ДИПЛОМНА РОБОТА
БАКАЛАВРА

ДРБ. ІП - 49.00.00.000 ПЗ

Група ІП-19-1К

Скула Аліна

2022

Міністерство освіти і науки України
Івано-Франківський національний технічний університет нафти і газу

Інститут інформаційних технологій

Кафедра інженерії програмного забезпечення

УДК 004.4

ДИПЛОМНА РОБОТА БАКАЛАВРА

Тема: «Інформаційне моделювання соціальної активності в мережі»

(назва згідно з наказом ректора)

Спеціальність – 121 “Інженерія програмного забезпечення”

ПОЯСНЮВАЛЬНА ЗАПИСКА

ДРБ.ІП - 49.00.00.000 ПЗ

(позначення)

Студент

ІП-19-1К /А.В.Скула/
(шифр групи) (підпис) (дата) (розшифровка підпису)

Керівник роботи

доц., к.т.н. /Р.Б.Вовк/
(посада) (підпис) (дата) (розшифровка підпису)

Консультанти:

з питань нормоконтролю

доц., к.т.н /В.Я. Піх/
(посада) (підпис) (дата) (розшифровка підпису)

Допускається до захисту
Завідувач кафедри

проф., д.т.н. /В.І.Шекета/
(посада) (підпис) (дата) (розшифровка підпису)

2022
(рік)

Івано-Франківський національний технічний університет нафти і газу

Інститут, факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Ступінь вищої освіти бакалавр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:
Зав. кафедрою **ІПЗ**
проф. В.І. Шекета
“ ____ ” _____2022 р.

ЗАВДАННЯ

НА ДИПЛОМНУ РОБОТУ БАКАЛАВРА СТУДЕНТОВІ

Скулі Аліні Володимирівні

(прізвище, ім’я, по-батькові)

1. Тема проекту (роботи) Інформаційне моделювання соціальної активності в мережі
керівник проекту (роботи) Вовк Роман Богданович, доцент, к.т.н.

затверджені наказом закладу вищої освіти від **“** __ ” _ травня__ 2022 р. № _____

2. Строк подання студентом проекту (роботи) 17 червня 2022 р.

3. Вихідні дані до проекту (роботи) Результати і матеріали отримані під час проходження переддипломної практики

4. Зміст розрахунково - пояснювальної записки (перелік питань, які потрібно розробити)

1. Огляд існуючих концепцій, рішень та сервісів в даній області
2. Побудова моделі або алгоритму власного рішення
3. Документування реалізації власного оригінального рішення вибраними засобами

5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)

1. Стратегія впровадження 1: обидва алгоритми працюють паралельно (рис. 2.8)
2. Стратегія впровадження 2: послідовність алгоритму SBSID та алгоритму FBSID (рис.2.9)
3. Стратегія впровадження 3: послідовність алгоритму FBSID та алгоритму SBSID (рис. 2.10)
4. Надсилання m повідомлень до вузла (рис.2.1)
5. Кількість ізольованих вузлів у різних налаштуваннях повідомлень (рис. 3.14)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання 25 травня 2022 р.

Керівник _____
(підпис)

Завдання прийняв до виконання _____
(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів проекту	Примітка
1	Визначення та обґрунтування теми роботи	26.05.2022	виконано
2	Огляд існуючих концепцій, рішень та сервісів в даній області	29.05.2022	виконано
3	Побудова моделі або алгоритму власного рішення	30.05.2022	виконано
4	Документування реалізації власного оригінального рішення моделювання соціальної мережі вибраними засобами	04.06.2022	виконано
5	Тестування ефективності розробленого програмного додатку	10.06.2022	виконано
6	Оформлення пояснювальної записки кваліфікаційної роботи	17.06.2022	виконано

Студент-дипломник _____
(підпис)

Керівник роботи _____
(підпис)

АНОТАЦІЯ

В дипломній роботі запропоновано обчислювальну модель для виявлення соціально ізольованих осіб у соціальній мережі. Для цього було запропоновано два оригінальних алгоритми. Основна мета цього дослідження полягає в тому, щоб запропонувати обчислювальну модель для вирішення проблеми виявлення соціальної ізоляції як зі структурної, так і з точки зору особливостей. Для досягнення цієї мети спроектовано три основні кроки. Першим кроком є представлення спільноти за допомогою соціального графа. Другий, полягає у визначенні набору показників для вимірювання якості та кількості зв'язків між людьми. Нарешті, останній крок – полягав в розробці алгоритмів для ефективного визначення соціально ізольованих осіб на основі топології мережі та їх власних особистих особливості.

