

Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

(повна назва)

Кафедра Штучного інтелекту

(повна назва)

АТЕСТАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти другий (магістерський)

(рівень вищої освіти)

МОДЕЛЮВАННЯ ТА АНАЛІЗ МОДЕЛЕЙ МЕРЕЖ, КЕРОВАНИХ

ПРАВИЛОМ ПРИЄДНАННЯ, ЗАСНОВАНИМ НА ПОСЕРЕДНИЦТВІ

(тема)

Виконав:

студент 2 курсу, групи СШМ-18-3

Черніюк Антон Сергійович

(прізвище, ініціали)

Спеціальність 122 – Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми

(освітньо-професійна або освітньо -наукова)

Освітня програма Системи штучного

інтелекту (СШ)

(повна назва освітньої програми)

Керівник проф. Гвоздинський А.М.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

(підпис)

В.О. Філатов

(прізвище, ініціали)

2020 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____

Кафедра _____ Штучного інтелекту _____

Рівень вищої освіти _____ другий (магістерський) _____

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

Тип програми _____
(освітньо-професійна або освітньо-наукова)

Освітня програма _____ Системи штучного інтелекту (СШІ) _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

« _____ » _____ 20 ____ р.

ЗАВДАННЯ

НА АТЕСТАЦІЙНУ РОБОТУ

студентові _____ Черніюку Антону Сергійовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи Моделювання та аналіз моделей мереж, керованих правилом приєднання, заснованим на посередництві

затверджена наказом по університету від _____ р. № _____

2. Термін подання студентом роботи до екзаменаційної комісії _____ 20__ р.

3. Вихідні дані до роботи Науково-технічні публікації, дані Інтернет-джерел щодо безмасштабних мереж та, зокрема, моделей, заснованих на правилі приєднання вузлів за посередництвом

4. Перелік питань, що потрібно опрацювати в роботі Вступ, 1 Огляд предметної області, 1.1 Графи та мережі, 1.2 Визначення та властивості складних мереж, 1.3 Концепції зростання та приєднання, 1.4 Огляд моделей безмасштабних мереж, 1.5 Постановка задачі дослідження, 2 Дослідження моделі приєднання за посередництвом, 2.1 Опис моделі приєднання за посередництвом, 2.2 Динамічні властивості MDA-мережі, 2.3 Оцінювання показника скейлінгу для моделі MDA, 2.4 Види ступеневих розподілів, 3 Числове моделювання MDA-мережі, 3.1 Обґрунтування програмного засобу моделювання, 3.2 Моделювання MDA-мережі, 3.3 Емпіричне дослідження показнику скейлінгу моделі приєднання за посередництвом, Висновки

5. Перелік графічного матеріалу із зазначенням схем, плакатів, комп'ютерних ілюстрацій (слайдів) Слайд 1 – Титульний слайд. Слайд 2 – Об'єкт, предмет, мета досліджень. Слайд 3 – Складні мережі. Слайд 4 – Типові розподіли вузлів мережі за ступенями. Слайд 5 – Правило приєднання нових вузлів. Слайд 6 – Приклади правил приєднання.

Слайд 7 – Еволюція дерева Барабаші-Альберт. Слайд 8 – Дерево Барабаші-Альберт, розподіл Юла-Саймона. Слайд 9 – Постановка задачі дослідження. Слайд 10 – Модель приєднання за посередництвом (MDA). Слайд 11 – Порівняння правил приєднання БА та MDA у разі $m=1$. Слайд 12 – Еволюція мережі MDA у разі $m=1$. Слайд 13 – Еволюція мережі MDA у разі $m=2$. Слайд 14 – Математичне очікування ступенів вузлів у MDA-моделі. Слайд 15 – Блок-схеми алгоритмів функцій. Слайд 16 – Ранговий розподіл вузлів за ступенями при $m=1$. Слайд 17 – Ранговий розподіл вузлів за ступенями при $m=5$. Слайд 18– Розподіл вузлів за віком при $m=100$. Слайд 19– Залежність показника скейлінгу частотного розподілу ступенів вузлів від керуючого параметра m . Слайд 20 – Висновки

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Основний розділ	проф. Гвоздинський А.М.		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на дипломну роботу		
2	Аналіз предметної області		
3	Постановка завдання та узгодження з керівником		
4	Аналіз моделей графів та мереж, числових характеристик		
5	Дослідження правил приєднання нових вузлів до мереж		
6	Дослідження моделі мереж, заснованих на правилі приєднання через посередництво		
7	Експериментальні дослідження моделей мереж		
8	Написання пояснювальної записки		
9	Попередній захист		
10	Захист перед ЕК		

Дата видачі завдання _____ 2020 р.

Студент _____ Черніюк А.С.

(підпис)

Керівник роботи _____ проф. Гвоздинський А.М.

(підпис)

(посада, прізвище, ініціали)

РЕФЕРАТ

Записка пояснювальна: 66 с., 17 рис., 3 дод., 25 джерел.

ПЕРЕЛІК КЛЮЧОВИХ СЛІВ: MDA-модель; безмасштабні мережі; показник скейлінгу; преференційне приєднання; приєднання за посередництвом; розподіл Юла-Саймона; ступеневий розподіл.

Об'єкт дослідження – безмасштабні мережі.

Предмет дослідження – модель безмасштабних мереж з приєднанням за посередництвом.

Мета роботи – аналіз моделей складних мереж, дослідження правил приєднання вузлів, визначення показнику скейлінгу, програмна реалізація мереж, заснованих на правилі приєднання вузлів за посередництвом, проведення експериментальних досліджень згенерованих мереж щодо перевірки безмасштабності та визначення показника скейлінгу.

Методи дослідження – математичне моделювання, програмна реалізація, імітаційне моделювання мереж.

РЕФЕРАТ

Пояснительная записка: 66 с., 17 рис., 3 прил., 25 источников.

ПЕРЕЧЕНЬ КЛЮЧЕВЫХ СЛОВ: MDA-модель; масштабно-инвариантные сети; показатель скейлинга; предпочтительное присоединение; присоединение-через-посредника; распределение Юла-Саймона; степенное распределение.

Предмет исследования – масштабно-инвариантные сети.

Объект исследования – модель масштабно-инвариантных сетей с присоединением-через-посредника.

Цель работы – анализ моделей сложных сетей, исследование правил присоединения узлов, определение показателя скейлинга, программная реализация сетей, основанных на правиле присоединения узлов через посредника, проведение экспериментальных исследований сгенерированных сетей по проверке масштабной инвариантности и определению показателя скейлинга.

Методы исследования- математическое моделирование, программная реализация, имитационное моделирование сетей.

ABSTRACT

Explanatory note: 66 p., 17 fig., 3 ann., 25 sources.

KEYWORDS: MDA-model; mediator-driven attachment rule; power distribution; preferential attachment; scale-free networks; scaling factor; Yule-Simon distribution.

Subject – scale-free networks.

Object – model of scale-free networks with mediator-driven attachment rule.

The purpose of the work are complex network models analysis, studies the attachment rules for nodes, analytical determination the scaling factor, program implementation of networks based on the MDA-rule, numerical studies of the generated networks to verify its scale invariance and determine the scaling factor.

Methods – mathematical modeling, software implementation, numerical simulation of networks.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ ТА ТЕРМІНІВ	8
ВСТУП	9
1 ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ	10
1.1 Графи та мережі	10
1.2 Визначення та властивості складних мереж	13
1.3 Концепції зростання та приєднання.....	17
1.4 Огляд моделей безмасштабних мереж.....	18
1.5 Постановка задачі дослідження.....	22
2 ДОСЛІДЖЕННЯ МОДЕЛІ ПРИЄДНАННЯ ЗА ПОСЕРЕДНИЦТВОМ	23
2.1 Опис моделі приєднання за посередництвом.....	23
2.2 Динамічні властивості MDA-мережі	28
2.3 Оцінювання показника скейлінгу для моделі MDA.....	31
2.4 Види ступеневих розподілів	33
3 ЧИСЛОВЕ МОДЕЛЮВАННЯ MDA-МЕРЕЖІ	40
3.1 Обґрунтування програмного засобу моделювання	40
3.2 Моделювання MDA-мережі.....	46
3.3 Емпіричне дослідження показнику скейлінгу моделі приєднання за посередництвом	51
ВИСНОВКИ	55
ПЕРЕЛІК ПОСИЛАНЬ	56
ДОДАТОК А	59
ДОДАТОК Б	62
ДОДАТОК В	64

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ ТА ТЕРМІНІВ

MDA – приєднання, засноване на посередництві (mediator-driven attachment);

SF-мережа – scale-free network (безмасштабна мережа);

БА – мережа Барабаші-Альберт;

MIM – масштабно-інваріантна мережа;

ПП – переважне приєднання (переважне зв'язування).

ВСТУП

Теорія складних мереж є відносно новою науковою сферою, яка вивчає складні мережі. Більшість мереж реального світу є безмасштабними, тобто такими, що розподіл вузлів за кількістю зв'язків є ступеневим, або наближається до ступеневого. Наприклад, це стосується соціальних мереж, мереж співпраці, мереж цитування, всесвітньої павутини, нейронних та білкових мереж, деяких транспортних мереж тощо. Таким чином, дослідження безмасштабних мереж має суттєву практичну цінність.

Наразі існує багато моделей безмасштабних мереж, проте жодна з них не є універсальною, що зумовлює актуальність проблеми, яка досліджується. З іншого боку, ступеневий розподіл параметру, в даному випадку - кількості зв'язків вузлів, є характерною особливістю фрактальних властивостей системи та її здатності до самоорганізації. Це ще одна причина теоретичного інтересу до безмасштабних мереж.

В основі більшості моделей безмасштабних мереж лежить модель Барабаші-Альберт. Вона ґрунтується на двох основних ідеях: концепції зростання та концепції переважного приєднання. Хоча в даний час існує багато специфікацій та узагальнень для генеративного механізму, що використовується для створення моделей складних мереж, їхні загальні риси успадковані саме від цієї моделі.

Однією з ключових властивостей переважної більшості моделей є потреба в доступності інформації щодо властивостей всіх вузлів мережі, зокрема, значень ступенів вузлів. В той же час для переважної більшості мереж реального світу використання такої інформації неможливе через її великий обсяг, конфіденційність та ін. Тому більш перспективними є моделі приєднання нових вузлів за посередництвом, у яких статистика існуючих вузлів не використовується явним чином під час додавання нових. Властивості однієї з таких моделей – моделі MDA – є предметом досліджень у атестаційній роботі магістра.

1 ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Графи та мережі

Графом $G(V, L)$ називають математичний об'єкт, який представляє собою впорядковану пару: множину V вершин (вузлів), та множину L зв'язків (з'єднань) між ними. Термін "граф" був вперше введений американським математиком Джеймсом Сільвестром в 1878, хоча основи теорії графів було закладено Леонардом Ейлером ще в 1736 (задача про Кенігсбергські мости).

Існує багато різновидів та, відповідно, ознак класифікації графів.

Поперше, графи бувають орієнтовані (спрямовані) та неорієнтовані. Відповідно, з'єднання графічно відображаються дугами зі стрілками, або ребрами, тобто лініями без стрілок (рис. 1.1).

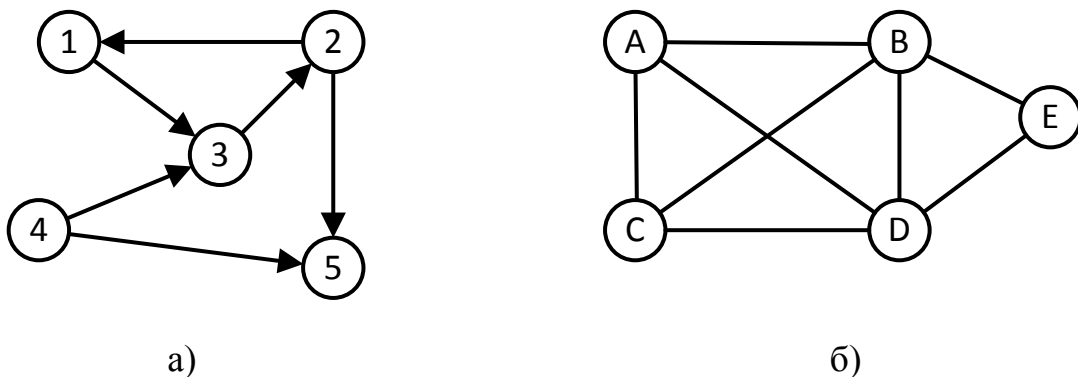


Рисунок 1.1 – Приклади графів: а) орієнтований; б) неорієнтований

Важливою характеристикою графу є його розмір, тобто кількість вузлів $n = |V|$. Відповідно, графи є скінченими та нескінченими. Під нескінченістю в даному випадку розуміється те, що немає такого числа n_{\max} , щоб $n < n_{\max}$. Під час дослідження чи моделювання граф, що оброблюється (який завжди скінчений) розглядається як підграф загального нескінченного графа.

Найчастіше словом "граф" називають простий неорієнтований граф,

тобто граф, який не має петель та кратних ребер.

Ребро називається петлею, якщо його кінці збігаються.

Ребра називаються кратними, якщо множини їх кінцевих вершин збігаються.

Вершини v_i та v_j називаються кінцевими вершинами (або просто кінцями) ребра l_{ij} . Дві кінцеві вершини одного і того ж ребра називаються сусідніми. Цей факт позначають як $i \square j$.

Два ребра називаються суміжними, якщо вони мають загальну кінцеву вершину.

Максимальна кількість дуг простого графу складає $L_{\max} = n(n-1)/2$. Простий граф з такою кількістю дуг називається повним; кожна вершина в ньому з'єднана з усіма іншими. Мінімальна кількість ребер зв'язаного графу дорівнює $E_{\min} = n-1$. Такий граф не має циклів, тобто є деревом.

Мірами щільності графу можуть слугувати:

- лінійна щільність – відношення фактичної кількості дуг графа до максимально можливої $\rho = L / L_{\max}$;
- скорегована лінійна щільність – $\rho_c = (L - L_{\min}) / (L_{\max} - L_{\min})$;
- логарифмічна щільність – $\rho_L = \log(L) / \log(L_{\max})$;
- кликове число $\omega(G)$ (clique number) – розмір максимального за розміром повного підграфа.

Слід зазначити, що неорієнтований зв'язок (ребро) еквівалентний двом дугам, тому для неорграфу завжди $L = 2E$.

Ступенем вершини k_i називають кількість інцидентних їй ребер. Для будь-якого графа виконується тотожність

$$\sum_{i=1}^n k_i = L. \quad (0.1)$$

У разі, якщо граф орієнтований, то розрізняють вхідну та вихідну

ступені вершин (k_i^+, k_i^-) . При цьому для кожної з них виконується тотожність (0.1).

Вершина називається ізольованою, якщо вона не є кінцем ні для одного ребра; висячою (або листом), якщо вона є кінцем рівно одного ребра.

Маршрутом в графі називають кінцеву послідовність вершин, в якій кожна вершина (крім останньої) з'єднана з наступною вершиною ребром. Ланцюгом називається маршрут без повторення ребер. Простим ланцюгом називається маршрут без повторюваних вершин (звідки випливає, що в простому ланцюзі немає повторюваних ребер).

Простим циклом називають ланцюг, в якому ребра не повторюються, а перша і остання вершини співпадають. Довжиною циклу (тобто шляху) називають число ребер в ньому.

Елементарним циклом називають такий простий цикл, в якому не тільки ребра, а й вершини не повторюються.

