичайного випадкового графа. У моделі $G(n,p)$ степінь кожної вершини випадкового графа біноміальна з параметрами $n-1$ і p, і при таких p, які забезпечують більш-менш природню розрідженість, зазначений біноміальний розподіл апроксимується розподілом Пуассона, а зовсім не степеневим.

Самі Барабаші та Альберт запропонували дуже розумний погляд на процес формування інтернету. Вони припустили, що в кожний момент часу з'являється новий сайт, і цей сайт ставить фіксовану кількість посилань на своїх попередників. На кого він віддасть перевагу послатися? Напевно, на тих, хто й так уже популярний. Можна припустити, що ймовірність, з якою

новий сайт поставить посилання на один із існуючих сайтів, пропорційна числу посилань, що вже були на той сайт. Моделі випадкових графів, що ґрунтуються на описаній ідеї, називаються моделями переважного приєднання. У своїх роботах Барабаші та Альберт ніяк не конкретизували, яку саме з цих моделей вони пропонують розглядати. А ці моделі виключно різнорідні за своїми властивостями. Адже можна ставити посилання незалежно одне від одного, а можна враховувати ще й залежність між різними посиланнями з одного сайту.

Нещодавно отримано свідчення того, що функціональні зв'язки у мозку людини утворюють безмасштабні мережі. У мозку тварин та людини постійно йдуть процеси самоорганізації та розпаду, функціональних нейромережевих структур. Кожна така мережа, що поєднує нейронні ансамблі з різних відділів мозку, створюється для реалізації певної функції мозку – поведінкової, когнітивної тощо. Саме такі функціональні нейронні мережі стали об'єктом вивчення групи дослідників на чолі з Д. Чіалво (Dante Chialvo), метою яких було перевірити можливість застосування ідей та методів сучасної теорії мереж (безмасштабні мережі, мережі тісного світу тощо) до топології реальних функціональних структур мозку.

Використовуючи метод магнітно-резонансного зображення, вимірювалася активність мозку при здійсненні простої моторної дії на кожному часовому інтервалі в 2.5 секунд протягом 400 таких інтервалів. Вивчалася область мозку розміром 36×64×64 просторових осередків з розміром кожного осередку 3×3.475×3.475 мм^3. Фіксувалась активність кожного такого осередку при виконанні таких простих функцій, як, наприклад, стукіт пальцем по столу. При реєстрації активності кожного такого осередку можна визначити, які осередки взаємопов'язані між собою (якщо активність осередків у різних частинах мозку виявляє властивість синхронності, це вказує на те, що між ними є зв'язок). Отже, кожні дві таких комірки вважались функціонально пов'язаними між собою, якщо значення просторово-часової кореляції між ними перевищує деякий

поріг. Виявилося, що такі зв'язані осередки утворюють безмасштабну мережу, яка містить 31203 вузла, з властивостями тісного світу та з степеневим розподілом степенів вузлів: $P(k) \approx c \cdot k^{-\gamma}$, де показник скейлінгу становив $\gamma \approx 2$.

Також вивчались структурні властивості продуктів художньої творчості. Наприклад, партитура музичного твору для рояля легко перетворюється на мережеву структуру, якщо вузлами такої мережі призначити музичні ноти всіх можливих тривалостей. Число вузлів у такій мережі не буде перевищувати 1800. Дійсно, число клавіш у рояля дорівнює 88 і, множачи це число на 20 (кількість можливих тривалостей ноти: цілі, половинні, чверті, восьмі, тощо), отримуємо 1760. Лі (Xiaofan Liu), Тсе (Chi K. Tse) та Смол (Michael Small) з політехнічного університету Гонконгу проаналізували статистичні властивості мереж, побудованих за описаним вище принципом, для творів Баха, Моцарта, Шопена та сучасних китайських композиторів, що працюють у жанрі поп-музики. Усі ці мережі виявилися безмасштабними. Середня кількість кроків між вузлами у цих мережах варіюють у діапазоні від 2.8 до 4.2.

Барабаші та Альберт запропонували просту та елегантну модель виникнення та еволюції безмасштабних мереж. Вони показали, що для виникнення безмасштабних мереж необхідні дві умови:

– зростання. Починаючи з невеликого числа m_0 вузлів, на кожному часовому кроці додається один новий вузол з m зв'язками, які з'єднують цей новий вузол з m різними вже існуючими вузлами;

– переважне приєднання (preferencial attachment). Коли вибираються вузли, до яких приєднується новий вузол, то вважається, що ймовірність, з якою новий вузол буде з'єднуватися з існуючим вузлом, пропорційна до кількості зв'язків, якими цей вузол вже пов'язаний з іншими вузлами.

Цей принцип був раніше запропонований Прайсом і Саймоном для пояснення походження степеневого закону у соціальних явищах. Фізики розглядають такі явища як властивості критичних станів.

Обчислення показують, що принцип переважного приєднання призводить до безмасштабної мережі з показником скейлінгу $\gamma = 3$. Саме таке значення скейлінгу досить рідко зустрічається в реальних мережах, але цінність цієї моделі Барабаші-Альберт полягає в тому, що вона показує принципову можливість генерації безмасштабної мережі на базі простих правил.

Розглянемо, як узагальнити цю модель, щоб отримати степеневий закон із довільним показником степені. Нехай можливість приєднання нового вузла до існуючого вузла, який має степінь q, пропорційна деякій функції $f(q)$, яка називається функція переваги. У моделі Барабаші-Альберт $f(q) = q$. Трохи модифікуємо цю функцію, нехай вона буде лінійною: $f(q) = q + A$, де $A = a \cdot m$ має сенс «додаткової привабливості», $a \geq -1$, m – число зв'язків кожного нового вузла. Можна показати, що така лінійна модифікація призводить до безмасштабної мережі з $\gamma = 3 + a$, отже діапазон показника скейлінгу становитиме від 2 до нескінченості.

Якщо функція переваги має вигляд q^{θ}, де θ – деяке невід'ємне дійснозначне число, то можна показати, що механізм переважного приєднання призводить до степеневого розподілу і, відповідно, безмасштабності мережі, тільки при $\theta = 1$, тобто тільки коли функція переваги є лінійною.

Принцип переважного приєднання вперше використовувався для пояснення появи степеневого розподілу у дослідженні Юла щодо видового-родового різноманіття класифікації рослин в 1925 році. Сучасний метод основного кінетичного рівняння, який дає той самий висновок, що й у Юла, але прозорішим чином, був запропонований Гербертом Саймоном в 1955 та був ним застосований до проблеми дослідження розмірів міст. До зростання мереж цей принцип вперше був застосований Дерек де Солла Прайсом в 1976, який досліджував мережі цитування між науковими публікаціями. Назва «переважне приєднання» та сьогоднішню популярність моделей безмасштабних мереж пов'язана з роботою Барабаші та Альберт, які

застосували його до розподілу посилань між інтернет-сторінками.

У безмасштабних мережах вузли з більшим степенем мають більше можливостей забирати собі зв'язки, що додаються до мережі. Інтуїтивно принцип переважного приєднання може бути зрозумілий, якщо ми думаємо в термінах соціальних мереж, які об'єднують людей. Тут зв'язок від A до B означає, що людина A знайома з людиною B. Вузли з великим степенем відповідають відомим людям з великою кількістю зв'язків. Коли новачок потрапляє до спільноти, для нього краще зв'язатися з одним з відомих людей, ніж з відносно невідомим. Подібним чином у всесвітній мережі сторінки пов'язуються з хабами, наприклад, із добре відомими сайтами, як Google або Вікіпедія, ніж зі сторінками, які мало, кому відомі. Якщо вибирати для зв'язку нову сторінку випадковим чином, то можливість вибору певної сторінки буде пропорційна її степені. Це пояснює принцип переважного приєднання. Цей принцип є прикладом позитивного зворотного зв'язку, де спочатку випадкові варіації (один вузол спочатку має більше посилань або починає збирати посилання раніше за інші) автоматично посилюються, тим самим значно збільшуючи розрив. Це також іноді називають ефектом Матвія, «багаті стають багатшими», або автокаталізом у хімії.

1.4 Постановка задач досліджень

Метою кваліфікаційної роботи є дослідження моделей складних мереж та розробка такої генеративної моделі, в якій була б можливість роздільної та незалежної параметризації показників скейлінгу та еластичності генерованої мережі.

Досягнення поставленої мети забезпечується виконанням наступних задач:

– провести аналіз предметної галузі;

– провести порівняльну характеристику існуючих моделей складних мереж, зокрема, безмасштабних;

– провести дослідження щодо структурованості спільнот у безмасштабних мережах, міри гомофільності мереж;

– на основі зазначених моделей розробити генеративну модель безмасштабної мережі з гомофільною структурою спільнот;

– виконати програмну реалізацію запропонованої моделі та шляхом моделювання перевірити її працездатність та коректність.

2 РОЗРОБКА МОДЕЛІ БЕЗМАСШТАБНИХ МЕРЕЖ З ГОМОФІЛЬНОЮ СТРУКТУРОЮ СПІЛЬНОТ

2.1 Модулярність як міра гомофільності мереж

Відомо, що більшість мереж реального світу не є однорідними. Це проявляється, наприклад, в тому, що вузли мають суттєво різні степені (тобто кількість зв'язків). Дуже часто розподіл вузлів за степенями є степеневим (1.8) (вот така, начебто, тавтологія), тобто існує невелика частка вузлів, які концентрують у себе значний відсоток загальної кількості зв'язків мережі, та велика більшість вузлів, які мають степені, значно менші, за середню. Такі мережі називаються безмасштабними. Ключовим проявом безмасштабності є самоподібність: частка мережі статистично подібна всій мережі.