КЛЮЧОВІ СЛОВА: СОЦІАЛЬНА МЕРЕЖА, МОДЕЛЮВАННЯ, ОБЧИСЛЮВАЛЬНА МОДЕЛЬ, ТОПОЛОГІЯ, ЯКІСТЬ МЕРЕЖІ

ANNOTATION

This study proposes a new computational model for identifying socially isolated individuals in the social network. Two original algorithms have been proposed for this purpose. The main purpose of this study is to propose a computational model to solve the problem of identifying social isolation, both structurally and in terms of features. Three main steps have been designed to achieve this goal. The first step is to represent the community through a social graph. The second is to define a set of indicators to measure the quality and quantity of relationships between people. Finally, the last step - was to develop algorithms for the effective identification of socially isolated individuals based on network topology and their own personal characteristics.

KEY WORDS: KEY WORDS: SOCIAL NETWORK, SIMULATION, COMPUTER MODEL, TOPOLOGY, NETWORK QUALITY

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ 7
ВСТУП 8
1 ОГЛЯД ІСНУЮЧИХ ПРОГРАМНИХ РІШЕНЬ ТА СЕРВІСІВ В ПРЕДМЕТНІЙ ОБЛАСТІ 11
1.1 Аналіз соціальних мереж 11
1.2 Мотивація дослідження 11
1.3 Постановка проблеми 12
1.4 Заходи соціальної ізоляції 16
1.5 Виявлення соціальної ізоляції на основі опитувань 17
1.6 Виявлення викидів на основі графіка 19
1.7 Базові поняття для побудови соціальної мережі 20
2 АЛГОРИТМІЧНІ ТА МАТЕМАТИЧНІ РІШЕННЯ ДЛЯ ЗАДАЧ МОДЕЛЮВАННЯ СОЦІАЛЬНОЇ МЕРЕЖІ 25
2.1 Пропонована схема реалізації алгоритму 25
2.2 Пропонована модель соціальної мережі 29
3 АЛГОРИТМІЧНЕ ТА ПРОГРАМНЕ РІШЕННЯ ЗАДАЧІ МОДЕЛЮВАННЯ СОЦІАЛЬНОЇ АКТИВНОСТІ В МЕРЕЖІ 36
3.1 Складові програмного рішення 36
3.2 Налаштування моделі 36
3.3 Експериментальні дослідження та імітаційне моделювання 37
3.4 Порівняння результатів з точним методом 50
3.5 Перспективи вдосконалення запропонованого алгоритму 58
ВИСНОВКИ 60
СПИСОК ПОСИЛАНЬ НА ДЖЕРЕЛА 62
БІБЛІОГРАФІЧНА ДОВІДКА 66

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

СМ – соціальна мережа

БД – База даних

СУБД - Система управління базами даних

ІМ – інформаційне моделювання

ВСТУП

Люди за своєю суттю товариські як члени цього суспільства, і взаємодія та співпраця необхідні для підтримки життя суспільства. За допомогою комунікації люди можуть отримати величезну кількість інформації, щоб отримати цінні уроки з досвіду інших або розробити нові ідеї, щоб краще функціонувати різними способами. Незважаючи на цей факт, люди, особливо літні, час від часу почуватимуться самотніми та ізольованими, і це відчуття стане справжнім викликом, коли воно стане постійним.