Існують узагальнення поняття «граф»:

- мультиграф – граф, який містить кратні ребра, тобто ребра, що з'єднують одну і ту ж пару вершин;
- псевдограф – граф, який містить петлі, тобто ребра, які з'єднують вершини самі з собою;
- гіперграф – граф, в якому ребро може з'єднувати більше двох вершин;
- ультраграф і інші.

Існує ряд способів представлення графів для їх зберігання і комп'ютерної обробки. Вибір того, чи іншого способу залежить від розміру графа, його щільності, видів розв'язуваних задач. Основними способами є:

- матриця суміжності. Стівпці та рядки цієї матриці відповідають вершинам графа. Елементи матриці визначають наявність зв'язку від вершини-рядка до вершині-колонки (1), або його відсутність (0). Цей

спосіб є найбільш зручним у разі щільних графів. Недоліком матриць суміжності є витрати пам'яті ($O(n^2)$), які є занадто великими, якщо граф розріджений.

- матриця інцидентності. Її рядки відповідають вершинам графа, а стовпці – ребрам (чи дугам). Елемент матриці (на перетині рядка i зі стовпцем l) дорівнює 1, якщо дуга l входить в вершину i , -1, якщо вона виходить з неї, та 0 в інших випадках. Цей спосіб представлення графа є найбільш витратним по пам'яті ($O(n \cdot |L|)$), тому застосовується рідко;

- список суміжності. Кожній вершині графа відповідає рядок, в якому зберігається список суміжних вершин. Така структура даних не є таблицею в звичайному розумінні, а являє собою «список списків»;

- список ребер. Кожному ребру (чи дузі) графа відповідає рядок цього списку, в якому зберігаються номери двох вершин, інцидентних цьому ребру/дузі.

Для двох останніх способів витрати пам'яті на зберігання структури графа складають $O(|L|)$. Таким чином, ці способи є найбільш економними. Саме вони часто застосовуються для подання розріджених графів, при реалізації базових алгоритмів обходу графа, для зовнішнього зберігання або обміну даними тощо.

1.2 Визначення та властивості складних мереж

Терміни «мережа» і «граф» багато в чому схожі. Під час дослідження структури мережі, її топології, математичною моделлю мережі є граф, тому терміни «мережа» і «граф» в цьому сенсі можна вживати як синоніми. Втім, "чисті" математики надають перевагу терміну "граф", а спеціалісти з прикладних галузей частіше вживають слово "мережа". Оскільки чисельність математиків набагато менша, а також через популярність соціальних мереж та мережі інтернет, термін «мережа» зараз застосовується як більш сучасний та загальний термін, ніж «граф».

Аналогом поняття "вершина" графа є "вузол" мережі, а замість "дуга, ребро" частіше застосовують термін "зв'язок" (який може розглядатись як однонаправлений, чи двонаправлений).

В ХХ столітті сформувався новий науковий напрям: теорія складних мереж. Незважаючи на те, що в об'єктах цієї теорії є різні мережі – інформаційні, біологічні, електричні, транспортні, найбільший внесок в розвиток теорії складних мереж внесли дослідження соціальних мереж. Термін «соціальна мережа» позначає зосередження соціальних об'єктів, які можна розглядати як мережу, вузли якої – соціальні об'єкти (наприклад, люди), а зв'язки – соціальні відносини між ними. Цей термін був введений в 1954 році соціологом Дж. Барнсом.

Ключовими властивостями складних мереж є такі:

- великі розміри;
- елементи випадковості при формуванні;
- зростання (зміна) в часі.

Типовими задачами досліджень складних мереж є:

- дослідження класичних графових характеристик цих мереж;
- дослідження статистичних властивостей мереж;
- моделювання складних мереж, їх візуалізація;
- моделювання і вивчення динаміки різних процесів, "носієм" яких є складні мережі, – дифузії, епідемічних процесів, різних потоків (нафти, інформації, електричного струму і т.д.);
- вивчення прикладних характеристик мереж – уразливості мереж, методи відновлення, захисту та знищення мереж (наприклад, скільки вузлів потрібно «відключити» для того, щоб мережа втратила зв'язність).

Однією з найважливіших статистичних характеристик складних мереж є розподіл вузлів за ступенями p_k (тобто ймовірність того, що вузол має рівно k зв'язків). Для більшості складних мереж реального світу ця функція розподілу належить одному з трьох сімейств (рис. 1.2):

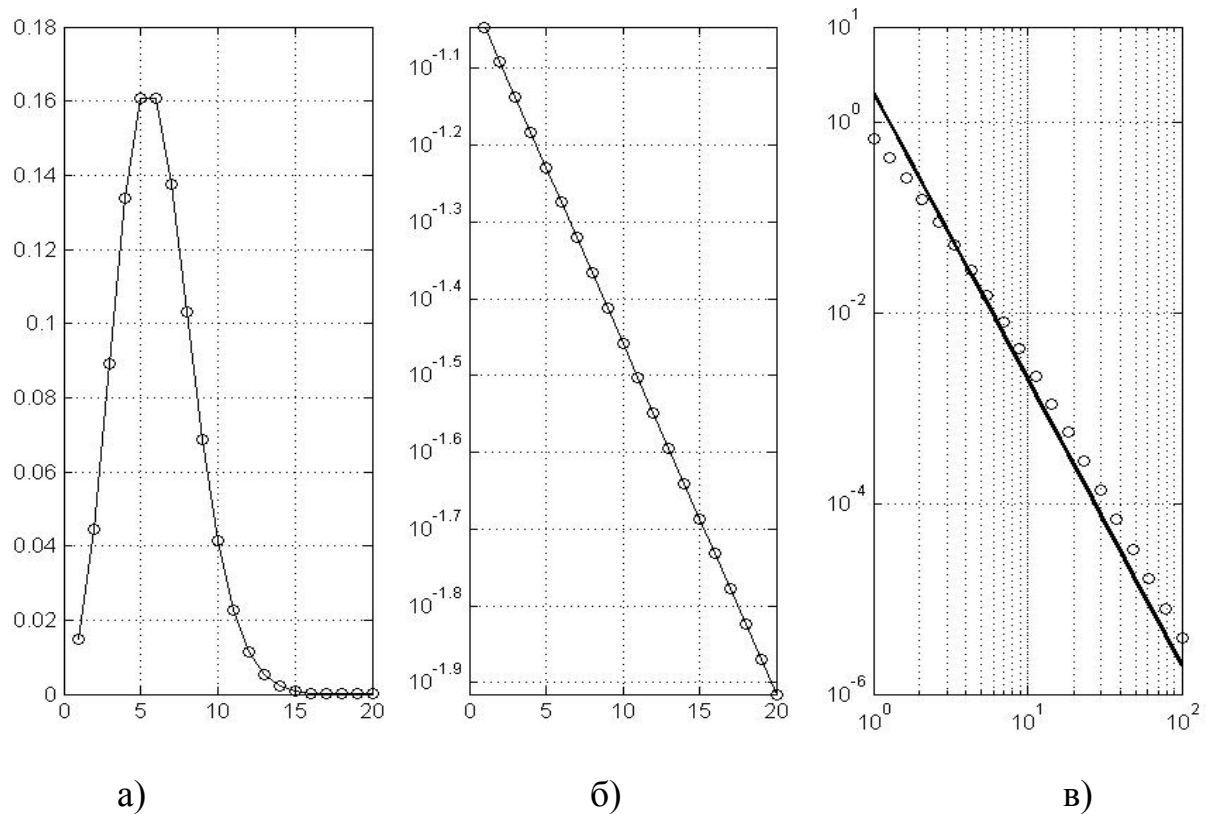


Рисунок 1.2 – Типові графіки розподілів вузлів мережі за ступенями:

а) пуасонівське, б) геометричне, в) ступеневе

а) Розподіл Пуасона $Poisson(\lambda)$:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda} \propto Poisson(\lambda). \quad (0.2)$$

Такий закон розподілу породжується моделлю випадкової мережі і характеризується немонотонністю, тобто наявністю "характерного", "середнього" ступеню вузлів (рис. 1.2(a)).

б) Геометричний розподіл $Geom(p)$:

$$p_k = pq^{k-1} \propto Geom(p) \quad (q = 1 - p). \quad (0.3)$$

Графік такого закону (рис.1.2(б)) в координатах $\log(p_k)$ від k має вигляд прямої лінії з негативним коефіцієнтом нахилу. Безперервним аналогом геометричного розподілу є експоненціальний. Якщо мережа має

такий розподіл вузлів за ступенями, то кажуть, що в ній "переможець отримує все", тобто мережа є гіперцентралізованою.

в) Ступеневий розподіл $Pow(\gamma, x_{min})$:

$$f(x) = C(\gamma, x_{min}) \cdot x^{-\gamma} \propto Pow(\gamma, x_{min}). \quad (0.4)$$

Графік ступеневого розподілу (рис.1.2(в)) також має вигляд прямої з негативним коефіцієнтом нахилу, але в подвійних логарифмічних координатах: $\log(p_k)$ від $\log(k)$. При цьому якщо на графіку позначення на вісях відсутні, то визначити масштаб неможливо, тому такий розподіл називають безмасштабним, або масштабно-інваріантним. Більше того, за будь-яким значенням $0 < P < 1$, доля зв'язків W , які інцидентні $n \cdot P$ найбагатшим на зв'язки вузлам, не залежить ані від розміру мережі n , ані від зсуву вибірки вузлів, та складає

$$W(P, \gamma) = P^{(\gamma-2)/(\gamma-1)}. \quad (0.5)$$

Частковим випадком масштабної інваріантності є широко відоме правило Парето «80/20», за яким 20 відсотків найбагатіших вузлів (хаби) контролюють 80% всіх зв'язків мережі. Навіть якщо видалити з мережі ці хаби та всі їхні зв'язки, то серед тієї частини мережі, що залишиться, 20% вузлів будуть мати так саме 80% зв'язків. Цьому частковому випадку відповідає рівняння (0.5) у вигляді $W(0.2) = 0.8$, тобто $\gamma \approx 2.161$.

В безмасштабній мережі (scale free network) виконується парадигма "переможець отримує більше", тобто мережа є централізованою (хаби існують), але не гіперцентралізованою (не зіркоподібною).

Більшість мереж реального світу (соціальні, інформаційні, технічні, деякі біологічні) є такими, що невпинно зростають. За умов зростання, властивості таких мереж визначаються правилами, за якими в мережу додаються нові вузли та зв'язки.

1.3 Концепції зростання та приєднання

Відповідно до концепції зростання, початкова мережа включає n_0 вузлів та $L_0 / 2$ ребер, а потім на кожному кроці до мережі додається новий вузол, який з'єднується з існуючими. Таким чином, розмір мережі n (тобто кількість вузлів) зазвичай використовується як міра часу. Крім того, номер вузла (i) вважається часом його народження (тобто $n - i$ є віком вузла).

Окрім часу народження, кожен вузол характеризується своїм ступенем (k_i) та іншими властивостями, наприклад, придатністю (fitness), тощо. Динаміка росту мережі визначається кількістю зв'язків, що з'єднують вхідний вузол з існуючими $m(n)$, або середнім ступенем вузлів $\bar{k}(n)$. Мережа також може мати інші керуючі параметри (такі як додаткова привабливість, еластичність тощо).

Властивості мережі по суті визначаються правилом приєднання, тобто ймовірністю π_i того, що нове ребро вказує на вузол i , як функцію вищезазначених параметрів:

$$\pi_i = f(i, k_i, n_0, L_0, \eta_i, A). \quad (0.6)$$

Ключовою властивістю безмасштабних мереж є розподіл вузлів за кількістю зв'язків $p(k)$, який повинен, принаймні, асимптотично, дотримуватися закону Юла-Саймона (який є дискретним аналогом ступеневого закону (0.4)):

$$p(k) = c \frac{\Gamma(k)}{\Gamma(k + \gamma)} \propto ck^{-\gamma}, \quad k \geq k_0. \quad (0.7)$$

Параметр $\gamma \geq 2$ називається коефіцієнтом скейлінгу.

Проблема полягає у пошуку правила приєднання (0.6), в якому ступінь вузла k_i використовується неявно, та яке дозволяє генерувати безмасштабну (0.7) мережу.

1.4 Огляд моделей безмасштабних мереж

Найпростішою та найвідомішою моделлю для безмасштабних мереж є модель Барабаші-Альберт (БА) [3], запропонована в 1999 р. Альбертом-Ласло Барабаші (A.-L. Barabási) та Рекою Альберт (R. Albert). В її основі лежать дві основні концепції: концепція зростання та концепція переважного приєднання, згідно з якими кожен вхідний вузол підключається до $m = const$ існуючих. Ймовірність π_i того, що нове ребро з'єднає вхідний вузол $(n + 1)$ з деяким існуючим вузлом i , пропорційна його ступеню k_i :

$$\pi_i = k_i / \sum_i k_i. \quad (0.8)$$

Таким чином, правило приєднання БА є найпростішим випадком загальної форми (0.6). Еволюція мережі БА у випадку $m = 1$ (дерево БА) показана на рис.1.3.

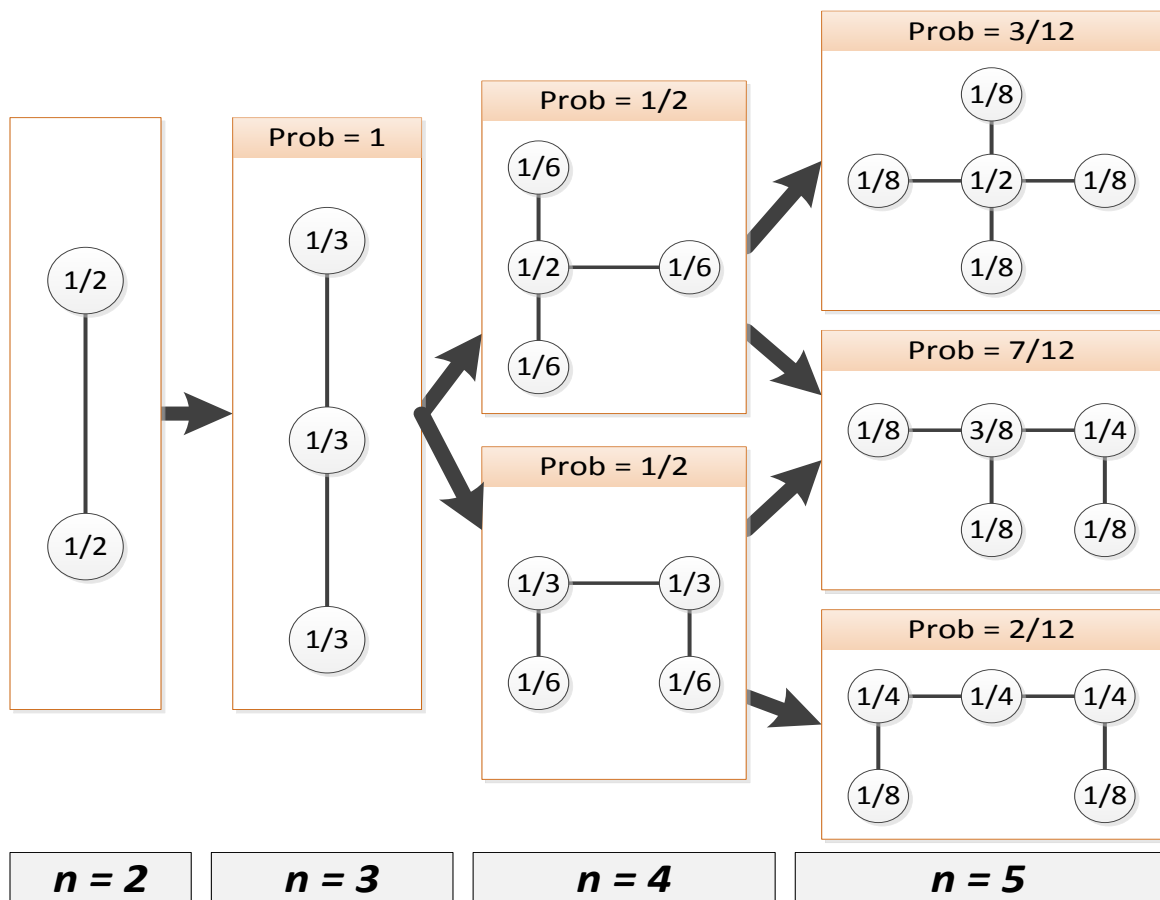


Рисунок 1.3 – Еволюція дерева Барабаші-Альберт

В вузлах мережі на рис. 1.3 показані ймовірності приєднання нового вузла. Також на цьому рисунку показані ймовірності (Prob), за якими буде реалізована відповідна топологія мережі. Можна бачити, що ймовірність утворення ланцюга, або зіркоподібної структури (гіперхаба), є невеликою та швидко зменшується з ростом розміру мережі. Найбільш ймовірною є така структура, за якою хаби (тобто вузли, "багаті" на зв'язки) існують, але вони контролюють не всі зв'язки. Тобто реалізується парадигма "чим більше маєш – тим більше отримаєш".