Іншим проявом неоднорідності мереж є кластеризація: вузли мають тенденцію з'єднуватись у більш-менш чітко виражені спільноти. Так, що густина зв'язків між вузлами, які входять до однієї спільноти, значно вище, ніж густина зв'язків між вузлами різних спільнот. Зазначена тенденція називається гомофільністю. Ідеально гомофільна мережа є незв'язною: вона складиметься з кількох відокремлених спільнот. Зрозуміло, що така модель не відповідає дійсності, тому зв'язність мережі обумовлює неідеальну, нестовідсоткову гомофільність. Більше того: безмасштабність / самоподібність мереж вимагає не тільки степеневого розподілу степенів вузлів мережі, а й такого ж самого (1.8) розподілу степенів вузлів кожної спільноти. З таким же самим показником скелінгу (γ), який діє для вузлів всієї мережі. При цьому цілком слушною є вимога щодо безмасштабності розподілу розмірів самих спільнот (можливо з таким же, а можливо з іншим іншим значенням показника скейлінгу, γ). Отже, накладання на мережу вимоги безмасштабності суттєво впливає на міру її гомофільності.

Задача оцінювання якості кластеризації мережі безпосередньо пов'язана із задачею її поділу на кластери (спільноти). Виявлення спільнот є однією з ключових задач аналізу структури складних мереж. Зокрема, поділ мережі на спільноти дає змогу виконувати її масштабування шляхом представлення кожної спільноти вихідної мережі окремим вузлом у новій, скороченій моделі. Такий підхід дозволяє краще зрозуміти структуру взаємозв'язків між спільнотами та оцінити їхній масштаб. Крім того, кластеризація вузлів сприяє виявленню нетипових елементів і зв'язків (мостів), які мають вирішальне значення для забезпечення цілісності мережі, її стійкості до втрати зв'язності, а також впливають на швидкість поширення вірусів, інформації чи чуток. Отже, створення ефективних методів виявлення спільнот у мережах є актуальним і важливим практичним завданням.

Попри значну практичну цінність цієї проблематики, універсальних і науково обґрунтованих методів її розв'язання на сьогодні не існує. Однією з основних причин є надзвичайно загальний характер задачі поділу мережі на спільноти, що зумовлює наявність різних підходів до її трактування та, відповідно, великої кількості критеріїв оцінювання якості отриманого розбиття. Однією з важливих класифікаційних характеристик такого розбиття є можливість або неможливість перетину спільнот. Інакше кажучи, йдеться про те, чи може окремий вузол одночасно належати кільком спільнотам. Найпростішим випадком є розбиття множини вузлів на підмножини, які попарно не перетинаються.

Отже, об'єктом дослідження є складні мережі, математичною моделлю яких є неорієнтований зважений граф $G(V,E)$. Найбільш поширеними підходами до поділу вузлів мережі на спільноти є методи, засновані на максимізації показника модулярності. Їхньою основною перевагою є відносна простота реалізації, що забезпечує високу обчислювальну ефективність. У зв'язку з цим модулярність також може застосовуватися як кількісна характеристика гомофільності мережі [11].

Припускається, що спільноти мережі не перетинаються попарно, тому розбиття на спільноти $C_u, u = 1,...,k$ називається припустимим, якщо ці спільноти утворюють повну групу: $Nodes = \bigcup_{u=1}^{k} C_u, \quad \forall u \neq v: \; C_u \bigcap C_v = \varnothing$.

Кількість шуканих спільнот (k) у загальному випадку вважається невідомою.

Якщо всі вузли мережі розбиті на k спільнот (розфарбовані k кольорами), які попарно не перетинаються, то модулярність мережі визначена як:

$$\mu(G_n) = \sum_{j=1}^{k} \left(e_{j,n}^{in} - e_{j,n}^{2} \right) = 1 - E_n^{ext} - \sum_{j=1}^{k} e_{j,n}^{2}, \quad \text{where} \quad E_n^{ext} = \sum_{j=1}^{k} e_{j,n}^{ext}, \; (2.1)$$

де $e_{j,n}$ – відносний обсяг спільноти j (тобто сума степенів вузлів кольору j, нормована на vol(G_n) – обсяг всієї мережі);

$e_{j,n}^{in}$ – відносна кількість зв’язків між вузлами всередині спільноти j;

$e_{j,n}^{ext}$ – відносна кількість зв’язків між вузлами спільноти j та вузлами інших спільнот (очевидно, що $e_{j,n} = e_{j,n}^{in} + e_{j,n}^{ext}$);

E_n^{ext} – відносна кількість зв’язків між вузлами різних спільнот по усій мережі, «відносність» всіх зазначених обсягів означає їх нормування на vol(G_n) – обсяг мережі (тобто загальну кількість зв’язків).

Очевидно, що значення модулярності, визначене формулою (2.1) істотно залежить від вибору нуль-моделі мережі. Відповідно до припущення, запропонованого авторами цієї міри [11], як нульову модель використовується модель випадкового графа Erdős–Rényi. Водночас у роботі [12] запропоновано альтернативний підхід, що ґрунтується на іншій моделі формування зв'язків. Згідно з нею критерій модулярності модифікується:

$$\mu^*(G_n) = \sum_{j=1}^{k} \left(e_{j,n}^{in} - p_{j,n} e_{j,n} \right), \quad \text{where} \quad p_{j,n} = \frac{n_j - 1}{n - 1}, \tag{2.2}$$

де n_j – розмір спільноти j (тобто кількість вузлів кольору j);

n –розмір мережі.

Модулярність мережі є показником здатності вузлів групуватися в спільноти. Якщо вузли, що належать до однієї спільноти, характеризуються високою щільністю взаємозв'язків, тоді як зв'язки між різними спільнотами трапляються рідко, значення показника модулярності є високим. У протилежному випадку модулярність набуває низьких значень.

Визначення модулярності базується на використанні як нуль-моделі випадкового графа Ердеша-Рен'ї зі збереженням ступенів вузлів досліджуваної мережі. У випадковому неорієнтованому графі $G(V, E)$ з N вузлами з степенями вузлів k_i, $i = 1,...,N$ ймовірність наявності ребра між вузлами i та j дорівнює $k_i k_j / (2m)$, де $m = size(G)$ це загальна кількість ребер мережі. Для кожної пари вузлів i, j можна визначити різницю між фактичною вагою зв'язку, тобто відповідним елементом матриці суміжності $A_{i,j}$ графа та її очікуваним значенням, розрахованим на основі обраної нуль-моделі:

$$\Delta_{i,j} = A_{i,j} - \frac{k_i k_j}{2m}. \tag{2.3}$$

Модулярність спільноти (C_u) дорівнює сумі значень (2.3), взятої по всіх вузлах, що входять до цієї спільноти:

$$q_u = \frac{1}{2m} \sum_{i,j \in C_u} \left(A_{i,j} - \frac{k_i k_j}{2m} \right) = \frac{1}{2m} \left(L_u^{in} - \frac{(L_u^{tot})^2}{2m} \right), \tag{2.4}$$

де L_u^{in} – вага всіх посилань у спільноті u (для незваженого графа), що

дорівнює подвоєній кількості ребер, обидва кінці яких належать спільноті;

$L_u^{tot} = \sum_{i \in C_u} k_i$ – загальна вага всіх вузлів, включених до спільноти C_u.

Модулярність мережі $Q(G)$ визначається як сума значень модулярності всіх спільнот, що входять до її складу. Завдяки цьому показник модулярності є простим для обчислення та адитивним відносно окремих спільнот. Саме ці властивості зумовили його широке використання як критерію оцінювання якості розбиття вузлів мережі на спільноти.

Водночас слід зазначити, що розв'язання оберненої задачі – знаходження такого розбиття множини вузлів на спільноти, яке забезпечує максимальне значення модулярності, – характеризується високою (субфакторіальною) обчислювальною складністю. Це пояснюється тим, що кількість можливих способів розбиття множини з n елементів на заздалегідь невідому кількість спільнот (тобто необхідністю перебору всіх можливих значень k від 1 до n) надзвичайно швидко зростає:

$$B_n = O\left((n / \log n)^n\right). \tag{2.5}$$

Таким чином, максимізація модулярності мережі $Q(G)$ серед усіх можливих варіантів розбиття методом повного перебору є практично нездійсненною. Варто зазначити, що кількість таких розбиттів, визначена формулою (2.5), відома як число Белла.

Ще однією проблемою використання модулярності є обмежена роздільна здатність [3] алгоритмів, які базуються на її максимізації. Іншими словами, оптимізація модулярності (2.4) має тенденцію до формування спільнот близького розміру, що не завжди відповідає реальній структурі мережі. Для усунення цього недоліку до формули обчислення модулярності спільноти (2.4) вводиться додатковий параметр — коефіцієнт роздільної здатності $\gamma > 0$ (за замовчуванням $\gamma = 1$):

$$q_u = \frac{1}{2m} \sum_{i,j \in C_u} \left(A_{i,j} - \gamma \frac{k_i k_j}{2m} \right) = \frac{1}{2m} \left(L_u^{in} - \gamma \frac{(L_u^{tot})^2}{2m} \right). \quad (2.6)$$

Якщо $\gamma < 1$, алгоритм віддає перевагу більшим спільнотам, інакше – меншим за розміром спільнотам.