Мережа може бути представлена графом, що включає вузли та ребра як зв'язки між парою вузлів . Отже, соціальна мережа — це складна підмножина мережі, яка являє собою сукупність соціальних та особистих взаємодій з різними цілями, що представляють стосунки в реальному світі, в яких люди можуть ділитися своїми думками, експериментами та почуттями . Соціальна мережа складається з соціальних акторів з різними культурами, професіями, характеристиками та освітніми рівнями. Зв’язок між цими акторами різний за ступенем сили та важливості. Деякі з цих акторів впливовіші за інших і значно впливають на коло своїх друзів. Іншими словами, ці видатні особи діють як лідери, щоб контролювати мережу і мотивувати інших членів підкорятися їм і встановлювати більше зв’язків.

Соціальна мережа — це складна структура з деякими особливими характеристиками, такими як структура спільноти, ефект малого світу та розподіл ступенів :

• Структура спільноти є невід'ємною частиною соціальної мережі, яка допомагає щоб краще розуміти поведінку користувачів і побудову мережі. На графі, що представляє соціальну мережу, коефіцієнт кластеризації допомагає виявити спільноти, вимірюючи тенденцію вузла приєднатися до певного кластера.

• Феномен малого світу також відомий як шість ступенів поділу. Основна ідея цієї характеристики полягає в тому, що короткий ланцюжок знайомств може зв’язати незнайомців.

• Мережа – це розподілена або безмасштабна мережа степеневого ступеня, коли невелика частина вузлів має багато з'єднань, а велика кількість вузлів мало.

Соціальна ізоляція концептуалізується кількома різними способами, наприклад небажанням брати участь у соціальних подіях, нездатністю встановити зоровий контакт з іншими , відсутність соціальної взаємодії між малими групами (сім'єю та друзями) або великою групою людей у суспільстві .

Крім того, існують деякі поняття, подібні до соціальної ізоляції, які можуть взаємозамінно використовуватися, хоча вони мають різні значення, наприклад, самотність, соціальна тривога, інтроверсія, соціальна ангедонія та соціальне відчуження.

• Самотність – це неприємне суб’єктивне відчуття, коли відбувається якісне чи кількісне зниження у стосунках особи з іншими. Можна сказати, що самотність поширена серед дорослих і пов’язана з поганим психічним і фізичним самопочуттям і нездоровим способом життя.

• Соціальна тривожність – це розлад, який здебільшого з’являється в період статевого дозрівання і зазвичай виникає, коли людина нервує і відчуває себе незручно в різноманітних соціальних ситуаціях, таких як співбесіда з IT-рекрутером, публічні виступи або спілкування з незнайомими людьми .

• Інтроверсія визначається як особистісна риса, що індивід воліє витрачати більше часу та енергії на свій внутрішній світ замість спілкування з іншими, що негативно вплине на соціалізацію.

• Соціальна ангедонія включає неприємні дії або думки і відома як зниження бажання шукати та брати участь у соціальних подіях, якими ми могли насолоджуватися раніше.

• Соціальна ізоляція – передбачає нехтування соціальними групами, такими як сім'я та друзі. Іншими словами, соціальна ізоляція – це неможливість створити

єдину групу членів суспільства . Як зазначалося вище, існують різні визначення соціальної ізоляції, засновані на різних точках зору; однак у цьому дослідженні ми визначаємо соціальну ізоляцію як міру, що відображає відсутність взаємодії між індивідом і спільнотою користувачів соціальних мереж.

1 ОГЛЯД ІСНУЮЧИХ ПРОГРАМНИХ РІШЕНЬ ТА СЕРВІСІВ В ПРЕДМЕТНІЙ ОБЛАСТІ

1.1 Аналіз соціальних мереж

Аналіз соціальних мереж (SNA) можна визначити як практику дослідження взаємодій між акторами соціальної мережі з метою вивчення її характеристик та поведінки. Можна сказати, що структура і склад є двома змінними в СНР. У багатьох реальних сценаріях структурні змінні важливіші за композиції, оскільки акцент зосереджений на різних зв'язках між соціальними акторами, таких як довіра, дружба тощо. Тим часом змінні композиції в основному стосуються атрибутів актора. Використовуючи SNA, дослідники можуть вивчати, як люди встановлюють зв'язки у більш широку соціальну структуру. Більше того, SNA можна використовувати, щоб показати, як люди намагаються створити відносини і як вони впливають один на одного.