При $n \rightarrow \infty$ модель БА призводить до розподілу Юла-Саймона за ступенем вузлів (0.7) з коефіцієнтом скейлінга $\gamma = 3$. Зовнішній вигляд дерева БА при $n \gg 1$ показаний на рис 1.4.

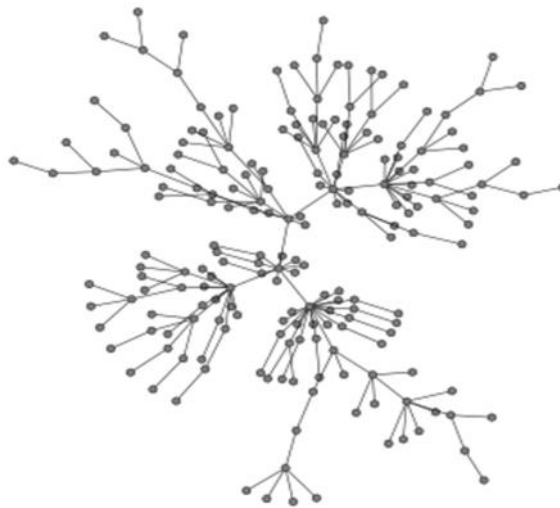


Рисунок 1.4 – Дерево Барабаші-Альберт

Відповідно до фітнес-моделі [4], кожен вузол асоціюється зі своїм значенням фітнесу η_i . Тому правило (0.8) узагальнюється до

$$\pi_i = f(k_i, \eta_i) = k_i \eta_i / \sum k_i \eta_i. \quad (0.9)$$

Існують також більш складні варіанти зважування [5-7]. Такий підхід дозволяє відтворити ступінь кореляції між вузлами мережі.

Інші розширення моделі БА засновані на застосуванні старіння [6],

тобто у залежності ймовірності прикріплення π_i в (0.6) безпосередньо від i (чи $n - i$):

$$\pi_i(n) = f(i, k_i(n)). \quad (0.10)$$

У моделі Прайса [2] мережа має глобальний параметр додаткової привабливості a , а правило приєднання (0.6) має форму

$$\pi_i = (k_i + a) / \sum (k_i + a). \quad (0.11)$$

Цей підхід був використаний для моделювання мереж цитування, які розглядаються як направлені. Тому в правилі приєднання під k_i розуміються лише вхідні посилання. Зрозуміло, що якщо $a = 0$ (тобто якщо використовується правило БА (0.8)), то нова (тобто вхідна) стаття, на яку ще не може бути посилань, не отримає їх і в майбутньому.

В даний час існує багато інших специфікацій та узагальнень для генеративного механізму (0.6), що використовується для створення мережевих моделей (нелінійне преференційне приєднання [8-9], перенаправлення зв'язків [10], приєднання на основі даних другого рівня сусідства [11] тощо).

Не всі перераховані вище моделі породжують саме безмасштабні мережі. Насправді [9] єдиний випадок, коли топологія мережі є безмасштабною, – це коли ймовірність приєднання є асимптотично лінійною відносно k_i , тобто $\pi_i \sim a_\infty k_i$. З іншого боку, незважаючи на широке поширення моделей безмасштабних мереж, мережі реального світу не завжди є безмасштабними [12].

Будучи ядром безмасштабних мережевих моделей, модель БА має певні первісні недоліки та обмеження (успадковані більшістю похідних моделей), які обмежують область її застосування:

- у реальних мережах коефіцієнт скейлінга γ не завжди дорівнює 3, а зазвичай має значення в діапазоні $2 < \gamma \leq 3$ (хоча, слід зазначити [1-2], що

процеси та явища з $\gamma \leq 2$ також існують);

- у реальних мережах середня ступінь вузлів $\bar{k}(n)$ має тенденцію до зростання з розміром мережі, в той час як вона є постійною в БА-моделі.

Таким чином, ця модель породжує дуже розріджені мережі;

- модель БА породжує нейтральні за асортативністю мережі [10, 14], в той час як соціальні мережі є позитивно асортативними, а біологічні та технічні – дизасортативними;

- інформація про ступінь вузлів використовується прямо;

- нарешті, здається дивним, що модель безмасштабної мережі базується на параметрі m , який залежить від масштабу.

Для подолання вищезазначених обмежень було запропоновано розширити перелік основних концепцій моделі БА новою: концепцією еластичності [15-16], тобто концепцією різних відносних темпів зростання чисельностей зв'язків та вузлів. Згідно з цим підходом використовується коефіцієнт еластичності $1 \leq \lambda < 2$ як відношення відносних темпів росту зв'язків та вузлів:

$$\lambda = \frac{\delta L(t)}{\delta n(t)} = \frac{\Delta L(t)}{L(t)} \bigg/ \frac{\Delta n(t)}{n(t)} = n \frac{L(n+1) - L(n)}{L(n)}. \quad (0.12)$$

Якщо параметр λ є фіксованим, загальна кількість зв'язків зростає асимптотично за ступеневим законом з показником λ :

$$L(n) = 2 \frac{\Gamma(n + \lambda - 1)}{\Gamma(n - 1)\Gamma(\lambda + 1)} \propto n^\lambda. \quad (0.13)$$

Також середній ступінь вузлів не є постійним m , а зростає:

$$\bar{k}(n) = L(n) / n \propto n^{\lambda-1}. \quad (0.14)$$

Якщо $\lambda > 1$, то кожен новий вузол приносить більше посилянь, ніж є у середньому на даний час (0.14). Наприклад, чим більше існує наукових

праць за даною тематикою, тим більше посилань має переглядати автор нової роботи.

Тому використання концепції еластичності узагальнює оригінальну модель БА та дає можливість створювати щільні мережі та мережі з коефіцієнтом скейлінга $\gamma = 1 + 2 / (2 - \lambda)$, який не перевищує 3.

1.5 Постановка задачі дослідження

Явне використання інформації про ступінь вузлів утворює ще одне суттєве обмеження як моделі БА, так і похідних моделей [13]. У реальних мережах вхідний вузол не може оцінити інформацію для всіх існуючих вузлів через великий розмір мережі, причини конфіденційності тощо. Дійсно, під час первісного вибору сайту чи наукової статті для ознайомлення з якоюсь темою, ми не цікавимося кількістю посилань на інші веб-сайти чи документи, які він/вона має. Аналогічно, майбутній бізнесмен не може отримати інформацію про те, хто з потенційних партнерів має скільки торгових зв'язків.

Модель приєднання за посередництвом (MDA - mediation-driven attachment) була запропонована Хасаном та ін. [13] для подолання цього обмеження. Ця модель може розглядатися як версія більш ранньої моделі копіювання [17].

Дослідження цієї моделі є метою атестаційної роботи. Її досягнення реалізується через розв'язання наступних задач:

- аналітичне дослідження моделі MDA, визначення її основних властивостей;
- програмна реалізація моделі;
- імітаційне моделювання, чисельне визначення характеристик та властивостей сформованих моделей складних мереж;
- порівняння результатів аналітичного дослідження з даними комп'ютерного моделювання.

2 ДОСЛІДЖЕННЯ МОДЕЛІ ПРИЄДНАННЯ ЗА ПОСЕРЕДНИЦТВОМ

2.1 Опис моделі приєднання за посередництвом

Згідно з правилом MDA, кожен вхідний вузол обирає якийсь існуючий рівноймовірно. Цей вузол називається медіатором. Потім вхідний вузол посилається на вхідний вузол сусідів цього посередника.

Так, під час первісного вибору сайту чи наукової статті для ознайомлення з якоюсь темою, ми не цікавимося кількістю посилань на інші веб-сайти чи документи, які він/вона має. Аналогічно, майбутній бізнесмен не може отримати інформацію про те, хто з потенційних партнерів має скільки торгових зв'язків.

Проте, після первісного відвідування веб-сайту чи ознайомлення з науковою статтею, читач, можливо, захоче перейти за деякими доступними посиланнями з цього сайту/статті на інші джерела. Цей первісний сайт/стаття саме й виступають у якості посередника.

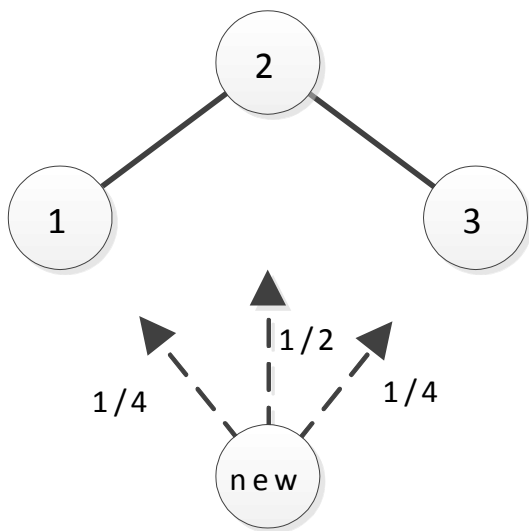
Ймовірність π_i того, що вузол i , який має ступінь k_i , приєднається до вхідного, можна знайти наступним чином: спочатку пронумеруємо сусідів вузла i : $1, 2, \dots, k_i$. Ймовірність того, що кожен з них буде обраний медіатором, дорівнює $1/n$. Ймовірність з'єднання вузла i з вхідним через посередництво вузла j дорівнює $1/k_j$. Таким чином,

$$\pi_i = \frac{1}{n} \sum_{j=1}^{k_i} \frac{1}{k_j}. \quad (0.15)$$

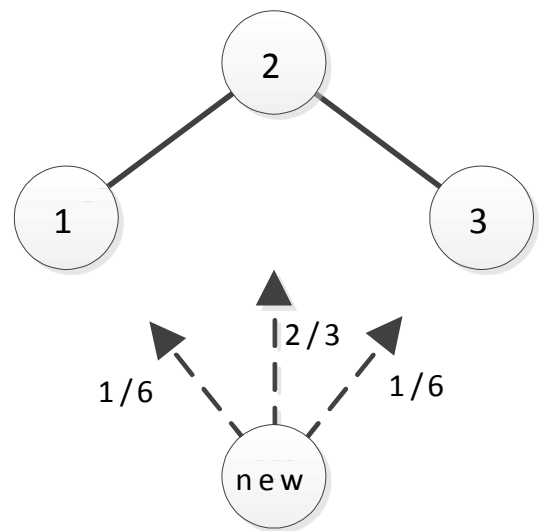
Відповідно до (0.15), ймовірність з'єднання вхідного вузла з існуючими залежить лише від їх ступеня (так, як і в моделі БА), але, на відміну від моделі БА, ця залежність є неявною.

Порівняння правил БА та MDA проілюстровано на рис. 2.1. Згідно з

(0.8), у моделі БА вхідний вузол "знає", що $k_1 = k_3 = 1$, та $k_2 = 2$. Таким чином, $\pi_i = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ відповідно. Згідно з правилом MDA, вхідний вузол зв'язується з вузлом 1 лише у тому випадку, якщо вузол 2 обраний як посередник (що відбувається з ймовірністю $1/3$), а вузол 1 обраний серед сусідів вузла 2 (ймовірність - $1/2$). Таким чином, $\pi_1 = \frac{1}{6}$. Аналогічно, $\pi_3 = \frac{1}{6}$. Вузол 2 буде з'єднаний з вхідним, якщо вузол 1 або вузол 3 вибрано як посередника, тому $\pi_2 = \frac{2}{3}$.



BA attachment rule

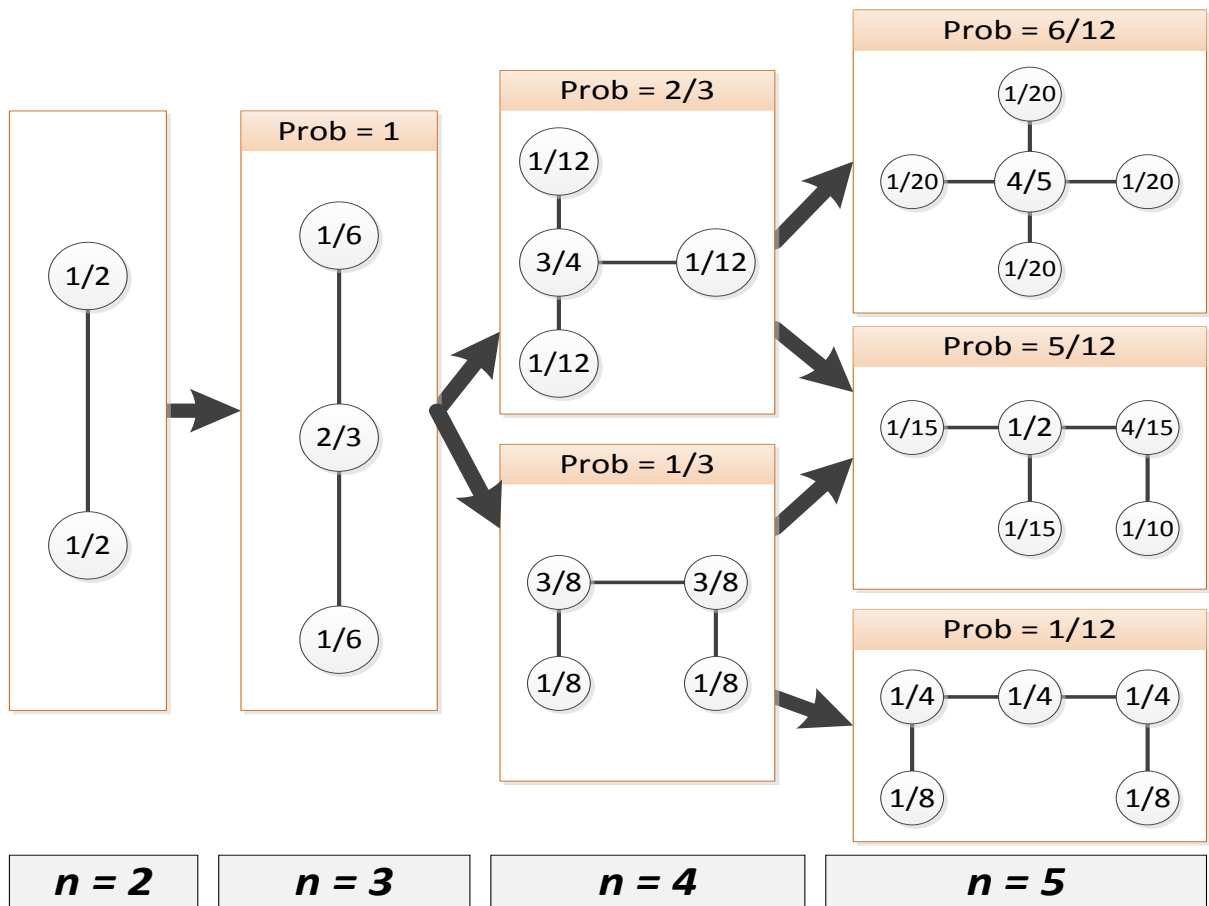


MDA attachment rule

Рисунок 2.1 – Порівняння між правилами приєднання Барабаші-Альберт та MDA у разі $m = 1$

Еволюція мережі MDA у разі $m = 1$ показана на рис. 2.2. Так саме, як і для мережі БА (рис.1.3), в вузлах мережі MDA (рис. 2.2) показані ймовірності приєднання нового вузла. Також на цьому рисунку показані ймовірності (Prob), за якими буде реалізована відповідна топологія мережі.