З урахуванням (2.6) формула розрахунку модулярності мережі набуває такого вигляду:

$$Q(G) = \sum_{u=1}^{K} q_u = \frac{1}{2m} \sum_{u=1}^{K} \left(L_u^{in} - \gamma \frac{(L_u^{tot})^2}{2m} \right). \quad (2.7)$$

Слід зазначити, що використання значення $\gamma \neq 1$ змінює статистичний зміст показника модулярності, оскільки він перестає характеризувати відхилення фактичної кількості зв'язків між вузлами, що належать до однієї спільноти, від їх очікуваної кількості.

2.2 Розробка генеративної моделі безмасштабних мереж

Нехай мережа представляється у формі неорієнтованого незваженого графа $G(V, E)$. Модель є генеративною (еволюційною), якщо вона дозволяє моделювати покрокове зростання мережі:

$$E\{x_{n+1}\} = x_n + \Delta_n, \quad (2.8)$$

де x – досліджуваний параметр мережі;

Δ_n – його очікуваний приріст на кроці n;

$E\{\cdot\}$ – оператор математичного очікування (зазвичай, для скорочення запису, його опускають; так само буде робитись і в даній квалфікаційній роботі).

Крок *(n)* як правило, ототожнюється з номером вузла. Таким чином,

генеративна модель (2.8) описує динаміку зміни певного параметра *x* мережі зі збільшенням кількості вузлів. Найчастіше таким параметром виступає ступінь вузла.

Модель (2.8) є різницевим рівнянням першого порядку, а її використання як узагальненої форми генеративної моделі мережі є одним із ключових припущень цього дослідження.

Для того щоб залежність очікуваного значення досліджуваного параметра x від n мала степеневий характер, необхідно, щоб приріст Δ_n у виразі (2.8) задовольняв таку умову:

$$\Delta_n = x_n \cdot \frac{a}{n+b}. \tag{2.9}$$

Залежність (2.9) має нульову глибину пам'яті, тобто Δ_n залежить лише від поточного значення x_n та не залежить від передісторії досліджуваного параметру x. Це є суттєвим, проте стандартним, припущенням під час побудови моделей мереж.

З (2.8) та (2.9) витікає рівняння динаміки параметру x:

$$x_{n+1} = x_n + x_n \cdot \frac{a}{n+b}, \tag{2.10}$$

розв'язком якого за початковою умовою x_k є:

$$x_n = x_k \cdot \frac{\Gamma(k+b)}{\Gamma(k+a+b)} \cdot \frac{\Gamma(n+a+b)}{\Gamma(n+b)}, \tag{2.11}$$

де $\Gamma(y)$ – гамма-функція Ейлера; для натуральних чисел m, $\Gamma(m+1)=m!$

Варто зазначити, що залежність (2.11) часто називають степеневою, оскільки при $n \to \infty$ виконується співвідношення $x_n \sim n^a$. При цьому параметр a інтерпретується як показник скейлінгу. Однак таке трактування пов'язане не з наближеним характером залежності (2.11) або недосконалістю

моделі (2.8)–(2.10), а з використанням термінів «степенева залежність» і «степеневий розподіл», які є природними для неперервних функцій часу *x(t)*, тоді як у цьому випадку розглядаються дискретні часові послідовності x_n. З огляду на це залежності виду (2.11) доцільніше називати «дискретно-степеневими».

Важливим окремим випадком моделі (2.8)–(2.11) є широко відома модель переважного приєднання та її найпростіша реалізація — модель Barabási–Albert (BA-модель) [7], яка описує процес зростання ступеня вузлів мережі. Саме BA-модель використовується як основа моделі, що розробляється. При цьому приймаються припущення, характерні для моделей переважного приєднання: кожен новий вузол з'єднується з уже існуючими за допомогою $m>1$ ребер, а початковою структурою є повний граф K_m, усі вузли якого належать до однієї (первинної) спільноти. За таких умов загальна кількість зв'язків у мережі з $n>m$ вузлів визначається як:

$$\mathrm{vol}(G_n) = m(m-1) + 2m(n-m) = 2m(n-n'), \quad \text{where} \quad n' = \frac{m+1}{2}. \quad (2.12)$$

Відповідно до принципу переважного приєднання ймовірність того, що вільний кінець новоствореного ребра буде приєднаний до вузла *i* мережі *G*, яка містить *n* вузлів, є пропорційною ступеню цього вузла $deg_{i,n}$ на кроці *n*:

$$\pi_{i,n} = \frac{deg_{i,n}}{\mathrm{vol}(G_n)}. \quad (2.13)$$

Одним із ключових припущень BA-моделі є те, що приєднання кожного з *m* новостворених ребер розглядається як незалежна подія. За такого припущення, з урахуванням співвідношення (2.12) очікуваний приріст ступеня вузла *i* визначається як:

$$\Delta_{i,n} = m \cdot \pi_{i,n} = m \cdot \frac{deg_{i,n}}{\mathrm{vol}(G_n)} = \frac{deg_{i,n}}{2(n-n')}. \quad (2.14)$$

Базове рівняння динаміки зростання ступенів вузлів у ВА-моделі має вигляд:

$$deg_{i,n+1} = deg_{i,n} + \frac{deg_{i,n}}{2(n-n')}, \quad (2.15)$$

з початковими умовами $deg_{i,i} = m$ for $i > m$, $deg_{i,m} = m$ for $i \leq m$.

Згідно з моделлю (2.12)–(2.15), яка є частковим випадком більш загальної моделі (2.8)–(2.11), очікувані значення ступенів вузлів мережі дорівнюють

$$deg_{i,n} = m \cdot \frac{\Gamma(i-n') \cdot \Gamma(n-n'+1/2)}{\Gamma(i-n'+1/2) \cdot \Gamma(n-n')}, \quad i \geq m. \quad (2.16)$$

Залежність (2.16) є дискретно-ступеневою, при великих i та n наближається до звичайної ступеневої $deg_{i,n} \sim \sqrt{n/i}$, тобто має показник рангового скейлінгу ρ=1/2.

Придатність моделі ВА (2.12) – (2.16) для моделювання кожної із спільнот гомофільної безмасштабної мережі (із збереженням значення скейлінгу $\rho = 1/2$) приймається як гіпотеза дослідження.

2.3 Розробка генеративної моделі гомофільних мереж

Моделі, побудовані на принципі переважного приєднання, зокрема ВА-модель (2.12)–(2.16), не є гомофільними. Проте їх можна використовувати для моделювання окремих спільнот. Припустимо, що під час створення кожному вузлу i присвоюється певний колір (тобто номер спільноти c_i): або один із K уже наявних, або новий колір, який відповідає

новоствореній спільноті $K+1$. У такому разі наступні $m-1$ вузлів також отримують колір $K+1$. Якщо ж новий вузол $n+1$ належить до однієї з уже існуючих спільнот ($c_{n+1} \leq K$), то вибір вузлів для приєднання здійснюється за правилом переважного приєднання, але лише серед вузлів тієї самої спільноти. За таких умов модифіковане правило переважного приєднання (2.13) набуває такого вигляду:

$$\pi_{i,n} = \frac{deg_{i,n}}{L_{k,n}} \cdot \delta(c_i, c_{n+1}), \tag{2.17}$$

де $L_{k,n}$ – сума степенів вузлів кольору k в мережі з n вузлів, тобто обсяг цієї спільноти; в свою чергу, $\sum_k L_{k,n} = \mathrm{vol}(G_n)$.

Якщо кількість вузлів кольору c_{n+1} є меншою за m (що має місце під час приєднання перших m вузлів новоствореної спільноти), то відсутні вузли для приєднання обираються відповідно до загального правила (2.13), тобто без урахування їхнього кольору.

Таким чином, генерація запропонованої моделі гомофільної мережі виконується відповідно до алгоритму, псевдокод якого наведено на рисунку 2.1.

Вхідними параметрами моделі (і відповідного алгоритму на рисунку 2.1) є розмір мережі, кількість ребер, що додаються новим вузлом, а також функції *pnew* та *select_color*. Перша з них визначає ймовірність $p_{K,n}^{new}$ віднесення нового вузла $n+1$ до нової спільноти $K+1$. Функція *select_color* призначає цьому вузлу один з наявних кольорів $k \in \{1, \dots, K\}$, використовуючи ймовірності $p_{k,n+1}$.

Саме значення ймовірностей $p_{K,n}^{new}$, $p_{k,n+1}$ визначають, чи буде згенерована мережа безмасштабною, який показник скейлінгу вона матиме, а також які очікувані значення показників гомофільності (2.1), (2.2) будуть для неї характерними.

```
GenerateBAwithHomophily(nMax, m, pnew, select_color):
G = complete_graph(m)
K = 1 # number of communities
colors[1:m] = 1 # initialize seed nodes color
S[1] = list(1,…,m) # initial community
n = m
while n < nMax-1 :
    with probability pnew(n,K) :
        K = K + 1 # adding new color
        # adding m new nodes n+1, n+2,…,n+m
        for j = 1 to m :
            colors[n+j] = K
            S[K] = [n+j]
            targets = pref_attach(G, m+1-j)
            targets += [n+j]*(j-1)
            # add node and edges to graph
            G.add_node_and_edges(n+j, targets)
            n++
    otherwise :
        # add new node with number n+1
        col = select_color(G,S,K) # choose existing color
        colors[n+1] = col
        S[col].append(n+1)
        targets = pref_attach (S[col], m)
        # add node and edges to graph
        G.add_node_and_edges(n+1, targets)
        n++
return G
```

Рисунок 2.1 – Псевдокод алгоритма генерації гомофільної мережі

Нехай у мережі G_n існує K кольорів (тобто спільнот). Тоді можливі два варіанти розвитку подій: (i) новий вузол стає первістком нової спільноти (ймовірність цього дорівнює $p_{K,n}^{new}$); (ii) новому вузлу призначається один із уже існуючих кольорів.