Для рішення поставленої задачі в роботі:

- Виконано моделювання соціальної ізоляції в соціальних мережах за допомогою зважених орієнтованих соціальних графів
- Визначено пари математичних формул для вимірювання соціальної ізоляції в мережі
- Розроблено алгоритми на основі повідомлень для визначення ізольованих вузлів у зваженому орієнтованому графі

1.2 Мотивація дослідження

Зв'язок психічного та фізичного здоров'я беззаперечна, і від цього зв'язку залежить якість життя людей. Соціальна ізоляція може мати негативний вплив на фізичне, соціальне та психологічне благополуччя на різних рівнях.

Наприклад, на індивідуальному рівні розглядається психічне та фізичне здоров'я. Деякі з руйнівних наслідків для людей – це передчасна смерть, депресія, деменція та інвалідність внаслідок хронічного захворювання. На рівні громади вищий рівень злочинності, низький дохід, обмежена доступність громадського транспорту та неадекватні можливості для соціальної взаємодії є деякими прикладами . Крім того, соціальна ізоляція може призвести до тривалого періоду госпіталізації та неналежного зайняття лікарняних ліжок та медичних ресурсів.

Тим часом, незважаючи на важливість та актуальність цієї проблеми, існує дуже мало дослідницьких робіт у цій галузі, які використовують технології для вирішення проблем, пов'язаних із соціальною ізоляцією в громадах, і, наскільки нам відомо, не існує обчислювальних алгоритмів для вирішення проблем. цю проблему ефективно. Тому розробка ефективного та здійсненного обчислювального підходу для виявлення соціальної ізоляції у великій популяції може бути використана як система підтримки прийняття рішень, що веде до покращення як на індивідуальному, так і на громадському рівнях. У той же час це зменшить ймовірність супутніх розладів у людей, які страждають від ізоляції.

1.3 Постановка проблеми

Як зазначалося раніше, дослідження показали, що методи раннього превентивного втручання є більш вигідними та цінними з точки зору підходу до проблеми соціальної ізоляції [22]. Тому нашою основною метою є розробка обчислювальної моделі та набору алгоритмів із застосуванням методів аналізу соціальних мереж для виявлення ізольованих осіб у мережі.

У цьому дослідженні проблема пошуку соціально ізольованих вузлів визначається з двох різних точок зору на основі структури мережі та особливостей індивідів.

Структурна соціальна ізоляція

Припустимо, що спільнота представлена зваженим орієнтованим графом G(V, E,w). V — набір вузлів, що представляють наших індивідів у спільноті, а E — набір ребер між кожною з них парою, e = (vi,vj) ∈ E, де vi,vj ∈ V . Крім того, вагова функція w : E → R призначає кожному ребру реальне значення від 0 до 1. Вага використовується для вимірювання якості та міцності з’єднань між парами. Тому більша вага означає міцніший зв’язок. Тим часом, якщо вага ребра між vi і vj дорівнює x, вага ребра між vj і vi буде x0. Отже, значення x не обов’язково дорівнює значенню x0.