Можна бачити, що на відміну від моделі БА, у подолі приєднання за посередництвом, яка розглядається, ймовірність утворення зіркоподібної структури (гіперхаба), є великою, тобто модель MDA (при $m = 1$) схильна до парадигми "переможець отримує все".

Рисунок 2.2 – Еволюція мережі MDA у разі $m = 1$

Цей результат можна легко довести аналітично. Якщо мережа з n вузлами є зіркоподібною, то новий вузол приєднується до центрального у тому разі, якщо посередником буде обрано будь який вузол крім центрального (бо в них немає ніяких інших сусідів). Ймовірність цього дорівнює

$$\pi_*^{(n)} = (n - 1) / n . \quad (0.16)$$

Таким чином, ймовірність $P_*^{(n+1)}$ того, що мережа з $n + 1$ вузлом буде зіркоподібною є добутком ймовірностей того, що вона була такою, маючи розмір n , (тобто $P_*^{(n)}$) та того, що новий вузол приєднується до центрального (0.16):

$$P_*^{(n+1)} = P_*^{(n)} \cdot \pi_*^{(n)} = P_*^{(n)} \cdot \frac{n - 1}{n} . \quad (0.17)$$

З урахуванням початкової умови $P_*^{(3)} = 1$ (деревовидна мережа з трьома вузлами завжди зіркоподібна – рис. 2.2), розв'язок (0.17) має вигляд:

$$P_*^{(n)} = \frac{2}{n-1}. \tag{0.18}$$

Для порівняння, ймовірність утворення зіркоподібної мережі за умов дії явного переважного з'єднання (БА модель) складає лише $P_*^{BA} = 2^{3-n}$.

Розглянемо тепер еволюцію мережі MDA, за умов що кожний новий вузол приєднується до $m = 2$ різних існуючих вузлів мережі. Графічне зображення перших двох кроків дано на рис. 2.3. Вважаємо, що кратні ребра неприпустимі.

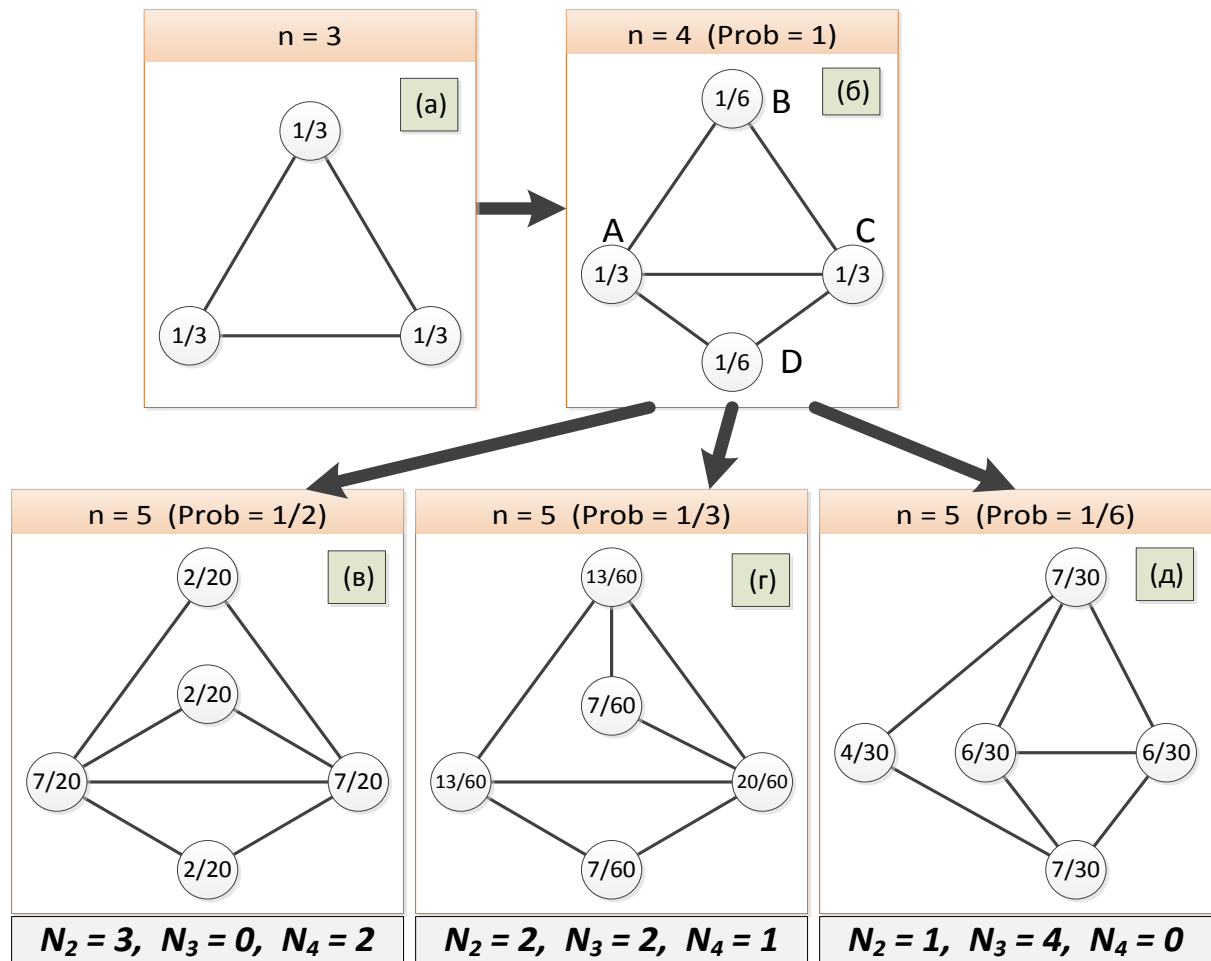


Рисунок 2.3 – Еволюція мережі MDA у разі $m = 2$

Спочатку мережа являє собою повний граф з трьома вузлами та трьома ребрами (рис.2.3(а) – кожен вузол має рівно $m = 2$ сусідів). Ймовірності приєднання нового вузла до кожного з трьох існуючих, вочевидь, рівні. Незалежно від того, до яких саме він приєднається, структура мережі з чотирма вузлами буде мати такий вигляд, як на рис. 2.3(б)

Другий крок розглянемо докладніше. Поперше, розрахуємо ймовірності нового вузла приєднання до існуючих, позначених як A, B, C, D . Вузол A може бути приєднаний через посередництво вузлів B, C, D , які мають відповідно 2, 3 та 2 сусідів. Тому згідно з (0.15) $\pi_A = \frac{1}{4}(\frac{1}{2} + \frac{1}{3} + \frac{1}{2}) = \frac{1}{3}$. Аналогічним чином, $\pi_C = \pi_A = \frac{1}{3}$. Вузли B та D можуть бути приєднані через посередництво A , або C . Згідно з (0.15) $\pi_B = \pi_D = \frac{1}{4}(\frac{1}{3} + \frac{1}{3}) = \frac{1}{6}$.

По-друге, визначимо можливу структуру мереж з $n = 5$ вузлами та розрахуємо ймовірності реалізації кожної з цих структур. Якщо п'ятий вузол обере у якості посередника вузол B , то нові зв'язки безальтернативно встановляться з обома сусідами B , тобто з A та C (рис.2.3(в)). Теж саме стане і в тому разі, якщо посередником буде вузол D . Ймовірність вибору вузлів-посередників однакова та дорівнює $1/4$, тому ймовірність утворення структури рис.2.3(в) становить $1/2$.

Якщо посередником буде обрано вузол A , чи вузол C , то можуть утворитись структури, показані на рис.2.3(г)-(д). Вузли A та C мають по три сусіди. Умовна ймовірність того, що для зв'язків будуть обрані вузли B та D (за умовою обрання A у якості посередника) становить $1/3$. Утворення зв'язків з B та D через посередництво C також має ймовірність $1/3$. Таким чином, ймовірність реалізації структури рис.2.3(д) становить $\frac{1}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{6}$. У всіх інших випадках буде утворено структуру, яка з точністю до симетрії подібна до зображеної на рис.2.3(г). Ймовірність цього складає $2/3$.

Таким чином, ймовірності реалізацій кожної з можливих структур

мережі з п'ятьма вузлами (рис.2.3(в)-(д)) дорівнюють $\frac{1}{2}$, $\frac{1}{3}$ та $\frac{1}{6}$ відповідно.

На рис. 2.3(в)-(д) також наведені розподіли вузлів по ступеням для кожної з трьох структур мережі з п'ятьма вузлами та ймовірності приєднання нового вузла до існуючих.

Аналізуючи рис. 2.2 та рис. 2.3, можна зробити висновок, що ймовірність високої концентрації зв'язків (тобто наявності "суперхабів") при $m = 2$ трохи нижче, ніж у випадку деревовидної мережі з $m = 1$, але значно вище, ніж для мережі Барабаші-Альберт.

2.2 Динамічні властивості MDA-мережі

Розглянемо загальний випадок динаміки мережі у разі довільного (але постійного) значення керуючого параметра $m = const$.

Будемо вважати, що початковою мережею є повний граф, в якому ступені всіх вузлів дорівнюють m . Такий граф складається з $n_0 = m + 1$ вузлів та містить $L_0 = m \cdot n_0 = m(m + 1)$ зв'язків ($e_0 = m(m + 1) / 2$ ребер). Обрання зазначеної початкової конфігурації обумовлено тим, що згідно з правилом MDA кожний з вузлів повинен мати не менш, ніж m сусідів. Якщо сусідів буде менше, то новий вузол не зможе встановити всі свої m зв'язків. Повний граф з $n_0 = m + 1$ вузлів є мінімальним за кількістю вузлів серед усіх графів, для яких ступені вузлів не менші за m .

Оскільки на кожному кроці в мережу додається один вузол, то є природним ототожнити розмір мережі (тобто кількість вузлів n) з номером кроку (тобто з мірою часу). Оскільки кожен новий вузол додає m ребер, тобто $2m$ зв'язків, то загальна їхня кількість у мережі розміром n та середня ступінь вузла складуть відповідно

$$L(n) = m \cdot (2n - n_0), \quad (0.19)$$

$$\bar{k}(n) = m \left(2 - \frac{n_0}{n}\right), \quad \lim_{n \rightarrow \infty} \bar{k}(n) = 2m. \quad (0.20)$$

Будемо розглядати процес встановлення зв'язків між новим вузлом та існуючими як послідовність з m незалежних спроб. Тоді ймовірність отримання зв'язка вузлом i дорівнює $m \cdot \pi_i(n)$, а значить, що саме на цю величину зросте математичне очікування ступені i -го вузла:

$$k_i(n+1) - k_i(n) = m \cdot \pi_i(n). \quad (0.21)$$

Перетворимо вираз (0.15) для $\pi_i(n)$:

$$\pi_i(n) = \frac{1}{n} \sum_{\substack{j=1, \\ j \square i}}^{k_i(n)} \frac{1}{k_j(n)} = \frac{k_i(n)}{m \cdot n} \left(\frac{m}{k_i(n)} \sum_{\substack{j=1, \\ j \square i}}^{k_i(n)} \frac{1}{k_j(n)} \right), \quad (0.22)$$

та позначимо вираз у дужках формули (0.22) як $b_i(n)$:

$$b_i(n) = \frac{m}{k_i(n)} \sum_{\substack{j=1, \\ j \square i}}^{k_i(n)} \frac{1}{k_j(n)}. \quad (0.23)$$

Припустимо, що для мережі великого розміру ($n \rightarrow \infty$) параметр $b_i(n)$ (0.23) прагне до деякої константи

$$\lim_{n \rightarrow \infty} b_i(n) = \beta. \quad (0.24)$$

Тоді, підставивши (0.24) (замість (0.23)) до (0.22), а потім (0.22) до (0.21), отримаємо рівняння динаміки зростання ступеню довільного вузла i :

$$k_i(n+1) - k_i(n) = \beta \cdot \frac{k_i(n)}{n}. \quad (0.25)$$

Це різницеве рівняння має загальний розв'язок

$$k_i(n) = C_i \cdot \frac{\Gamma(n + \beta)}{\Gamma(n)}, \quad (0.26)$$

де $\Gamma(x)$ – гамма-функція Ейлера (яка є узагальненням факторіалу на дійсні числа: $\Gamma(n) = (n - 1)!$ для натуральних n).

З (0.26) випливає, що ступені вузлів зростають асимптотично як ступенева функція з показником β : $k_i(n) \propto n^\beta$.

Для визначення констант C_i , які входять у (0.26), будемо вважати, що вузли відсортовані за віком. Тоді на кроці $i \geq n_0$ мережа складається з i вузлів (нагадуємо, що номер кроку ототожнений з розміром мережі). Новий вузол отримує номер $i + 1$, та, згідно з умовами правила приєднання, встановить рівно m зв'язків з вже існуючими. Таким чином, $k_i(i) = m$, з чого випливає, що

$$C_i = m \frac{\Gamma(i)}{\Gamma(i + \beta)}, \quad i \geq m. \quad (0.27)$$

Для вузлів, які є "батьками-засновниками" мережі, тобто для перших $n_0 = m + 1$ вузлів, початкові умови мають вигляд $k_i(n_0) = m$, тому

$$C_i = m \cdot \frac{\Gamma(n_0)}{\Gamma(n_0 + \beta)}, \quad 1 \leq i \leq n_0. \quad (0.28)$$

З урахуванням (0.27)-(0.28), математичне очікування ступенів вузлів (0.26) дорівнює

$$k_i(n) = m \cdot \frac{\Gamma(i^*)}{\Gamma(i^* + \beta)} \cdot \frac{\Gamma(n + \beta)}{\Gamma(n)}, \quad \text{де } i^* = \max(i, n_0). \quad (0.29)$$

У асимптотиці при великих значеннях i та n залежність (0.29) є ступеневою з показником β :

$$k_i(n) \propto m \cdot (n/i)^\beta. \quad (0.30)$$

З виразів (0.29) та (0.30) легко бачити, що залежність ступеня вузла від його номера є монотонно спадаючою, тому номер вузла є математичним очікуванням його рангу.