Згідно з алгоритмом (рисунок 2.1) генерації мережі у випадку (ii), який є основним, застосовується запропоноване модифіковане (2.17) PA-правило. Тоді очікуваний приріст степені вузла i дорівнює:

$$\Delta_{i,n} = m \cdot \pi_{i,n} \cdot p_{c_i,n+1} = m \cdot \frac{deg_{i,n}}{L_{c_i,n}} \cdot p_{c_i,n+1}, \qquad (2.18)$$

де $p_{c_i,n+1}$ – ймовірність того, що вузол $n+1$ отримає колір c_i.

Для того, щоб виконувалась базова умова (2.9) безмасштабності мережі та щоб зберігся показник рангового скейлінгу ВА-мереж ($\rho = 1/2$), необхідно обирати $p_{c_i,n+1}$ так, щоб вираз (2.18) співпав з (2.14). З цієї вимоги витікає, що правило фарбування нового вузла одним з існуючих кольорів (тобто сутність функції *select_color* з рисунку 2.1) має вигляд:

$$p_{k,n+1} = \frac{L_{k,n}}{\mathrm{vol}(G_n)}. \tag{2.19}$$

Таким чином, доки кількість спільнот (кольорів) у мережі не змінюється, спільне застосування правила приєднання (2.17) та правила фарбування (2.19) приводить до того самого виразу (2.13) для очікуваного приросту ступеня вузла i, відповідно, до того самого рівняння динаміки зростання ступенів вузлів (2.15), що й звичайна ВА-модель.

Якщо ж вузол $n + 1$ стає першим елементом нової спільноти (випадок (i)), то послідовно формується не один, а відразу m вузлів цього кольору. Після завершення m ступінь кожного з них дорівнюватиме m. При цьому перший вузол матиме m вільних (тобто не з'єднаних із вузлами своєї спільноти) кінців, другий – $m - 1$, останній – лише один. Усі ці $m(m + 1)/2$ вільних кінців приєднуються до вузлів мережі відповідно до стандартного правила переважного приєднання (2.13). У результаті очікуваний приріст ступеня вузла i після створення m вузлів нової спільноти зменшується порівняно зі звичайним випадком (ii) у $(m + 1)/(2m)$ разів, однак його дискретно-степеневий характер зберігається.

Оскільки поява нової спільноти є відносно рідкісною подією (тобто $K \ll n$), залежність ступеня вузлів від розміру мережі асимптотично наближається до дискретно-степеневої залежності ВА-моделі (2.16). Отже, за умови призначення кольорів вузлам відповідно до правила (2.19), можна вважати, що сформована гомофільна мережа (2.17) є асимптотично безмасштабною за розподілом ступенів вузлів.

Оскільки класичні моделі, побудовані на принципі переважного приєднання, формують мережі, які не характеризуються гомофільністю, їхню властивість безмасштабності розглядають лише стосовно розподілу ступенів вузлів. Водночас під час моделювання структурованих мереж цілком природно аналізувати також розподіл розмірів спільнот $L_{k,n}$ і, відповідно, висувати вимогу щодо його безмасштабності.

Позначимо t_k момент часу, який передує появі кольору k. Оскільки відлік часу синхронізований з кількістю вузлів, то вузли з номерами $t_k + 1, \ldots, t_k + m$ стануть зародком спільноти k.

На проміжку часу між $t_k + m + 1$ до t_{k+1} (тобто після завершення формування зародку кольору k та до появи кольору $k + 1$) прирости степенів всіх вузлів мережі визначаються згідно (2.14). Просумувавши їх вздовж спільнот, можна отримати рівняння динаміки приросту обсягів спільнот:

$$L_{j,n+1} = L_{j,n} + 2m \cdot p_{j,n+1} = L_{j,n} + \frac{L_{j,n}}{n - n'}, \qquad \begin{matrix} j = 1,\ldots,k, \\ n = t_k + m + 1,\ldots,t_{k+1}. \end{matrix} \tag{2.20}$$

Рівняння (2.20) задовольняє загальній моделі (2.10) (при *a=1, b=–n'*), тому його розв'язок, згідно (2.11), має вигляд:

$$L_{j,n} = L_{j,t_k+m} \cdot \frac{n - n'}{t_k + m - n'}. \tag{2.21}$$

Як було показано вище, під час формування наступної спільноти (з номером $k + 1$), $m(m + 1)/2$ кінців ребер приєднуються до попередніх спільнот за модифікованим правилом переважного приєднання (2.17). Таким чином, обсяги цих спільнот на кроці $t_{k+1} + m$ становитимуть

$$L_{j,t_{k+1}+m} = L_{j,t_{k+1}} + \frac{m(m+1)}{2} \cdot p_{j,t_{k+1}+1} = L_{j,t_{k+1}} \left(1 + \frac{m+1}{4(t_{k+1} - n')} \right). \tag{2.22}$$

З урахуванням (2.21) обсяги спільнот $j = 1, \dots k$ на момент завершення формування кольору $k + 1$ дорівнюватимуть:

$$L_{j,t_{k+1}+m} = L_{j,t_k+m} \cdot u(k), \quad j = 1,\dots,k, \tag{2.23}$$

де

$$u(k) = \left(1 + \frac{m+1}{4(t_{k+1} - n')}\right) \cdot \frac{t_{k+1} - n'}{t_k + m - n'} = \frac{t_{k+1} - n'/2}{t_k + m - n'}. \tag{2.24}$$

Обсяги зародків відразу після завершення їх формування дорівнюють:

$$L_{j,t_j+m} = 2m^2 - m(m+1)/2 = m(3m-1)/2. \tag{2.25}$$

Розгортаючи рівняння (2.23), та враховуючи (2.21) та (2.25), знайдемо обсяг спільноти $j \le k$ у довільний момент часу $n = t_k + m + 1, \dots, t_{k+1}$:

$$L_{j,n} = \frac{m(3m-1)}{2} \cdot \frac{n - n'}{t_k + m - n'} \cdot \prod_{i=j}^{k-1} u(i). \tag{2.26}$$

Тепер можна сформулювати умову безмасштабності спільнот відносно їх обсягу. Згідно з (2.9) співвідношення між обсягами послідовних спільнот має задовольняти умові:

$$\frac{L_{j+1,n} - L_{j,n}}{L_{j,n}} = \frac{-\alpha}{j + \alpha + \delta}, \tag{2.27}$$

для деяких $\alpha > 0,\ \delta > -1$.

Підставивши (2.26) у (2.27), отримаємо:

$$\frac{L_{j+1,n}-L_{j,n}}{L_{j,n}}=\frac{1-u(j)}{u(j)}=\frac{-\alpha}{j+\alpha+\delta}\quad\Rightarrow\quad u(j)=\frac{j+\delta+\alpha}{j+\delta}. \qquad (2.28)$$

З (2.28) та (2.24) випливає вимога щодо співвідношення між моментами t_k та t_{k+1}:

$$t_{k+1}=t_k+\frac{\alpha(t_k+c_1)}{k+\delta}+c_2,\quad \text{where}\quad c_1=\frac{m-1}{2},\quad c_2=\frac{3m-1}{4}. \qquad (2.29)$$

При цьому значення t_1 слід умовно встановити таким, щоб обсяг першої спільноти на кроці t_1+m був таким же самим, як і для всіх інших спільнот t_k+m, тобто (2.25). З цього випливає, що $t_1=(m+1)/4$. З цією початковою умовою розв'язок різницевого рівняння (2.29) має вигляд:

$$t_k=\begin{cases}-c_1+\dfrac{c_2}{\alpha-1}\left(\dfrac{\Gamma(\delta+1)\cdot\Gamma(k+\delta+\alpha)}{\Gamma(\delta+\alpha)\cdot\Gamma(k+\delta)}-(k+\delta)\right), & \alpha\neq 1,\\ -c_1+c_2(k+\delta)\left(\psi^{(0)}(k+\delta+1)-\psi^{(0)}(\delta+1)\right), & \alpha=1,\end{cases} \qquad (2.30)$$

де $\psi^{(0)}(x)$ – дігамма-функція (логарифмічна похідна від гамма-функції).

Залежності (2.30) при $k\to\infty$ набувають асимптотичного вигляду:

$$t_k\approx\begin{cases}const\cdot\left(k+\delta\right)^{\alpha}, & \alpha>1,\\ c_2(k+\delta)\cdot\ln(k+\delta+1), & \alpha=1,\\ c_2/(1-\alpha)\cdot(k+\delta), & 0<\alpha<1.\end{cases} \qquad (2.31)$$

Вираз (2.31) наочно демонструє, що при $\alpha>1$ послідовність моментів появи нових кольорів асимптотично підпорядковується степеневій залежності від номера кольору. При цьому α визначає показник скейлінгу, тоді як параметр δ задає не лише зсув у степеневих залежностях, а й їх масштаб, тобто характеризує інтенсивність появи нових кольорів порівняно зі швидкістю зростання кількості вузлів мережі.