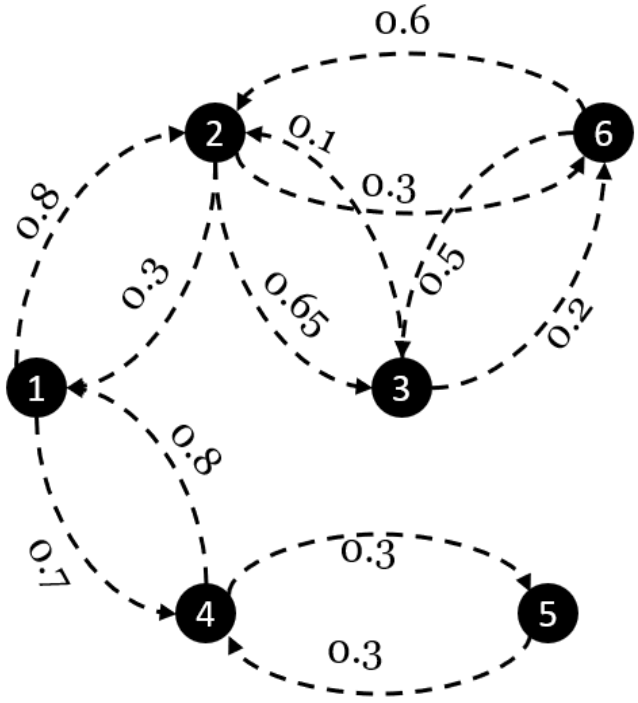


Рисунок 1.1 - Зважений орієнтований соціальний графік

Наприклад, на рисунку 1.1 ми маємо шість вузлів, що представляють шість осіб нашої спільноти. Вага зв'язку між вузлами один і два дорівнює 0,8, але вага зв'язку між вузлами два до одного дорівнює 0,3. Вага можна розрахувати на основі частоти з'єднань між парою вузлів, тривалості з'єднання, кількості спільних друзів, схожого членства в групах, рівня залежності, ступеня впливу один на одного або інших показників. Таким чином, ступінь вузла може означати, що якщо людина не має тенденції до взаємодії з іншими, ваги краю

можна використовувати для обчислення ступеня впливу та соціального схвалення вузла його сусідами.

Припустимо, що кількість активних з'єднань вузла (враховуючи ваги краю) менше, ніж на одне стандартне відхилення нижче середнього значення спільноти (віднімання середнього та стандартного відхилення для даної спільноти); у цьому випадку ми класифікуємо цей вузол як соціально ізольований вузол. Наприклад, якщо припустити, що середня кількість активних з'єднань у спільноті дорівнює 2 зі стандартним відхиленням 0,63, будь-який вузол з менш ніж 1,37 активних з'єднань можна класифікувати як ізольований вузол.

Наприклад, на рисунку 1.1 п'ятий вузол можна ідентифікувати як ізольований вузол.

Соціальна ізоляція на основі ознак

Цей алгоритм використовує той самий зважений орієнтований граф G(V, E, w) у попередньому алгоритмі, що представляє спільноту людини. Однак у цій частині кожен індивід, ai, має набір з n ознак, які можна визначити як Fai = {f1, f2, ..., fn}. Ці ознаки являють собою двійковий вектор фіксованого розміру, де значення клітинки дорівнює 1, якщо людина має відповідну функцію, і дорівнює 0 в іншому випадку. Наприклад, рис. 1.2 показує вектор ознак, якщо розглядати десять характеристик для кожної людини. У таблиці 1.1 наведено перелік різноманітних функцій. Цей набір функцій не є основними характеристиками. Іншими словами, дослідження показали, що деякі з наших характеристик або поведінки можуть посилити ізоляцію. Наприклад, припустимо, що перша та друга клітини цього вектора представляють ознаки віку старше 65 років і статус куріння відповідно до таблиці 1.1. Отже, цій людині не виповнилося 65 років і вона курить. Решта клітинок можна інтерпретувати аналогічно.

0	1	1	0	0	0	1	1	0	1
f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{1_0}

Рисунок 1.2 - Зразок двійкового вектора

Таблиця 1.1 - Двійкові значення для об’єктів

Особливості	Value	Features	Value
Понад 65	1	До 65	0
Курець	1	некурящий	0
Гендерна меншість	1	Трансгендер	0
Фізичні недоліки	1	Бути здоровим	0
Бідність	1	Багатство	0
Сільський	1	Урбанізувати	0
Жити самотньо	1	Не сам	0
Ожиріння	1	Не страждають ожирінням	0
Безробіття	1	Мають роботу	0
Неодружений	1	Не самотній	0

Зразок нашої гіпотетичної мережі, яка має як ваги, так і набори ознак показано на рисунку 1.3.