Якщо ж розглядати розподіл (0.29), чи його наближення (0.30), як рангові, то згідно з [2], ступеневому ранговому розподілу з показником β відповідає ступеневий частотний розподіл $p(k) \propto ck^{-\gamma}$, (0.7) з показником скейлінгу

$$\gamma = 1 + \frac{1}{\beta}. \quad (0.31)$$

Таким чином, можна підвести проміжний підсумок досліджень: застосування правила MDA (приєднання за посередництвом) так само, як і у разі використання звичайного переважного з'єднання (БА-модель) має приводити до безмасштабної мережі. Єдина різниця полягає у значенні показника скейлінгу: для БА-моделі $\gamma = 3$, а для MDA-моделі він має форму (0.31).

В той же час слід зазначити, що проведений аналіз та зроблений вище висновок ґрунтуються на припущенні (0.24), тобто що параметр β дійсно є константою. Доведення цього факту розглядатиметься в наступному розділі.

2.3 Оцінювання показника скейлінгу для моделі MDA

Згідно з (0.31) оцінювання показника скейлінгу розподілу вузлів MDA-мережі за ступенями зводиться до оцінювання параметра β (0.24). Нагадаємо, що у точне рівняння динаміки зростання ступенів вузлів (0.21) входить параметр $b_i(n)$ (0.23) – середньоарифметичне значення величин, обернених до ступенів вузлів, з'єднаних з вузлом i . При цьому розв'язання

рівняння (0.21) та його аналіз ґрунтувались на припущенні (0.24), що параметр $b_i(n)$ є асимптотично незалежним ані від номера вузла i (та його ступеня), ані від розміру мережі n , та прагне до константного значення β .

Припущення щодо незалежності $b_i(n)$ від i (та, відповідно, від k_i) означає, що асимптотично середнє значення величин, обернених до ступенів вузлів, поєднаних з вузлом i , є таким самим, як і середнє значення величин, обернених до ступенів вузлів, по всій мережі:

$$b_i(n) = \frac{m}{k_i(n)} \sum_{\substack{j=1, \\ j \sim i}}^{k_i(n)} \frac{1}{k_j(n)} \approx \frac{m}{n} \sum_{i=1}^n \frac{1}{k_i(n)}. \quad (0.32)$$

Суму, яка знаходиться в наближеному виразі (0.32) легко знайти, просуммувавши $k_i(n)$, знайдене в (0.29). Вона дорівнює

$$\sum_{i=1}^n \frac{1}{k_i(n)} = \frac{n + \beta}{m(\beta + 1)} + \frac{\Gamma(n)}{\Gamma(n + \beta)} \cdot \frac{\Gamma(n_0 + \beta)}{\Gamma(n_0)} \cdot \frac{\beta}{\beta + 1}. \quad (0.33)$$

При $n \rightarrow \infty$ друга складова в (0.33) спадає до нуля зі швидкістю $O(n^{-\beta})$, тому оцінка параметра $b_i(n)$ (0.32) має вигляд:

$$\lim_{n \rightarrow \infty} b_i(n) = \lim_{n \rightarrow \infty} \left(\frac{m}{n} \cdot \frac{n + \beta}{m(\beta + 1)} \right) = \frac{1}{\beta + 1}. \quad (0.34)$$

З іншого боку, згідно з (0.24) саме ця границя була позначена як β , тому значення β можна знайти прирівнюючи вирази (0.24) та (0.34) між собою:

$$\beta = 1 / (\beta + 1). \quad (0.35)$$

Рівняння (0.35) є широко відомим рівнянням золотого перетину, розв'язком якого є відношення золотого перетину

$$\lim_{n \rightarrow \infty} b_i(n) \approx \beta = \varphi, \quad (0.36)$$

де

$$\varphi = \frac{\sqrt{5} - 1}{2} \approx 0.618. \quad (0.37)$$

З (0.31) та (0.36)-(0.37) випливає, що параметр скейлінгу MDA-мережі асимптотично має дорівнювати

$$\lim_{n \rightarrow \infty} \gamma = 2 + \varphi \approx 2.618. \quad (0.38)$$

Отака фігня, малята.

2.4 Види ступеневих розподілів

Як було зазначено в оригінальному дослідженні [13], для малих m правило MDA призводить до супер-преференційного приєднання, тобто виникає ефект "переможець отримає все". Однак, для великих значень m цей ефект замінюється ефектом "переможець отримає більше", що породжує лише прості хаби, а не супер-хаби. За результатами чисельного моделювання з [13], розподіл вузлів за ступенями є асимптотично ступеневим, тобто відповідає розподілу Юла-Саймона (0.7) з коефіцієнтом скейлінгу

$$\gamma \approx 3(1 - 0.43531e^{-0.11 \cdot m}). \quad (0.39)$$

Слід зазначити, що для більшості реальних мереж механізм приєднання за посередництвом виглядає набагато природніше, ніж механізм явного преференційного приєднання.

Втім, залежність (0.39) є суто емпіричною. Вона, вочевидь, не відображає механізму впливу керуючого параметру (m) на показник скейлінгу, та протирічить результатами аналітичного дослідження, за яким $\gamma \approx 2.618$ (0.38). Таким чином, визначення порогу, за яким суперпреферен-

ційне приєднання змінюється на звичайне преференційне, та більш точне оцінювання показника скейлінгу має суттєве практичне значення.

Але перш ніж переходити до чисельного експерименту, необхідно уточнити, що саме слід оцінювати та яким чином. Під час аналітичних досліджень виникли щонайменш три розподіли: ймовірність $p(k)$ того, що вузол має певну ступінь k , математичне очікування $k_i(n)$ ступеня вузла в залежності від розміру мережі та віку (точніше – "кроку народження" вузла) та залежність тих же ступенів вузлів від їхнього рангу. Всі ці розподіли, як прогнозується, є ступеневими, але мають різні показники скейлінгу. Тому треба визначитись, яку саме ступеневу залежність (та який саме показник скейлінгу) слід вимірювати, та як визначити інші показники, маючи один.

Взагалі, тема ступеневих розподілів сьогодні досить популярна в науці [1,12,14,19]. Щорічно з'являється безліч статей, присвячених опису ступеневих розподілів в параметрах різних природних і соціальних явищ. Вони досить поширені і при цьому у багатьох випадках їх появу важко пояснити (як розподіл міст по населенню). Пошук спільних причин утруднений ще й тим, що в численних публікаціях на тему ступеневих розподілів досі немає усталеного методичного стандарту. Зокрема, різні дослідники використовують різні типи ступеневих розподілів, через що виникає чимало плутанини. Тому треба дещо прояснити це питання.

В літературі по ступеневим законам часто використовується три різних типи розподілів:

- частотний ступеневий розподіл;
- кумулятивний ступеневий розподіл (розподіл Парето);
- ранговий ступеневий розподіл.

Одні й ті ж дані – наприклад, дані по населенню міст – можна подати у формі будь-якого з цих трьох розподілів. Зазвичай, якщо ми використовуємо один з трьох типів і отримуємо в результаті ступеневу

криву, то, використовуючи будь-який інший тип ми теж отримаємо ступеневу криву. Однак, від вибору типу розподілу залежать конкретні параметри цієї кривої і, зокрема, показник скейлінгу. Наприклад, частотний розподіл міст по населенню зазвичай відповідає ступеневим законам з показником ~ 2 , а кумулятивний та ранговий розподіли – ступеневим законам з показником ~ 1 . Тому, зустрічаючи, наприклад, десь повідомлення, що "розподіл землетрусів за кількістю жертв відповідає ступеневому закону з показником 1.8", завжди необхідно уточнювати: який саме розподіл було використано.

Далі, щоб не плутатися, будемо позначати відповідні кожному розподілу показники скейлінгу так:

- показник частотного розподілу γ ;
- показник кумулятивного розподілу χ ;
- показник рангового розподілу β .

Частотний розподіл – мабуть, найпоширеніший, але не тому, що він зручніше або корисніше інших, а за звичкою: саме такий тип розподілів традиційно застосовується в статистичній фізиці і теорії ймовірностей. Звідти він перекочував в статистику і тепер широко поширений під назвою "гістограма". Такий розподіл є вибірковою щільністю ймовірності.

Нехай, наприклад, ми вивчаємо населені пункти держави з точки зору їх населеності [19]. Для побудови частотного розподілу вздовж вісі абсцис (x) відкладаються значення досліджуваної величини за зростанням (тобто населення міст k), а вздовж вісі ординат (y) – кількості міст, які мають саме таке значення населення (n_k), або доля таких міст від загальної кількості – $p_k = n_k / n$. Зазвичай, під час побудови частотного розподілу дані групуються: шкала населеності ділиться її на якісь рівні проміжки, "кошики", а потім підраховується кількість міст, які попали в кожний з кошиків.

Частотний розподіл, зазвичай, відображається у подвійній

логарифмічній шкалі (рис.2.4). Якщо він підпорядковується ступеневому закону, то графік повинен мати вигляд спадаючої прямої з кутовим коефіцієнтом " $-\gamma$ ".

Так саме, якщо досліджується розподіл вузлів мережі за ступенями, то вздовж вісі абсцис відкладаються ступені вузлів мережі k від k_{\min} до k_{\max} (у логарифмічній шкалі), а вздовж вісі ординат – кількості вузлів n_k , які мають ступінь k (теж в логарифмічній шкалі).

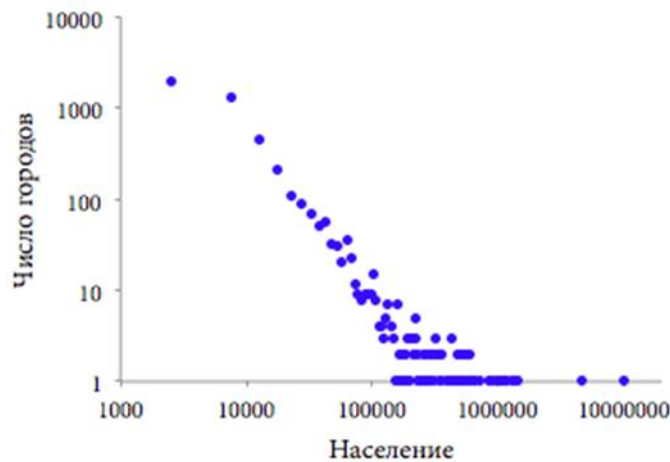


Рисунок 2.4 – Приклад частотного розподілу

В наведеному прикладі (рис.2.4), [19], пряма лінія, навколо якої групуються точки, проглядається. Її нахил відповідає приблизно $\gamma \in [1.7; 1.9]$. Однак, ми бачимо, що права частина розподілу зашумлена. Ця "борода" виникає через те, що в області високих значень статистичного параметра перестає діяти усереднення, яке ефективно згладжує криву в області низьких значень. Так, нижня лінія точок (рис.2.4) відповідає унікальним об'єктам, наступна – об'єктам кількістю 2 тощо. Іншими словами, причиною появи "бороди" є дискретність величини "кількість об'єктів" (безвідносно їхньої природи: міста, вузли мережі тощо).

Крім того, слід зазначити, що й в області малих значень досліджуваного параметру графік, зазвичай, відхиляється вниз від прямої. Це зумовлено тим, що саме в цій області наближення розподілу Юла (0.7)

ступеневим має найбільшу похибку.

Ці проблеми, змушують придивитися до інших способів подання статистики ступеневих розподілів.

Близьким родичем рангового розподілу є кумулятивний розподіл або, як ще його називають, розподіл Парето (на честь італійського економіста Вільфредо Парето, який на початку ХХ-го століття зауважив, що 80% землеволондінь в Італії знаходяться в руках 20% населення, відкуди й пішла славетна формула "80/20").

На кумулятивному розподілі по осі x відзначається величина досліджуваного параметра, у нас це населення міста, а по осі y – кількість елементів вибірки, значення цього параметра для яких, більше або дорівнює поточному x (або доля таких елементів у вибірці). Ця залежність також відображається у подвійних логарифмічних координатах (рис. 2.5).

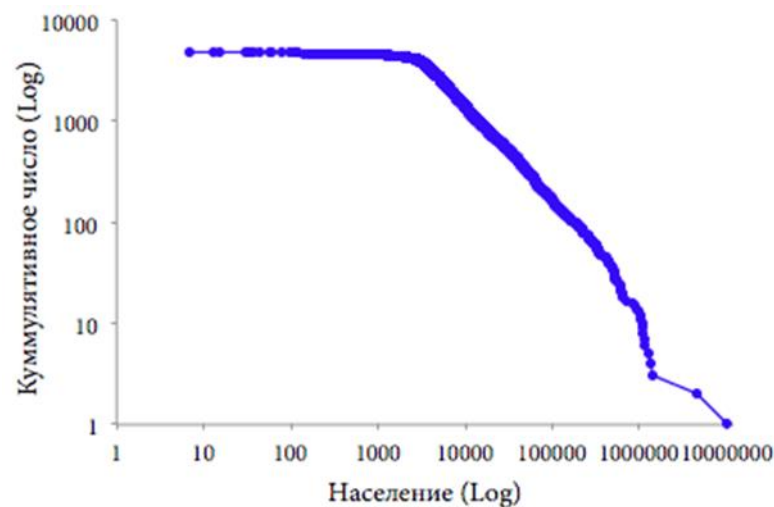


Рисунок 2.5 – Приклад кумулятивного розподілу (розподілу Парето)

З математичної точки зору, такий розподіл є емпіричною додатковою функцією розподілу $ccdf(x) = 1 - cdf(x)$, тому показники скейлінгу частотного та кумулятивного розподілів пов'язані простими співвідношеннями:

$$\chi = \gamma - 1, \quad \gamma = \chi + 1. \quad (0.40)$$

Ранговий розподіл пов'язують з ім'ям лінгвіста Джорджа Зіпфа, який

в середині 20-го століття виявив, що частота вживання слів у природних мовах відповідає ступеневим законам. Однак, в своєму аналізі Зіпф використав не частотний розподіл, а ранговий.

Повернемося до того ж прикладу – розподіл міст за населенням. Відсортуємо міста за зменшенням їх населення. Номер, який отримує в цьому списку кожне місто називається його рангом. На графіку вздовж вісі абсцис відкладається ранг, а вздовж вісі ординат – значення досліджуваного параметру (в даному прикладі – населення міст). Як і раніше, графік будується у подвійних логарифмічних координатах (рис.2.6).

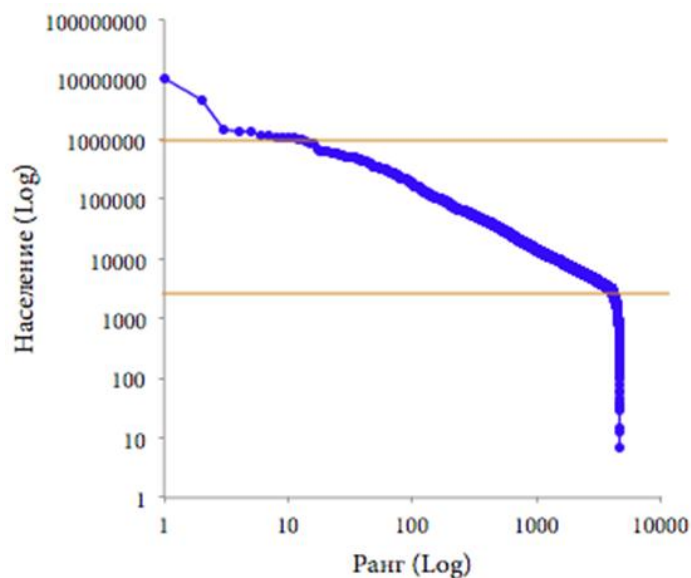


Рисунок 2.6 – Приклад рангового розподілу

Теоретично, якщо розподіл досліджуваної величини є ступеневим (0.4), то між показниками скейлінгу частотного, кумулятивного та рангового розподілів існують прості співвідношення (0.31), тобто

$$\gamma = 1 + 1 / \beta, \quad \beta = 1 / (\gamma - 1) = 1 / \chi, \quad \chi = 1 / \beta. \quad (0.41)$$

Втім, порівнюючи рис. 2.4 та рис. 2.6, легко бачити, що ранговий розподіл є набагато наочніший, ніж частотний. Можна стверджувати, що значущими залишаються всі наявні дані. Завдяки цьому існування

ступеневого закону та відповідно показник скейлінгу можна встановити з набагато більшою точністю, ніж при частотному розподілі. Так, у даному прикладі є ясно виражений проміжок (відзначений помаранчевими межами), для якого ступеневий закон виконується з великою точністю. За даними, які знаходяться на цьому проміжку, $\beta \approx 1.09$, що згідно з (0.41) відповідає $\gamma \approx 1.92$.