Мірою інтенсивності появи кольорів є різниця часу між моментами появи наступного та поточного кольорів. Вона дорівнює:

$$\Delta t_k = t_{k+1} - t_k = \begin{cases} \dfrac{c_2}{\alpha - 1}\left(\dfrac{\alpha \cdot \Gamma(\delta+1) \cdot \Gamma(k+\delta+\alpha)}{\Gamma(\delta+\alpha) \cdot \Gamma(k+\delta+1)} - 1 \right), & \alpha \neq 1, \\ c_2\left(\psi^{(0)}(k+\delta+1) - \psi^{(0)}(\delta+1) + 1 \right), & \alpha = 1, \end{cases} \tag{2.32}$$

або асимптотично (при $k \to \infty$):

$$\Delta t_k \approx \begin{cases} const \cdot (k+\delta)^{\alpha-1}, & \alpha > 1, \\ c_2 \cdot \ln(k+\delta) \quad , & \alpha = 1, \\ c_2 / (1-\alpha) \quad , & 0 < \alpha < 1. \end{cases} \tag{2.33}$$

Графіки Δt_k (2.32) для $k = 1, \dots 20,\ m = 5$ та різних значень α та δ наведені на рисунку 2.2, рисунку 2.3.

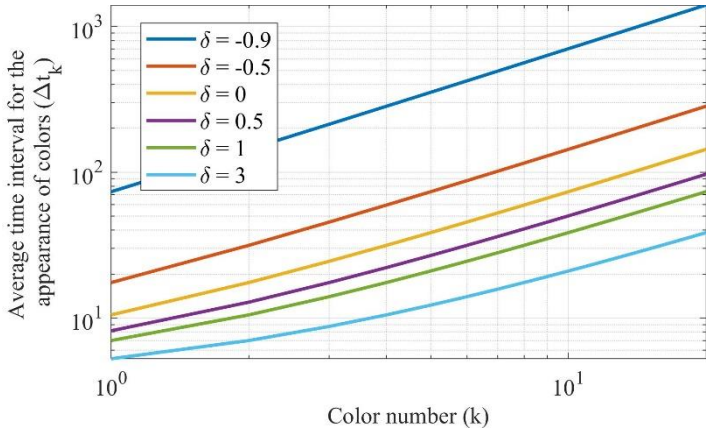


Рисунок 2.2 – Залежність проміжків часу між моментами появи спільнот від їх порядкового номеру для $m = 5,\ \alpha = 2$

Процес виникнення нового кольору відповідає схемі Бернуллі – послідовність спроб до першої вдалої. Тоді з (2.32), (2.33) випливає, що при $\alpha > 1$ ймовірність виникнення нового кольору на кроці n є асимптотично степеневою:

$$p_{k,n}^{new} = \frac{1}{\Delta t_k} \approx const \cdot \left(k + \delta\right)^{1-\alpha}, \tag{2.34}$$

а у випадку $\alpha < 1$ кольори виникають з асимптотично постійною ймовірністю, яка дорівнює $p_{k,n}^{new} \approx \left(1 - \alpha\right) / c_2$.

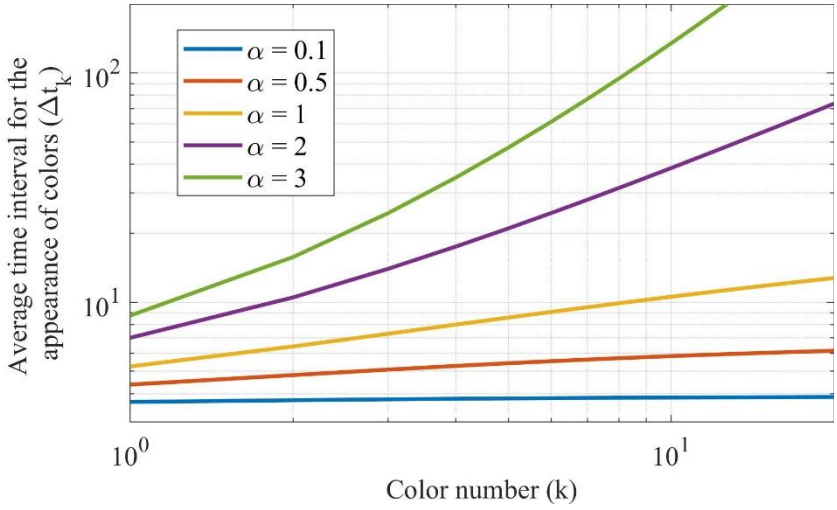


Рисунок 2.3 – Залежність проміжків часу між моментами появи спільнот від їх порядкового номеру для $m = 5,\ \delta = 1$

2.4 Оцінка модулярності модельованої мережі

Згідно з (2.28), (2.29) вираз (2.26), який описує кількість зв'язків у спільноті $j \le k$ у довільний момент часу $t_k + m < n \le t_{k+1}$, можна представити у формі:

$$L_{j,n} = c_2 \frac{2m(n - n')}{t_k + c_1} \cdot \frac{\Gamma(k + \delta + \alpha) \cdot \Gamma(j + \delta)}{\Gamma(k + \delta) \cdot \Gamma(j + \delta + \alpha)}. \tag{2.35}$$

З цього випливає, що відносні обсяги спільнот дорівнюють:

$$e_{j,n} = \frac{L_{j,n}}{\mathrm{vol}(G_n)} = \frac{c_2}{t_k + c_1} \cdot \frac{\Gamma(k + \delta + \alpha) \cdot \Gamma(j + \delta)}{\Gamma(k + \delta) \cdot \Gamma(j + \delta + \alpha)}. \tag{2.36}$$

Легко бачити, що через саму конструкцію алгоритму побудови гомофільної безмасштабної мережі (рисунок 2.1) відносні обсяги спільнот (2.36) одночасно є ймовірностями $p_{j,n+1}$ (2.19) отримання новим вузлом кольору j.

У разі, якщо $\alpha > 1$, то зазначені обсяги/ймовірності асимптотично збігаються до

$$e_j^{as} = (\alpha - 1) \cdot \frac{\Gamma(\delta + \alpha) \cdot \Gamma(j + \delta)}{\Gamma(\delta + 1) \cdot \Gamma(j + \delta + \alpha)}. \tag{2.37}$$

Зокрема, відносний обсяг першої (та найбільшої) спільноти становить $e_1^{as} = \frac{\alpha - 1}{\alpha + \delta}$. З цього витікає, що задаючи вагу першої спільноти e_1^{as} та показник скейлінгу обсягів спільнот $\alpha > 1$, можна визначити параметр δ:

$$\delta = \frac{\alpha - 1}{e_1^{as}} - \alpha. \tag{2.38}$$

Оскільки кожна спільнота, крім першої, з'єднується $m(m+1)/2$ ребрами з вузлами попередніх спільнот (і це є єдиною причиною виникнення міжспільнотних зв'язків), то відносна кількість таких зв'язків у мережі дорівнює

$$E_n^{ext} = \frac{(k-1)m(m+1)}{\mathrm{vol}(G_n)} = \frac{(k-1)(m+1)}{2(n-n')}. \tag{2.39}$$

Таким чином, модулярність (2.1) модельної мережі становить:

$$\mu(G_n) = 1 - \frac{(k-1)(m+1)}{2(n-n')} - \sum_{j=1}^{k} e_{j,n}^2. \tag{2.40}$$

На жаль, у випадку, коли α не є цілим числом, сума квадратів

відносних обсягів спільнот не має аналітичного виразу через елементарні математичні функції (хоча досить громіздким чином виражається через гіпергеометричну функцію ${}_3F_2$).

Графічний вигляд залежності (2.40) для деяких наборів параметрів моделі показаний на рисунку 2.4, рисунку 2.5.

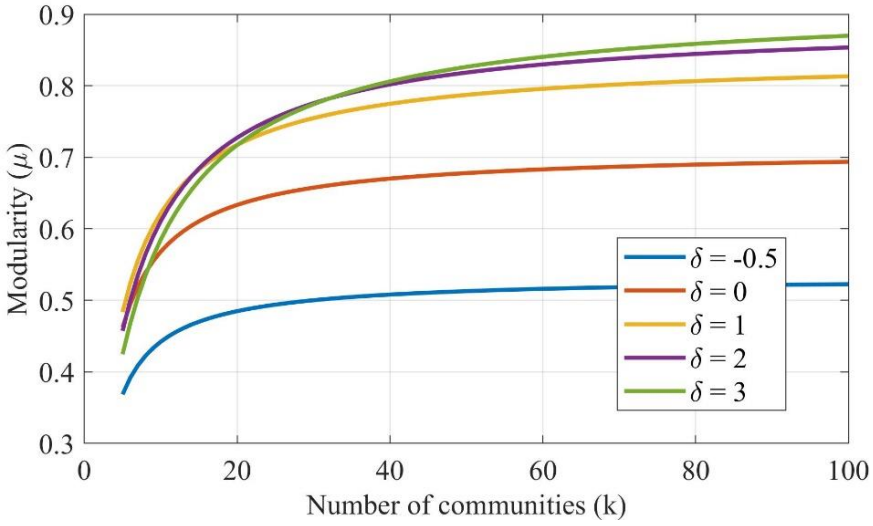


Рисунок 2.4 – Залежність модулярності мережі від кількості спільнот при $\alpha = 3$ та варіюванні δ

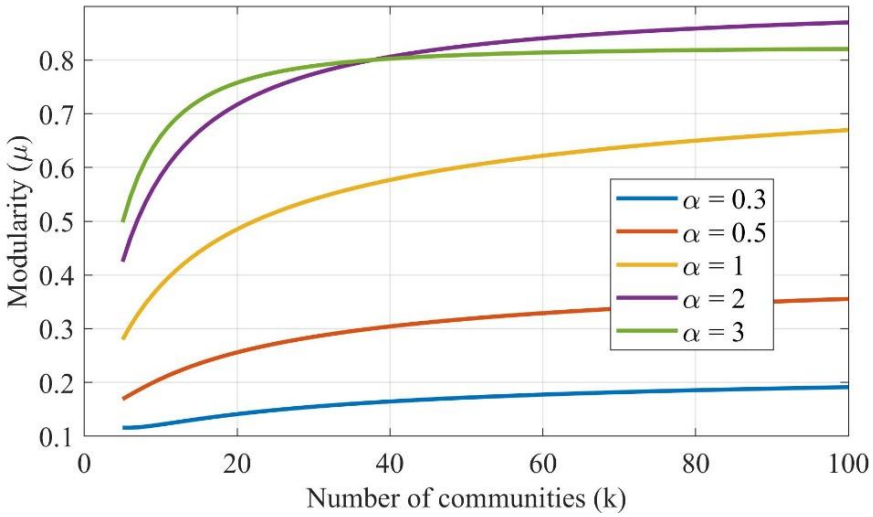


Рисунок 2.5 – Залежність модулярності мережі від кількості спільнот при $\delta = 0$ та варіюванні α

Обчислення модифікованої модулярності (2.2) пов'язане з аналогічною проблемою, оскільки розміри спільнот n_j асимптотично підкоряються дискретно-степеневому закону, подібному до (2.36), (2.37). Унаслідок цього підсумовування добутків $n_{j,n}e_{j,n}$ також приводить до виникнення гіпергеометричних функцій.

Попри громіздкість залежності (2.40), наведені графіки дають змогу зробити висновок, що значення модулярності зростає зі збільшенням кількості спільнот (k) у мережі. Якщо ж значення k фіксоване, то модулярність зростає зі збільшенням параметрів α та δ.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ РОЗРОБЛЕНОЇ МОДЕЛІ, ПРОВЕДЕННЯ ЧИСЕЛЬНОГО ЕКСПЕРИМЕНТУ

3.1 Інструменти програмної реалізації моделі

Для реалізації програмної частини проєкту було обрано мову програмування Python. Python є високорівневою інтерпретованою мовою програмування загального призначення, яка підтримує об'єктно-орієнтовану, процедурну та функціональну парадигми програмування. Завдяки простому й зрозумілому синтаксису, динамічній типізації та широким можливостям стандартної бібліотеки ця мова широко застосовується для розробки програмного забезпечення, аналізу даних, машинного навчання та наукових обчислень.

Однією з основних переваг Python є наявність великої кількості бібліотек і фреймворків, що значно спрощують реалізацію складних алгоритмів та скорочують час розробки програмних продуктів. Підтримка модульної структури програм, повторного використання коду та інтеграції із зовнішніми бібліотеками забезпечує високу гнучкість і масштабованість програмних рішень.

Python має ефективні вбудовані структури даних, підтримує автоматичне керування пам'яттю та забезпечує високу продуктивність розробки. Інтерпретований характер мови дозволяє швидко тестувати окремі модулі та оперативно вносити зміни до програмного коду. Крім того, Python легко інтегрується з компонентами, реалізованими мовами C та C++, що дає змогу використовувати високопродуктивні обчислювальні бібліотеки без втрати зручності розробки.

Важливою перевагою Python є також багата екосистема спеціалізованих бібліотек для роботи з графами, математичними обчисленнями, статистичним аналізом і візуалізацією даних. Це робить мову особливо придатною для реалізації моделей складних мереж,

дослідження їхніх властивостей та проведення експериментального аналізу. Завдяки кросплатформеності, активній підтримці спільноти розробників і регулярному оновленню програмних засобів Python є одним із найбільш популярних інструментів для виконання наукових досліджень та створення прикладних систем.

Основною бібліотекою, використаною під час реалізації програмного забезпечення, стала NetworkX. Вона призначена для створення, аналізу та дослідження графів і складних мереж, а також широко застосовується при моделюванні мережевих структур та дослідженні їхніх властивостей.

Бібліотека NetworkX підтримує роботу з різними типами графів, зокрема неорієнтованими, орієнтованими, мультиграфами та зваженими графами. Її універсальна архітектура дозволяє використовувати як вузли практично будь-які об'єкти Python, що забезпечує гнучкість під час побудови математичних моделей.

До важливих можливостей бібліотеки належать засоби імпорту та експорту графів у найбільш поширених форматах, генерація типових моделей мереж, а також широкий набір алгоритмів для аналізу їхньої структури. Зокрема, NetworkX містить методи пошуку підграфів, клік, k-ядер (K-core), зв'язних компонент та інших структурних елементів мережі.

Окремою перевагою бібліотеки є наявність великої кількості інструментів для обчислення структурних характеристик графів. За її допомогою можна визначати ступені вузлів, діаметр і радіус графа, довжини найкоротших шляхів, центральність вузлів, коефіцієнти посередництва, асортативність, модулярність та інші показники, що використовуються під час аналізу складних мереж. Саме тому NetworkX було обрано як основний інструмент для реалізації моделі гомофільної мережі та проведення експериментальних досліджень.Наприклад, програмна реалізація пошуку центральності вузлів за близкістю виглядає наступним чином.

Імпортуємо бібліотеку:

```
import networkx as nx
```

За допомогою конструктору класу Graph() створюємо потрібні екземпляри:

```
G = nx.Graph()
G_1_2 = nx.Graph()
G_1_3 = nx.Graph()
G_2_3 = nx.Graph()
G_1_2_3 = nx.Graph()
```

Граф G – це повний граф, G_1_2, G_1_3, G_2_3 – допоміжні графи з коаліціями {1, 2}, {1, 3}, {2, 3}, G_1_2_3 – граф з шуканою коаліцією.

Для побудови вузлів та прокладання маршрутів між ними скористаємося вбудованими методамим add_edge() (лістинг 3.1). Цей метод приймає до себе назву двух вузлів, та створює ребро між ними.

Лістинг 3.1 – Вбудовані методи add_edge()

```
all_path = "1 4 1 5 1 6 1 7 1 8 2 8 2 12 3 9 3 10 6 11 7 11 8 11 9 11 10 11".split(" ")
coalition_1_2 = "1_2 4 1_2 5 1_2 6 1_2 7 1_2 8 1_2 12 3 9 3 10 6 11 7 11 8 11 9 11 10 11".split(" ")
coalition_1_3 = "1_3 4 1_3 5 1_3 6 1_3 7 1_3 8 1_3 9 1_3 10 2 8 2 12 6 11 7 11 8 11 9 11 10 11".split(" ")
coalition_2_3 = "1 4 1 5 1 6 1 7 1 8 2_3 8 2_3 9 2_3 10 2_3 12 6 11 7 11 8 11 9 11 10 11".split(" ")
coalition_1_2_3 = "1_2_3 4 1_2_3 5 1_2_3 6 1_2_3 7 1_2_3 8 1_2_3 9 1_2_3 10 1_2_3 12 6 11 7 11 8 11 9 11 10 11".split(" ")
def func_add_edge(G, path):
    a = 0
    while a < len(path) - 1:
        G.add_edge(path[a], path[a + 1])
        a += 2
func_add_edge(G, all_path)
func_add_edge(G_1_2, coalition_1_2)
func_add_edge(G_1_3, coalition_1_3)
func_add_edge(G_2_3, coalition_2_3)
func_add_edge(G_1_2_3, coalition_1_2_3)
```

3.2 Моделювання гомофільних безмасштабних мереж

Програмну реалізацію розроблених моделей мереж і методів оцінювання їхньої гомофільності виконано мовою програмування Python. Для візуалізації згенерованих мереж використовувалося програмне забезпечення Gephi.

З метою підтвердження властивостей запропонованої моделі щодо безмасштабності та гомофільності було реалізовано програмний прототип, основу якого становить алгоритм, наведений на рисунку 2.1. Призначення існуючого кольору новому вузлу (функція `select_color`) здійснювалося відповідно до (2.19), тоді як ймовірність появи нового кольору (функція `pnew`) обчислювалась за формулами (2.32) та (2.34).

Для перевірки властивостей моделі було згенеровано мережу, що містить $n = 10000$ вузлів, при розмірі початкової спільноти $m = 5$. Отримані результати (рисунок 3.1) показали, що ранговий розподіл вузлів за ступенями збігається з відповідним розподілом BA-моделі, тобто практично не залежить від параметрів структури спільнот. Як і передбачалося теоретично, цей розподіл має дискретно-степеневий характер із показником скейлінгу $\rho = 1/2$.

Для дослідження безмасштабності розподілу спільнот було згенеровано `nSamples=10` мереж із параметрами $\alpha = 4\, n = 10^4, m = 5, \alpha = 4, \delta = 0$. Рангові розподіли обсягів і розмірів спільнот, тобто кількості зв'язків і вузлів відповідно, наведено на рисунках 3.1 та 3.2.