Порівнюючи рис. 2.5 та рис. 2.6 можна зазначити, що форма рангового розподілу дуже схожа на форму розподілу Парето, тільки перевернутого, немов осі x і y помінялися місцями. І це не випадково. Дійсно, кумулятивний розподіл є нічим іншим, як зверненим ранговим розподілом. Важко сказати, чому Парето в своїх роботах надав перевагу кумулятивному розподілу замість більш простому в побудові та більш наочному ранговому. Ймовірно, тут зіграло свою роль, що в ньому по осі абсцис відкладаються значення статистичного параметра – так, як і в частотному розподілі. Як би там не було, кумулятивні розподіли набули поширення, тому важливо їх не плутати ані з частотними, ані з ранговими.

На доданок можна відзначити ще один спосіб візуалізації розподілів: криву Лоренца [14]. В цій кривій вздовж вісі абсцис відкладають долю від загального числа елементів вибірки $0 < p < 1$, а вздовж вісі ординат – долю від сукупного значення параметру, яку мають перші Np елементів.

Так, у випадку, коли досліджуваним параметром є ступені вузлів мережі, то вздовж вісі абсцис відкладаються значення $p_i = i/n$, а вздовж вісі ординат – значення $L_{>}(n)/L(n)$, де $L_{>}(n)$ – кількість зв'язків, які мають перші i (тобто i найбагатших) вузли. Зрозуміло, що крива Лоренца за сутністю є кумулятивним ранговим розподілом. Якщо досліджуваний параметр розподілений за ступеневим законом, то крива Лоренца теж буде ступеневою з показником скейлінгу $1 - \beta$.

Підводячи підсумок, можна зробити висновок, що з всіх способів візуалізації ступеневих законів розподілу, перевагу слід віддати ранговому

через, по-перше, його простоту наочність, по-друге, через відсутність "бороди" та більшу точність оцінювання показнику скейлінгу.

3 ЧИСЛОВЕ МОДЕЛЮВАННЯ MDA-МЕРЕЖІ

3.1 Обґрунтування програмного засобу моделювання

У якості програмного засобу / середовища для програмної реалізації моделі MDA-мережі було обрано середовище MatLab [19-20].

MatLab (скорочення від англ. «Matrix Laboratory – пакет прикладних програм для вирішення задач технічних обчислень, який має вбудовану мову програмування. MatLab працює на більшості сучасних операційних систем, включаючи Linux, Mac OS, Microsoft Windows.

MatLab як мова програмування був розроблений Клівом Моулером в кінці 1970-х років коли він був деканом факультету комп'ютерних наук в Університеті Нью-Мексико. Метою розробки було надати студентам факультету можливість використання програмних бібліотек Linpack і Eispack без необхідності вивчення Фортрана. Незабаром нова мова поширилася серед інших університетів і була з великим інтересом зустрінута вченими, які працюють в області прикладної математики. До сих пір в Інтернеті можна знайти версію 1982 року, написану на Фортрані, поширювану з відкритим вихідним кодом. Інженер Джон Літл познайомився з цією мовою під час візиту Кліва Моулера в Стенфордський університет в 1983 році. Зрозумівши, що нова мова має великий комерційний потенціал, він об'єднався з Клівом Моулером і Стівом Бангертом. Спільними зусиллями вони переписали MatLab на C і заснували в 1984 компанію The MathWorks для подальшого розвитку. Ці переписані на C бібліотеки довгий час були відомі під ім'ям JASCRAS. Спочатку MatLab призначався для проектування систем управління (основна спеціальність Джона Літла), але швидко завоював популярність у багатьох інших наукових і інженерних областях. Він також широко використовувався і в освіті, зокрема, для викладання лінійної алгебри та чисельних методів.

Мова MatLab є високорівневою мовою програмування, яка інтерпретується. Середовище MatLab підтримує засновані на матрицях структури даних, широкий спектр функцій, інтегроване середовище розробки, об'єктно-орієнтовані можливості і інтерфейси до програм, написаних на інших мовах програмування.

Програми, написані на MatLab, бувають двох типів - функції і скрипти. Функції мають вхідні і вихідні аргументи, а також власний робочий простір для зберігання проміжних результатів обчислень і змінних. Скрипти ж використовують загальний робочий простір. Як скрипти, так і функції не компілюються в машинний код, а зберігаються у вигляді текстових файлів. Існує також можливість зберігати так звані *pre-processed* програми - функції і скрипти, оброблені в вид, зручний для машинного виконання. У загальному випадку такі програми виконуються швидше звичайних, особливо якщо функція містить команди побудови графіків.

Основною особливістю мови MatLab є її широкі можливості щодо роботи з матрицями, які творці мови висловили в гаслі «думай векторно» (Think vectorized).

MatLab надає користувачеві велику кількість (кілька сотень) функцій для аналізу даних, які покривають майже всі області математики, зокрема:

- матриці і лінійна алгебра – алгебра матриць, лінійні рівняння, власні значення і вектора, факторизація матриць та інші;
- багаточлени і інтерполяція - корені многочленів, операції над многочленами, їх диференціювання, інтерполяція, екстраполяція кривих тощо;
- математична статистика і аналіз даних – статистичні функції, регресія, цифрова фільтрація, швидке перетворення Фур'є та інші;
- обробка даних – набір спеціальних функцій, включаючи побудову графіків, оптимізацію, пошук нулів, чисельне інтегрування та інші;
- диференціальні рівняння – рішення диференціальних і диференціально-алгебраїчних рівнянь, диференціальних рівнянь із запізненням,

рівнянь з обмеженнями, рівнянь в часткових похідних і інші;

- розріджені матриці – спеціальний клас даних пакету MatLab, що використовується в спеціалізованих додатках;

- цілочисельна арифметика – виконання операцій цілочисельний арифметики в середовищі MatLab.

MatLab надає зручні засоби для розробки алгоритмів, включаючи високорівневі з використанням концепцій об'єктно-орієнтованого та функціонального програмування. У ньому є всі необхідні засоби інтегрованого середовища розробки, включаючи відлагоджувач та профайлер.

У складі пакету MatLab є велика кількість функцій для побудови графіків, в тому числі тривимірних, візуального аналізу даних і створення анімованих роликів. Вбудована середовище розробки дозволяє створювати графічні інтерфейси користувача з різними елементами управління, такими як кнопки, поля введення і інші.

Програми MatLab, як консольні, так і з графічним інтерфейсом користувача, можуть бути зібрані за допомогою компоненти MatLab Compiler в незалежні від MatLab виконувани програми або динамічні бібліотеки, для запуску яких на інших комп'ютерах, проте, потрібна установка вільно розповсюдженого середовища MatLab Compiler Runtime (MCR).

Пакет MatLab включає різні інтерфейси для отримання доступу до зовнішніх підпрограм, написаним на інших мовах програмування, даних. Підтримуються технології Component Object Model (COM) та Dynamic Data Exchange (DDE). Багато з цих можливостей відомі під назвою MATLAB API.

Пакет MatLab в Microsoft Windows надає доступ до програмній платформі .NET Framework. Є можливість завантажувати .NET збірки (Assemblies) і працювати з об'єктами .NET класів з середовища MatLab. У версії MatLab 7 підтримується .NET Framework версій до 4.0 включно.

Пакет MatLab містить функції, які дозволяють йому отримувати доступ до інших додатків середовища Windows, так само як і цим програмам отримувати доступ до даних MatLab, за допомогою технології динамічного обміну даними (DDE).

В MatLab існує можливість викликати методи веб-сервісів. Спеціальна функція створює клас, ґрунтуючись на методах API веб-сервісу.

MatLab взаємодіє з клієнтом веб-сервісу за допомогою прийняття від нього посилки, їх обробки і посилки відповіді. Підтримуються наступні технології: Simple Object Access Protocol (SOAP) і Web Services Description Language (WSDL).

Пакет MatLab включає інтерфейс взаємодії з зовнішніми додатками, написаними на мовах C і Фортран. Здійснюється це взаємодія через MEX-файли. Існує можливість виклику підпрограм, написаних на C або Фортрані з MatLab, як ніби це вбудовані функції пакета. MEX-файли являють собою спільні бібліотеки, які можуть бути завантажені і виконані інтерпретатором, вбудованим в MatLab. MEX-процедури мають також можливість викликати вбудовані команди MatLab.

Інтерфейс MatLab, що відноситься до загальних DLL дозволяє викликати функції, що знаходяться в звичайних DLL-бібліотеках, прямо з MatLab. Ці функції повинні мати C-інтерфейс.

Крім того, в MatLab є можливість отримати доступ до його вбудованих функцій через C-інтерфейс, що дозволяє використовувати функції пакета в зовнішніх додатках, написаних на C. Ця технологія в MatLab називається C Engine.

Для MatLab є можливість створювати спеціальні набори інструментів (toolbox), що розширюють його функціональність. Набори інструментів є колекції функцій, написаних на мові MatLab для вирішення певного класу задач. Компанія Mathworks поставляє набори інструментів, які використовуються в багатьох областях, включаючи наступні:

- цифрова обробка сигналів, зображень та даних: DSP Toolbox, Image Processing Toolbox, Wavelet Toolbox, Communication Toolbox, Filter Design Toolbox – набори функцій, що дозволяють вирішувати широкий спектр завдань обробки сигналів, зображень, проектування цифрових фільтрів і систем зв'язку;

- системи управління: Control Systems Toolbox, μ -Analysis and Synthesis Toolbox, Robust Control Toolbox, System Identification Toolbox, LMI Control Toolbox, Model Predictive Control Toolbox, Model-Based Calibration Toolbox – набори функцій, що полегшують аналіз і синтез динамічних систем, проектування, моделювання та ідентифікацію систем управління, включаючи сучасні алгоритми управління, такі як Робастне управління, H_{∞} -управління, μ -синтез та інші;

- фінансовий аналіз: GARCH Toolbox, Fixed-Income Toolbox, Financial Time Series Toolbox, Financial Derivatives Toolbox, Financial Toolbox, Datafeed Toolbox – набори функцій, що дозволяють швидко і ефективно збирати, обробляти і передавати різну фінансову інформацію;

- аналіз і синтез географічних карт, включаючи тривимірні: Mapping Toolbox;

- збір і аналіз експериментальних даних: Data Acquisition Toolbox, Image Acquisition Toolbox, Instrument Control Toolbox, Link for Code Composer Studio – набори функцій, що дозволяють зберігати і обробляти дані, отримані в ході експериментів, в тому числі в реальному часі. Підтримується широкий спектр наукового і інженерного вимірювального обладнання;

- візуалізація і уявлення даних: Virtual Reality Toolbox дозволяє створювати інтерактивні світи і візуалізувати наукову інформацію за допомогою технологій віртуальної реальності і мови VRML;

- засоби розробки: MatLab Builder for COM, MatLab Builder for Excel, MatLab Builder for NET, MatLab Compiler, Filter Design HDL Coder – набори функцій, що дозволяють створювати незалежні програми з

середовища MatLab;

- взаємодія з зовнішніми програмними продуктами: MatLab Report Generator, Excel Link, Database Toolbox, MatLab Web Server, Link for ModelSim – набори функцій, що дозволяють зберігати дані в різних видів таким чином, щоб інші програми могли з ними працювати;

- бази даних: Database Toolbox – інструментальні засоби для роботи з базами даних;

- наукові та математичні пакети: Bioinformatics Toolbox, Curve Fitting Toolbox, Fixed-Point Toolbox, Fuzzy Logic Toolbox, Genetic Algorithm and Direct Search Toolbox, OPC Toolbox, Optimization Toolbox, Partial Differential Equation Toolbox, Spline Toolbox, Statistic Toolbox, RF Toolbox – набори спеціалізованих математичних функцій, що дозволяють вирішувати широкий спектр наукових і інженерних задач, включаючи розробку генетичних алгоритмів, вирішення завдань в приватних похідних, цілочисельні проблеми, оптимізацію систем та інші;

- нейронні мережі: Neural Network Toolbox – інструменти для синтезу та аналізу нейронних мереж;

- нечітка логіка: Fuzzy Logic Toolbox – інструменти для побудови та аналізу нечітких множин;

- символічні обчислення: Symbolic Math Toolbox – інструменти для символічних обчислень з можливістю взаємодії з символічним процесором програми Maple.

Існує багато пакетів, програмно сумісних з MatLab, або близьких за функціональністю: SciLab, GNU Octave, R та інші. Втім, Matlab залишається найбільш поширеним. Крім того, завдяки всебічній документованості та наявності великої кількості прикладів, MatLab якнайкраще підходить для навчальних цілей.

3.2 Моделювання MDA-мережі

Першим кроком під час реалізації моделі мережі з приєднанням, заснованим на посередництві, є вибір способу представлення мережі. Як було зазначено під час аналізу проблемної області, існує багато таких способів, зокрема, матриця суміжності, матриця інцидентності, список суміжності, список ребер.

Оскільки згідно з алгоритмом приєднання через посередництво типовою операцією є визначення переліку вузлів, інцидентних до посередника, та ступеню цих вузлів, то найбільш доцільним способом зберігання структури мережі міг би бути список суміжності.

Згідно з цим способом кожному вузлу мережі відповідає рядок списку, в якому зберігається список суміжних вузлів. Така структура даних являє собою «масив списків». Оскільки MatLab напряду не підтримує таку структуру, як список, чи масив списків, то структуру мережі доведеться зберігати у звичайному двовимірному масиві, тобто матриці.

Як було визначено під час аналізу проблемної області, витрати пам'яті на зберігання структури мережі у вигляді списку суміжності складають $O(|L|)$. Реалізація списку суміжності двовимірним масивом призведе до збільшення витрат пам'яті до $O(n \cdot k_{\max})$, де k_{\max} – максимальний ступінь для вузлів мережі. Згідно з аналітичними результатами досліджень (0.29), (0.36), ця величина дорівнює

$$M \approx nm(n/n_0)^\varphi = O(n^{1+\varphi}) \approx O(n^{1.618}). \quad (0.42)$$

Втім, поелементний доступ до масиву (під час додавання номерів вузлів у кінець рядка) є значно повільнішим, ніж векторні операції (заповнення рядка, чи стовпця). Тому для зберігання структури мережі було використано звичайну матрицю суміжності (з витратами пам'яті $O(n^2)$).

Схема алгоритма цієї програми наведена на рис. 3.1.