На рисунку 3.3 представлено ранговий розподіл усереднених розмірів спільнот у згенерованих безмасштабних гомофільних мережах залежно від значення показника скейлінгу α (при $n = 10^4, m = 5, \alpha = 4, \delta = 0$).

Аналіз отриманих результатів свідчить, що розподіли, наведені на рисунках 3.2 та 3.3, добре узгоджуються з теоретично прогнозованими дискретно-степеневими законами (2.35).

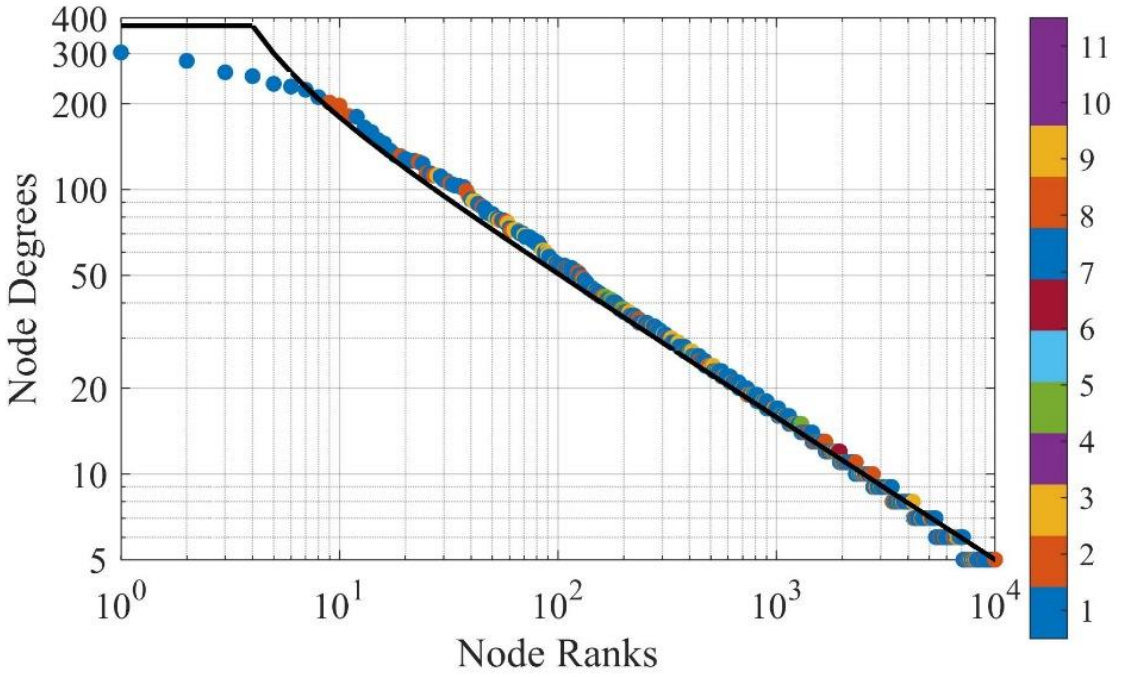
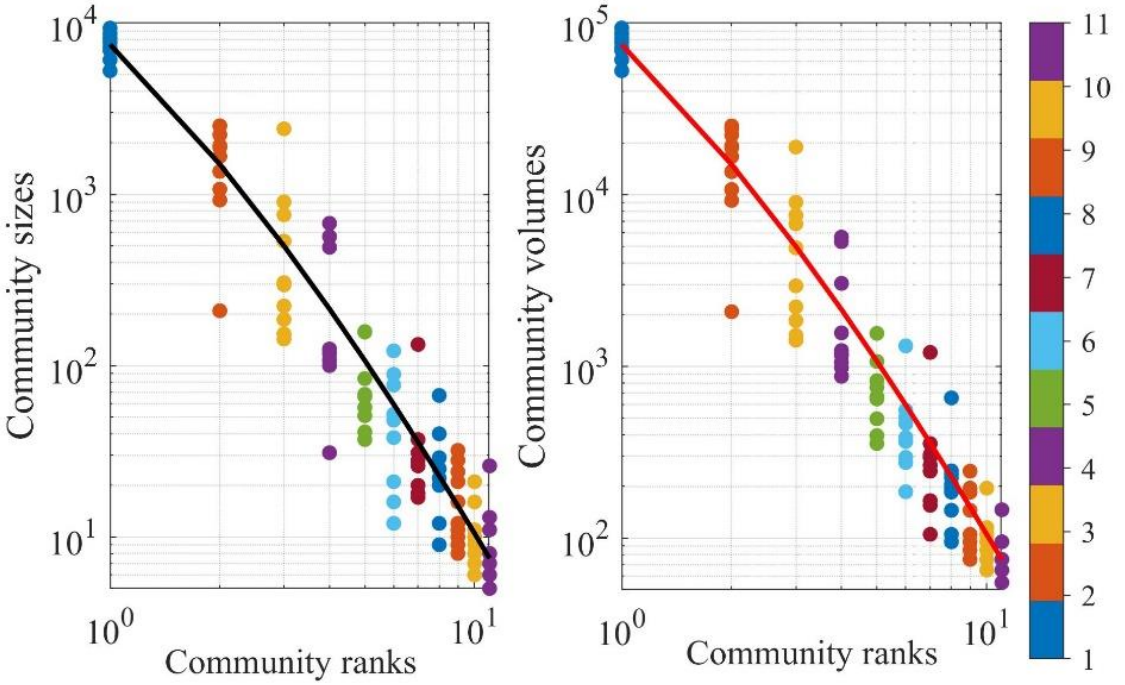


Рисунок 3.1 – Ранговий розподіл степенів вузлів гомофільної мережі



а) б)

Рисунок 3.2 – Рангові розподіли спільнот безмасштабної гомофільної мережі: *а* – за розміром спільнот; *б* – за обсягом спільнот

Додатково було виконано оцінювання модулярності згенерованих модельних мереж. Отримані значення перебувають у діапазоні від 0,4 до 0,8, що повністю узгоджується з чисельними оцінками, наведеними на рисунку 2.5.

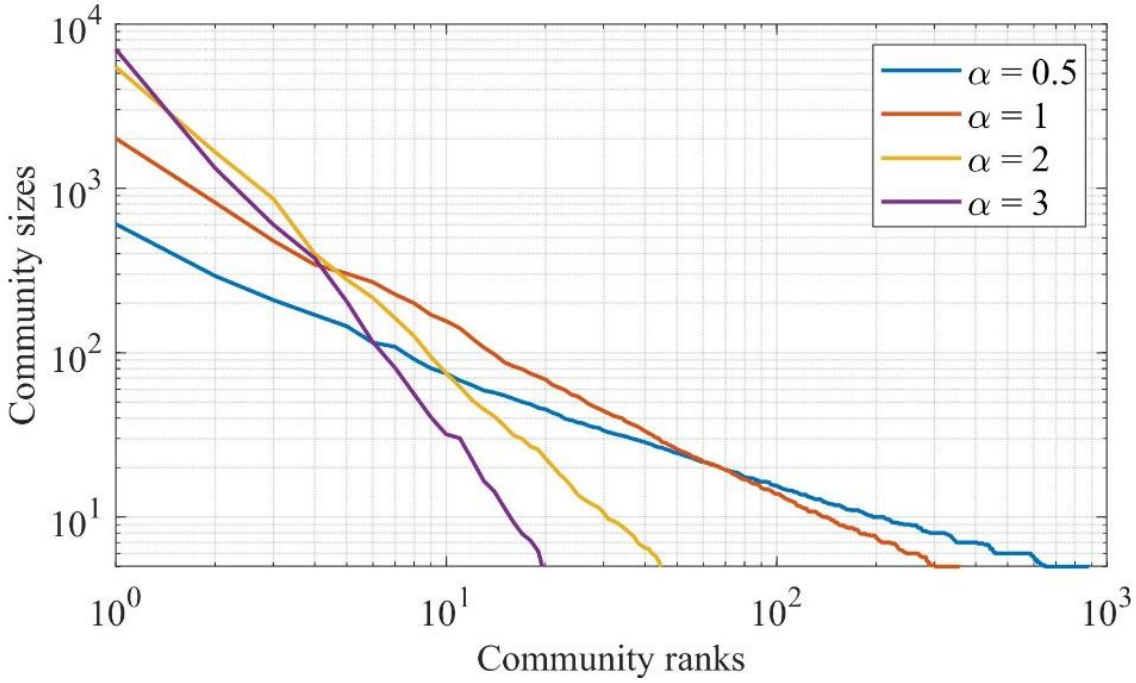


Рисунок 3.3 – Ранговий розподіл усередненого розміру спільнот безмасштабних гомофільних мереж при варіюванні показника скейлінгу

Для демонстрації структури сформованої мережі було згенеровано окремий приклад із $n = 500$ вузлами при параметрах $m = 5, \delta = 0, \alpha = 3$.

Візуалізацію мережі виконано за допомогою програмного забезпечення Gephi, а її результат наведено на рисунку 3.4. Згенерована мережа містить сім кольорових спільнот: найбільша (червона) включає близько 320 вузлів, тоді як дві найменші (зелена та чорна) складаються лише з п'яти вузлів кожна.

На рисунку 3.4 розмір кожного вузла пропорційний його ступеню, значення якого змінюється в межах від 5 до 79.

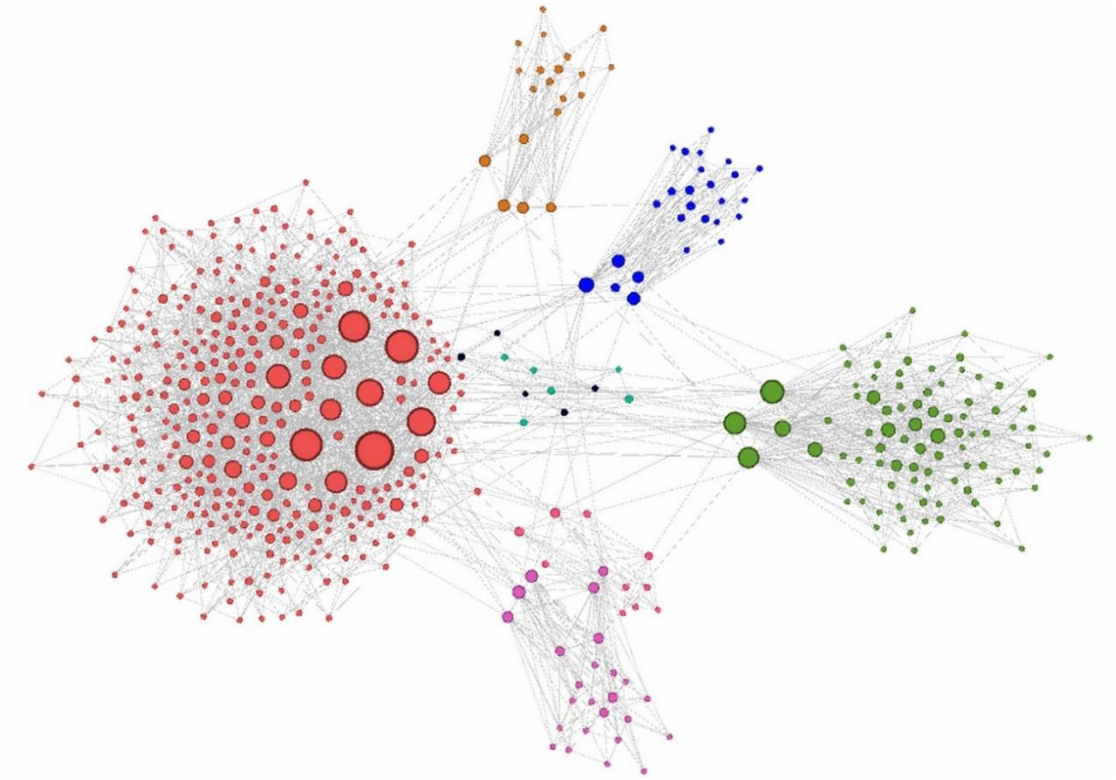


Рисунок 3.4 – Приклад гомофільної безмасштабної мережі з 500 вузлів

Наведені рисунки (рисунок 3.1 – рисунок 3.4) демонструють основні властивості розробленої моделі мереж, а саме гомофільність та безмасштабність відносно степенів вузлів та розмірів спільнот.

ВИСНОВКИ

Розглянуто проблему моделювання безмасштабних мереж з властивістю гомофільності. Лосліджено моделі складних мереж, властивості безмасштабності, асортативності, кластеризації, модулярності та гомофільності. Визначено, що мірою гомофільності мереж є добре відомий та широко застосовуваний коефіцієнт модулярності.

Розроблено алгоритм генерації моделі гомофільних мереж. Зазначений результат є важливим, оскільки запропонований алгоритм забезпечує можливість покрокової генерації гомофільної мережі будь-якого заданого розміру та кількості спільнот.

Доведено, що для одночасного підтримання гомофільності та збереження безмасштабності мережі (відносно степенів вузлів) необхідно відносити новий вузол до тієї чи іншої спільноти у відповідності до правила переважного приєднання. Таким чином, приєднання нового вузла мережі до існуючих відбувається за двокроковим застосуванням цього правила. Доведено, що у зазначеному випадку розподіл степенів вузлів буде ідентичним до розподілу степенів вузлів звичайних (однородних) безмасштабних мереж, тобто таких, які не містять спільнот.

Визначено ймовірності появи нових спільнот, за дотриманням яких генерована мережа є безмасштабною за обсягом та розміром спільнот. Ці ймовірності відповідають очікуваному проміжку часу між появою спільнот, який є асимптотично степеневим відносно поточної кількості спільнот.

Встановлено залежність основного показника гомофільності мереж – модулярності – від параметрів моделі. Вказана залежність не має виразу через елементарні функції, проте досліджена експериментально.

Здійснено програмну реалізацію моделі, перевірено відповідність властивостей мереж, які генеруються, заданим вимогам. Визначено, що генеровані мережі є гомофільними та безмасштабними відносно степенів вузлів та обсягів спільнот.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Albert R., Barabási A.-L. Statistical mechanics of complex networks. *Reviews of modern physics*. 2002. Vol. 74, no. 1. P. 47–97. URL: https://doi.org/10.1103/revmodphys.74.47 (date of access: 05.06.2026).

2. Watts D. J., Strogatz S. H. Collective dynamics of 'small-world' networks. *Nature*. 1998. Vol. 393, no. 6684. P. 440–442. URL: https://doi.org/10.1038/30918 (date of access: 05.07.2026).

3. Newman M. Power laws, Pareto distributions and Zipf's law. *Contemporary physics*. 2005. Vol. 46, no. 5. P. 323–351. URL: https://doi.org/10.1080/00107510500052444 (date of access: 05.07.2026).

4. Dorogovtsev S. N., Mendes J. F. F., Samukhin A. N. Structure of growing networks with preferential linking. *Physical review letters*. 2000. Vol. 85, no. 21. P. 4633–4636. URL: https://doi.org/10.1103/physrevlett.85.4633 (date of access: 05.07.2026).

5. Elastic scale-free networks model based on the mediaton-driven attachment rule / V. Shergin et al. *2020 IEEE third international conference on data stream mining & processing (DSMP)*, Lviv, Ukraine, 21–25 August 2020. 2020. URL: https://doi.org/10.1109/dsmp47368.2020.9204207 (date of access: 05.07.2026).

6. Identifying time dependence in network growth / M. Falkenberg et al. *Physical review research*. 2020. Vol. 2, no. 2. URL: https://doi.org/10.1103/physrevresearch.2.023352 (date of access: 05.07.2026).

7. Meng X., Zhou B. Scale-free networks beyond power-law degree distribution. *Chaos, solitons & fractals*. 2023. Vol. 176. P. 114173. URL: https://doi.org/10.1016/j.chaos.2023.114173 (date of access: 05.07.2026).

8. Shergin V., Chala L. The concept of elasticity of scale-free networks. *2017 4th international scientific-practical conference problems of infocommunications. science and technology (PIC S&T)*, Kharkov, 10–13

October 2017. 2017. URL: https://doi.org/10.1109/infocommst.2017.8246392 (date of access: 05.07.2026).

9. Unravelling the size distribution of social groups with information theory in complex networks / A. Hernando et al. *The european physical journal B*. 2010. Vol. 76, no. 1. P. 87–97. URL: https://doi.org/10.1140/epjb/e2010-00216-1 (date of access: 05.07.2026).

10. Li A., Li J., Pan Y. Homophyly networks – a structural theory of networks. *arXiv*. 2013. URL: https://doi.org/10.48550/arXiv.1310.8295 (date of access: 08.07.2026).

11. Newman M. E. J. Mixing patterns in networks. *Physical review E*. 2003. Vol. 67, no. 2. URL: https://doi.org/10.1103/physreve.67.026126 (date of access: 08.07.2026).

12. Mele A. A structural model of homophily and clustering in social networks. *SSRN electronic journal*. 2017. URL: https://doi.org/10.2139/ssrn.3031489 (date of access: 08.07.2026).

13. Choromański K., Matuszak M., Miękisz J. Scale-free graph with preferential attachment and evolving internal vertex structure. *Journal of statistical physics*. 2013. Vol. 151, no. 6. P. 1175–1183. URL: https://doi.org/10.1007/s10955-013-0749-1 (date of access: 08.07.2026).

14. Noldus R., Van Mieghem P. Assortativity in complex networks. *Journal of complex networks*. 2015. Vol. 3, no. 4. P. 507–542. URL: https://doi.org/10.1093/comnet/cnv005 (date of access: 08.07.2026).

15. Shergin V., Udovenko S., Chala L. Assortativity properties of Barabási-Albert networks. *Data-centric business and applications*. Cham, 2020. P. 55–66. URL: https://doi.org/10.1007/978-3-030-43070-2_4 (date of access: 08.07.2026).

16. Li A., Li J., Pan Y. Community structures are definable in networks, and universal in real world. *arXiv*. 2013. URL: https://doi.org/10.48550/arXiv.1310.8294 (date of access: 08.07.2026).

17. Girvan M., Newman M. E. J. Community structure in social and biological networks. *Proceedings of the national academy of sciences*. 2002. Vol. 99, no. 12. P. 7821–7826. URL: https://doi.org/10.1073/pnas.122653799 (date of access: 08.07.2026).

18. Network community detection using modified modularity criterion / V. Shergin et al. *Eastern-European journal of enterprise technologies*. 2024. Vol. 6, no. 4 (132). P. 6–13. URL: https://doi.org/10.15587/1729-4061.2024.318452 (date of access: 08.07.2026).