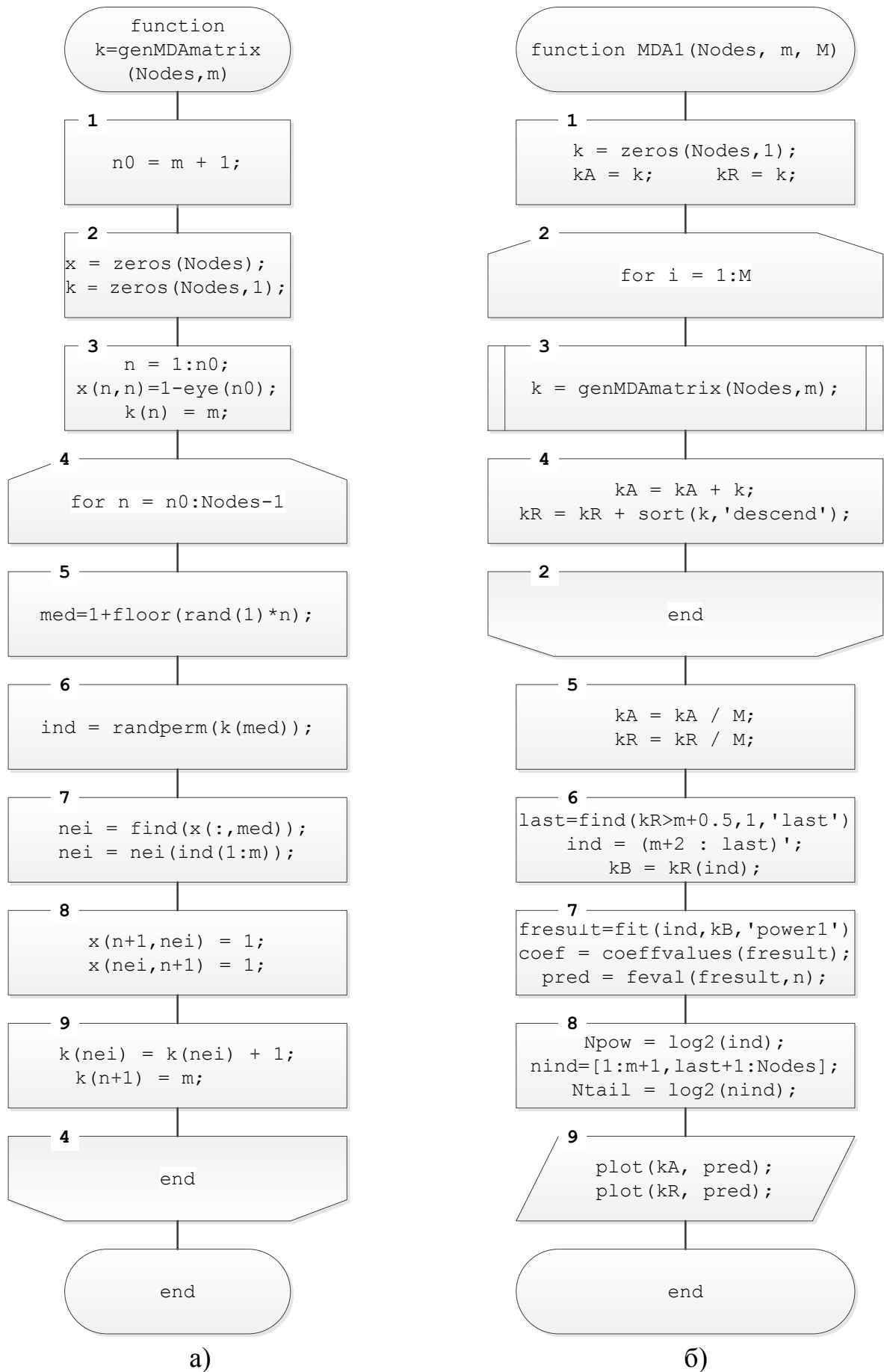


Рисунок 3.1 – Блок-схеми алгоритмів функцій: а) genMDAmatrix, б) MDA1

Функція безпосереднього генерування мережі має назву `genMDAmatrix`. Вхідними параметрами для неї є `Nodes` – кількість вузлів у мережі та керуючий параметр `m` – кількість зв'язків для нових вузлів.

Двовимірний масив $x(i, j)$ являє собою матрицю суміжності. В окремому масиві $k(i)$ зберігаються поточні значення ступенів вузлів.

Спочатку формується початкова мережа – повний граф з $n_0 = m + 1$ вузлів. Потім у циклі додаються нові вузли. Параметр циклу (n) дорівнює кількості існуючих вузлів (тобто додається вузол з номером $n + 1$). На кожній ітерації циклу обирається випадковий вузол-посередник `med`, визначаються його сусіди (тобто ненульові елементи стовпця з номером `med`), їхні номери переміщуються та визначаються m обраних, до яких й приєднується вузол $n + 1$. Для генерації випадкових чисел застосовується вбудована функція `rand()`.

Головна функція має назву `MDA1`. Вхідними параметрами для неї є `Nodes` – розмір мереж, керуючий параметр `m` – кількість зв'язків для нових вузлів та `M` – кількість мереж, що генеруються.

У циклі довжиною `M` генеруються мережі шляхом виклику функції. Отримані ступені вузлів (k) зберігаються та усереднюються у масиві `kA` – за спаданням їхнього віку. Також зберігаються та усереднюються ступені вузлів, відсортовані за спаданням рангу – у масиві `kR`.

Потім формується `kB` – фрагмент масива `kR`, який містить елементи з рангом, більшим за $m + 1$ та з значенням, більшим за $m + 0.5$. За допомогою вбудованих функцій `fit`, `coeffvalues` та `feval` визначається ранговий показник скейлінгу β (змінна `coef(:, 2)`) та обчислюються прогнозовані значення регресійної змінної (`pred`). Отримані результати відображаються графічно вбудованою функцією `plot`.

Реалізація інших елементів програми моделювання мережі не потребує зайвих коментарів. Текст програми наведений у додатку А.

За допомогою розробленої програми були проведені чисельні експерименти. Спочатку моделювались мережі з $N_{\text{Nodes}}=4096$ вузлів та досліджувалось, чи можна вважати їх безмасштабними (тобто чи дійсно розподіл вузлів за ступенями є ступеневим). Рангові розподіли ступенів вузлів для випадків $m=1$ та $m=5$ наведені на рис. 3.2 та рис. 3.3. На рис. 3.4 наведений розподіл ступенів вузлів за їхнім віком для $m=100$.

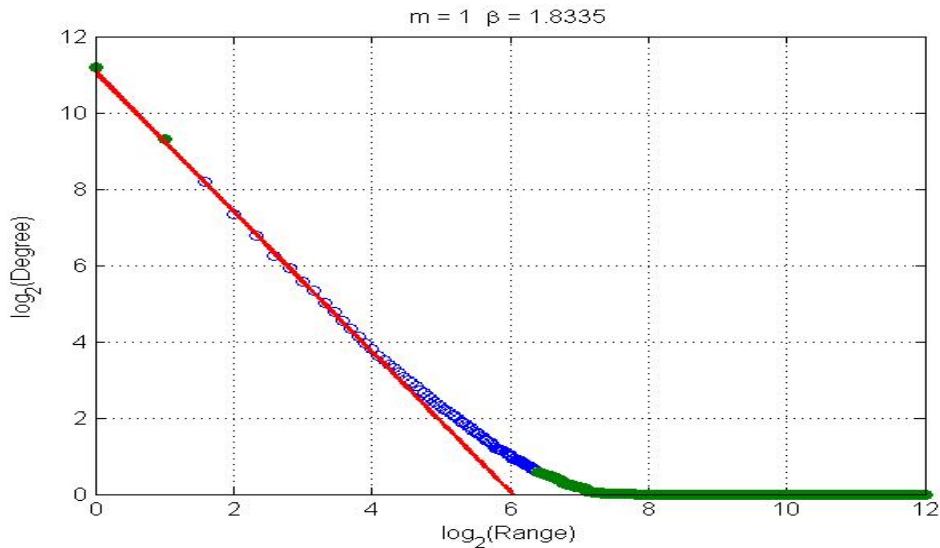


Рисунок 3.2 – Ранговий розподіл вузлів за ступенями при $m=1$

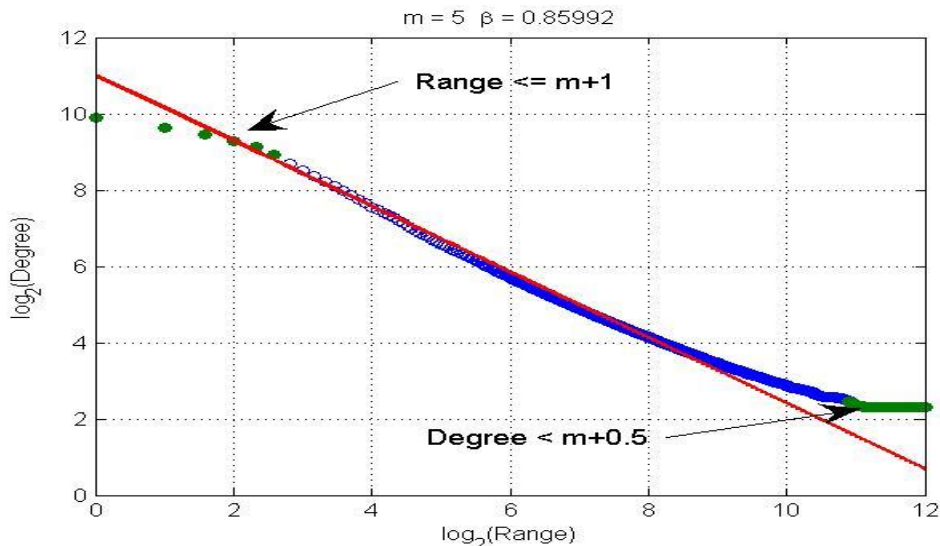


Рисунок 3.3 – Ранговий розподіл вузлів за ступенями при $m=5$

Згідно з отриманими результатами, графіки містять три частки: початкову, середню, та кінцеву. Вузли, що відносяться до середньої

частки, позначені синім кольором, а інші – зеленим. Легко бачити, що рангові розподіли вузлів середньої частки (синіх) з великою точністю підпорядковуються ступеневому закону. Регресійні прямі зображені на рис. 3.2 - рис. 3.4 червоним кольором, а відповідні показники скейлінгу вказані у заголовках графіків.

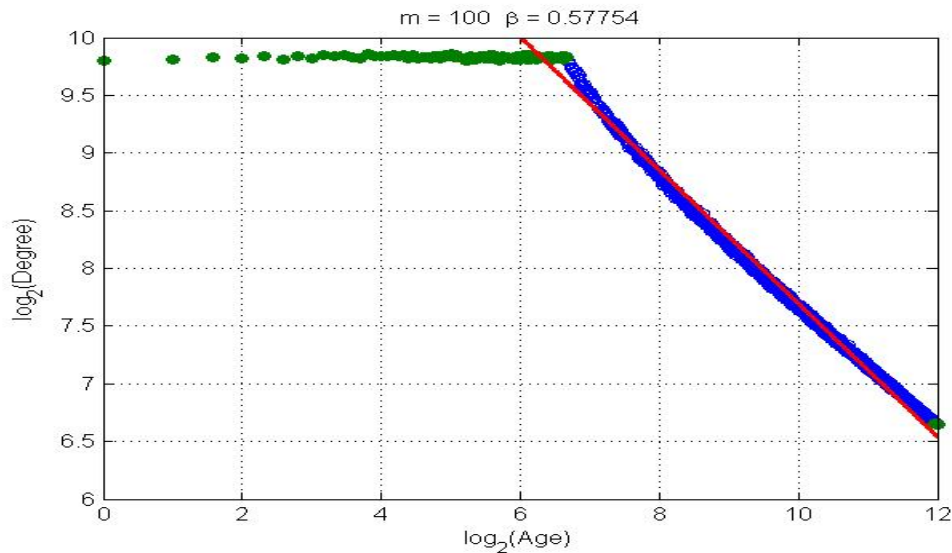


Рисунок 3.4 – Розподіл вузлів за віком при $m=100$

Початкові вузли (перші $n_0 = m + 1$) суттєво відхиляються від ступеневого розподілу, тобто регресійних прямих (рис. 3.2 - рис. 3.4), "недобираючи" зв'язків. Втім, це явище цілком природне, зумовлене тим, що "справжнім" розподілом є все ж таки розподіл Юла (0.7), який збігається з ступеневим (0.4) лише асимптотично. Крім того, згідно з (0.29) діє операція округлення маленьких значень номерів до n_0 . Саме тому регресійна пряма будувалася без врахування початкових (зелених) вузлів.

З іншого боку, кінцеві вузли, вочевидь, теж не підпорядковуються ступеневому розподілу. Наявність такого "хвоста" не має теоретичного обґрунтування, втім її можна пояснити суто емпірично: у супер-преференційній мережі вузли, що з'явилися пізніше, не встигають набрати зв'язки, додаткові, до тих m , які вони мають під час "народження". Під час чисельного моделювання поріг відсікання було встановлено рівним $m + 0.5$. Згідно з результатами моделювання (рис. 3.2 – рис. 3.4) таке значення

порогу дозволяє досить чітко відділити "молоді" вузли (зелений хвіст) від вузлів "середнього віку" (сині, які відповідають ступеневим розподілам). Можна зазначити, що доля хвостової частки вузлів швидко скорочується зі зростанням m : для $m = 1$ хвіст містить майже 98% від загальної чисельності вузлів, для $m = 5$ – трохи більше половини, а для $m = 100$ – менше 10%.

3.3 Емпіричне дослідження показнику скейлінгу моделі приєднання за посередництвом

Результати первинного дослідження (рис. 3.2 - рис. 3.4) показали, що розподіл вузлів за ступенем можна вважати ступеневим, проте показник скейлінгу по-перше, відрізняється від теоретично прогнозованого (0.36) $\varphi = 0.618$, а по-друге, суттєво залежить від керуючого параметра m : він дорівнює 1.8335, 1.2861, 0.8599, 0.6731, 0.5775 для $m = 1, 2, 5, 10, 100$ відповідно.

У зв'язку з цим було проведено дослідження залежності показника скейлінгу (β) від керуючого параметра m (кількості сусідів посередника, з якими встановлюються зв'язки з новими вузлами). Для цього було розроблено функцію MDA2, блок-схема якої представлена на рис. 3.5, а текст – у додатку Б.

Для заданого набору значень $m = [1, 2, 3, 4, 5, 8, 10, 12, 15, 20, 30, 50, 100]$ функцією `genMDAmatrix` генеруються мережі з `Nodes = 4096` вузлів (по $M = 100$ мереж для кожного значення m). Для кожної мережі (окремо) обчислюється показник скейлінгу (за допомогою вбудованих функцій `fit`, `feval` та `coeffvalues`). Результати досліджень показані на рис. 3.6. Сині точки відповідають окремим мережам, червоні хрестики позначають арифметичні середні (для кожного m). Оскільки для малих значень m скейлінг має дуже високу волатильність, то замість арифметичного усереднення більш доцільним є медіанне усереднення. Лінія пурпурового кольору з'єднує медіани значень показника скейлінгу.

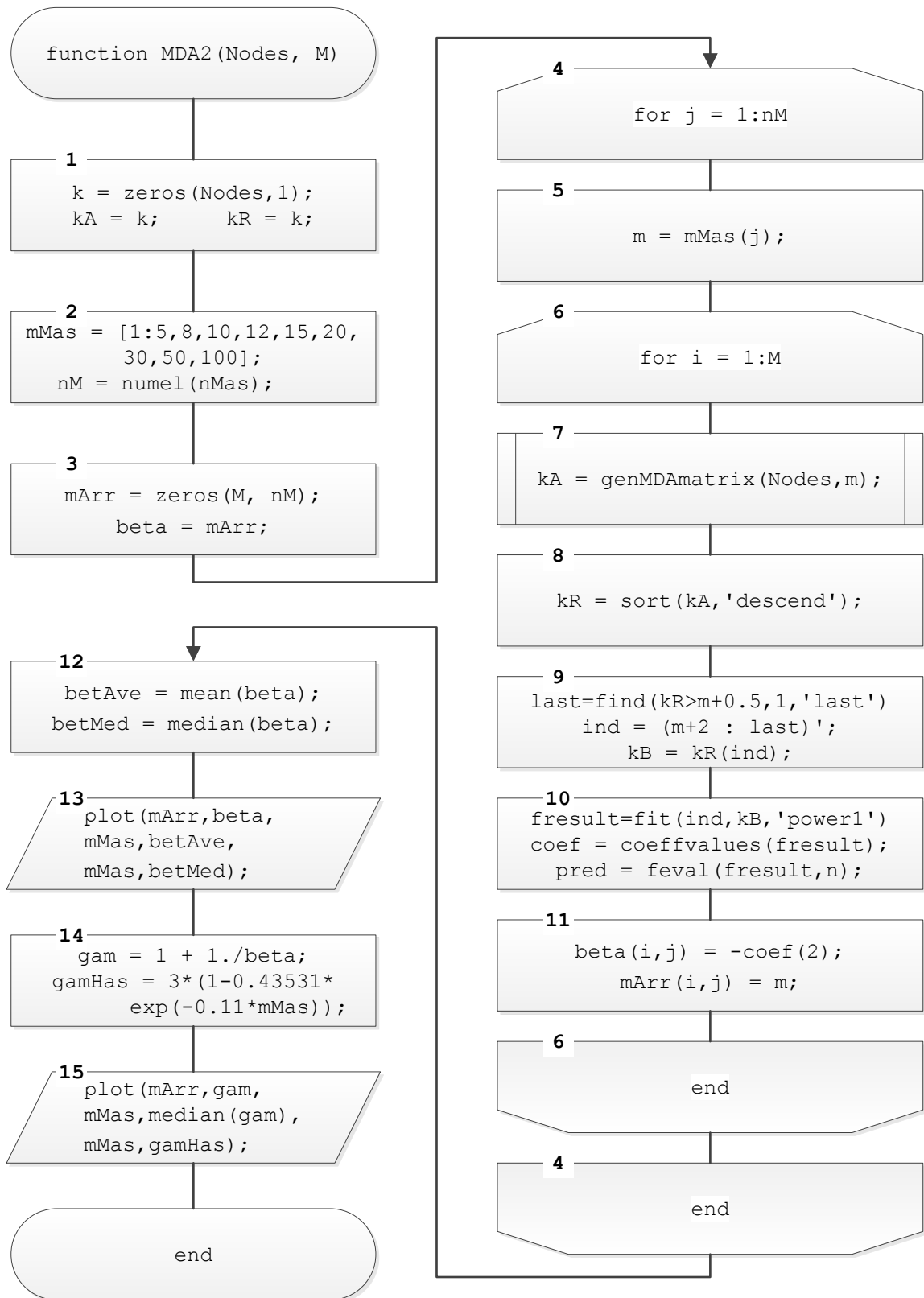


Рисунок 3.5 – Блок-схема функції MDA2

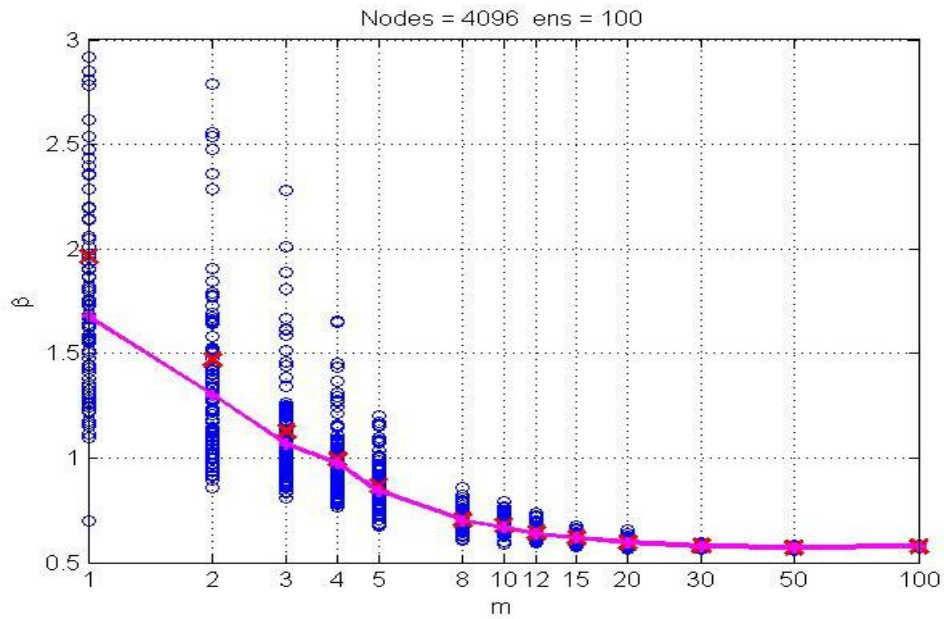


Рисунок 3.6 – Залежність показника скейлінгу рангового розподілу ступенів вузлів від керуючого параметра m

На рис. 3.7 показано залежність показника скейлінгу частотного розподілу ступенів вузлів $\gamma = 1 + 1/\beta$ від керуючого параметру m . Для співставлення на цьому ж рисунку чорним кольором показано емпіричну залежність (0.39) з оригінальної роботи [13].

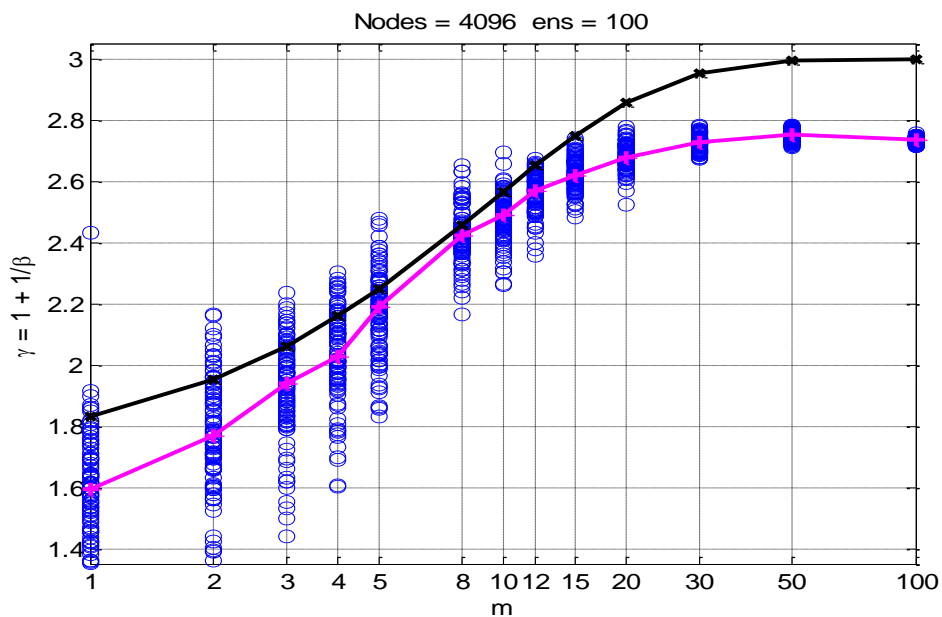


Рисунок 3.7 – Залежність показника скейлінгу частотного розподілу ступенів вузлів від керуючого параметру m

Аналізуючи отримані результати (рис. 3.6 – рис. 3.7), можна зробити певні висновки:

- при малих значеннях керуючого параметра варіативність показників скейлінгу є настільки високою, що не можна казати про більше-менше достовірне їхнє значення;

- більше того: при $m < 5$ ступінь вузлів мережі навіть не має математичного очікування (бо $\beta < 1$ та, відповідно, $\gamma < 2$). Це значить, що при малих значеннях m мережа є суперпреференційною, тобто невелика кількість вузлів-переможців перебирає на себе майже всі зв'язки у мережі;

- зі збільшенням m доля вузлів, розподіл ступенів яких є ступеневим, збільшується (рис. 3.2 – рис. 3.4), а варіативність показнику скейлінгу зменшується (тобто отримані оцінки стають все більше достовірними);

- залежність $\beta(m)$ (та, відповідно, $\gamma(m)$) не є монотонною; вона сягає мінімуму (відповідно, максимуму) при $m \approx 50$, втім, цей ефект може бути зумовлений обмеженістю розміру мереж, тому потребує додаткової перевірки;

- отримані емпіричні залежності $\beta(m)$, $\gamma(m)$ не відповідають ані розробленій у розділі 2.3 теоретичній моделі (за якою $\beta \propto 0.618$, $\gamma \propto 2.618$), ані моделі (0.39) з [13] (за якою $\beta \propto 0.5$, $\gamma \propto 3$).

З огляду на вищевказане, модель приєднання за посередництвом потребує більш ґрунтовного аналітичного дослідження її властивостей.

ВИСНОВКИ

Розглянуто проблему моделювання безмасштабних мереж. Властивості мереж суттєво залежать від правила приєднання нових вузлів до мережі. Був проведений огляд літературних джерел, проаналізовані існуючі моделі безмасштабних мереж та відповідні правила приєднання.

Найпростішою та найпоширенішою моделлю безмасштабних мереж є модель Барабаші-Альберт, яка ґрунтується на правилі переважного приєднання. Проте, це правило в свою чергу ґрунтується на доступності інформації щодо властивостей всіх існуючих вузлів мережі, зокрема ступенів вузлів. З іншого боку, для використання правил приєднання, заснованих на посередництві, така інформація не потрібна, бо вузол-посередник обирається випадково. Таким чином, правила приєднання, засновані на посередництві, є більш природними та краще відповідають властивостям реальних мереж, що є їх головною та безсумнівною перевагою.

Під час проведення досліджень з'ясовано, що однією з моделей приєднання, заснованих на посередництві, є модель MDA. Було проведено теоретичне дослідження властивостей мереж, породжених цим правилом, внаслідок якого встановлено, що ці мережі мають бути асимптотично безмасштабними. Визначено показник скейлінгу.

В рамках експериментальної частини досліджень було розроблено алгоритми генерації мереж за правилом MDA та здійснено їхню програмну реалізацію у середовищі MatLab. За результатами чисельного моделювання встановлено, що отримані мережі дійсно є безмасштабними, проте їхні властивості суттєво залежать від керуючого параметру – кількості зв'язків, між існуючими вузлами та новим. При малих значеннях керуючого параметра мережа є суперпреференційною, тобто невелика кількість вузлів-переможців перебирає на себе майже всі зв'язки у мережі. Зі збільшенням цього параметру доля вузлів, розподіл ступенів яких є ступеневим, збільшується, а варіативність показнику скейлінгу зменшується.

ПЕРЕЛІК ПОСИЛАНЬ

1. Newman, M.E.J. (2005) Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 2005, 46(5). p.323-351.
2. Dorogovtsev S. N., Mendes J. F. F. (2003). *Evolution of Networks: From Biological Networks to the Internet and WWW* – Oxford, USA: Oxford University Press, 2003. – 280 p.
3. Albert, R., Barabasi A.-L (2002). Statistical mechanics of complex networks: *Rev. Mod. Phys.* - V. 74. - p. 42-97.
4. Bianconi, Ginestra; Barabási, Albert-László (2001). "Competition and multiscaling in evolving networks". *Europhysics Letters*. 54 (4): 436–442. doi:10.1209/epl/i2001-00260-6
5. Dorogovtsev, S. N., Mendes, J. F. F. and Samukhin, A. N. (2000), Structure of growing networks with preferential linking, *Phys. Rev. Lett.*, 85, 4633.
6. Caldarelli, Guido; Catanzaro, Michele (2012). *Networks: A Very Short Introduction*. Oxford University Press. p. 78
7. Choromański, K.; Matuszak, M.; Miękisz, J. (2013). Scale-Free Graph with Preferential Attachment and Evolving Internal Vertex Structure. *Journal of Statistical Physics*. 151 (6): 1175–1183.
8. Kunegis, Jérôme & Blattner, Marcel & Moser, Christine. (2013). Preferential Attachment in Online Networks: Measurement and Explanations. *WebSci13 ACM Web Science Conference 2013*. 13. doi: 10.1145/2464464.2464514.
9. Krapivsky, P. L.; S. Redner; F. Leyvraz (2000). "Connectivity of Growing Random Networks". *Phys. Rev. Lett.* 85: 4629–4632. doi: 10.1103/physrevlett.85.4629
10. R. Noldus, P. Van Mieghem, "Assortativity in complex networks", *J. Complex Networks*, vol. 3, pp. 507-542, 2015.
11. Dangalchev Ch., Generation models for scale-free networks, *Physica*

A 338, 659 (2004).

12. Broido, Anna; Aaron Clauset: "Scale-free networks are rare". *Nature Communications*. 10 (1): 1017. (2019) doi:10.1038/s41467-019-08746-5.

13. Hassan, M. K.; Islam, Liana; Arefinul Haque, Syed (2017). "Degree distribution, rank-size distribution, and leadership persistence in mediation-driven attachment networks". *Physica A*. 469: 23–30. doi:10.1016/j.physa.2016.11.001.

14. Vadim Shergin, Serhii Udovenko, and Larysa Chala: Assortativity Properties of Barabási-Albert Networks. In *Data-Centric Business and Application*. Springer. (2020).

15. Shergin V. L., Chala L. E., Udovenko S. G. (2018). Fractal dimension of infinitely growing discrete sets. *Advanced Trends in Radioelectronics Telecommunications and Computer Engineering (TCSET)*, no. 348.

16. Shergin V. L., Chala L. E. (2017). The concept of elasticity of scale-free networks. *Proc. Problems of Infocommunications. Science and Technology (PIC S&T)*, no. 62.

17. Kumar, Ravi; Raghavan, Prabhakar (2000). *Stochastic Models for the Web Graph (PDF)*. *Foundations of Computer Science, 41st Annual Symposium on*. pp. 57–65. doi:10.1109/SFCS.2000.892065

18. Callaway, Duncan S.; Newman, M. E. J.; Strogatz, S. H.; Watts, D. J. (2000). "Network Robustness and Fragility: Percolation on Random Graphs". *Physical Review Letters*. 85 (25): 5468–71. doi:10.1103/PhysRevLett.85.5468.

19. Степенные распределения [Электронный ресурс]- Режим доступа: http://cognitivist.ru/er/kernel/prologi_04_powerlaw_distributions.xml-2013 р. - Загл. с экрана.

20. Шергін, В.Л., Удовенко, С.Г., Загребельна М.Ф. Структура екстремально асортативних масштабно-інваріантних мереж, Фізико-технологічні проблеми передавання, оброблення та зберігання інформації в інфокомунікаційних системах. Чернівці: 2018.

21. Кроновер, Р. Фракталы и хаос в динамических системах. Основы

теории [Текст] : пер. с англ. – М.: Постмаркет, 2000. –352с.

22. Шредер, М. Фракталы, хаос, степенные законы. Миниатюры из бесконечного рая [Текст] : пер. с англ. – Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. –528с.

23. Касьянов В.Н., Евстигнеев В.А. Графы в программировании: обработка, визуализация и применение.– СПб.: БХВ-Петербург, 2003. 1104 с.

24. MATLAB [Электронный ресурс] - Режим доступа: <https://ru.wikipedia.org/wiki/MATLAB>- 2019 г. - Загл. с экрана.

25. Matlab [Электронный ресурс] - Режим доступа: <http://matlab.ru/products/matlab> - 2019 г. - Загл. с экрана.