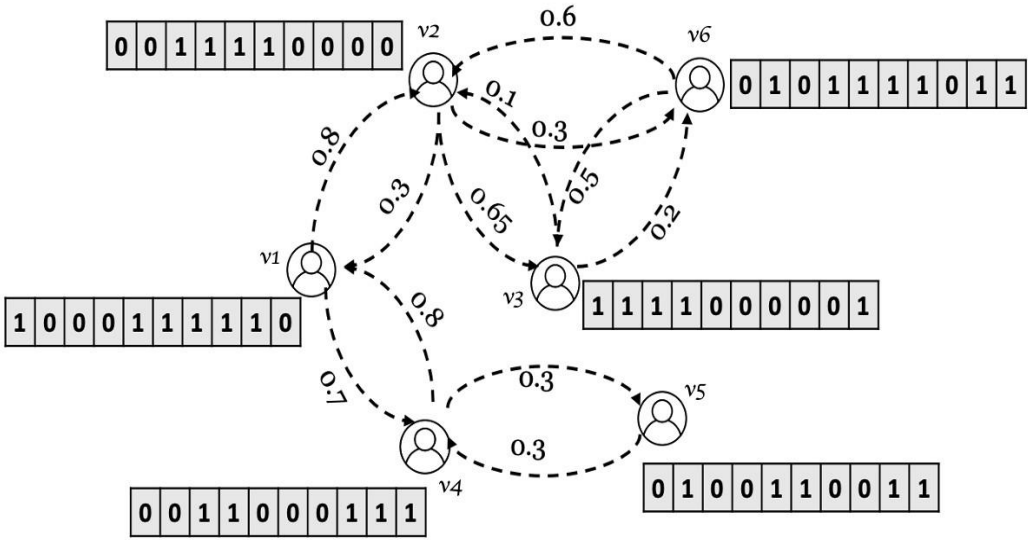


Рисунок 1.3: соціальний графік на основі пропозицій

Тому ми визначаємо задачу виявлення соціальної ізоляції як пошук набору вузлів у графі, який:

(a) не мають тенденції спілкуватися з іншими вузлами в графі; або

(b) не здатні ефективно спілкуватися, хоча їх може бути багато з'єднання; або

(c) не можуть встановити зв'язок з іншими через характеристики, які вони мають.

1.4 Заходи соціальної ізоляції

Через руйнівну дію соціальної ізоляції, за допомогою профілактичних методів, негативні наслідки ізоляції будуть зведені до нижчого рівня. На основі [22] існує шість типів втручань для боротьби із соціальною ізоляцією, таких як індивідуальні, групові, на основі надання послуг, технологічні, сусідські та структурні втручання.

Перший тип, індивідуальні втручання, — це зв’язок між літніми людьми та професіоналом чи волонтером, щоб допомогти їм. Зв’язок базується на спільних інтересах, а його позитивний вплив на старіючих дорослих доведено, хоча це економічно ефективний процес [40, 41].

У групових інтервенціях група людей зі спільними інтересами збирається разом, щоб провести соціальну, фізичну або освітню діяльність. Цей метод більш ефективний для осіб, які належать до груп меншин. Отже, вони можуть поділитися своїми проблемами з людьми тієї ж мови чи культури [40, 42].

Заходи щодо надання послуг пропонують довгострокову або середньострокову підтримку для людей похилого віку. Наприклад, місцеві служби намагалися залучити людей похилого віку, які страждають від проблем зі здоров’ям, до груп фізичних вправ . Згідно зі звітами, люди похилого віку відчували себе менш ізольованими після вправ. Інша послуга може включати студентів, які намагаються розвивати технологічні навички старших. Тому

ізольовані люди похилого віку почувалися б менш самотніми та пригніченими, зв’язуючись зі своєю сім’єю та друзями .

Втручання в контексті технологій спонукатимуть людей похилого віку використовувати більше сайтів соціальних мереж і смартфонів, щоб мати можливість покращити свої соціальні зв’язки через Інтернет .

Через обмеження ходіння у літніх людей втручання сусідів є менш популярним, ніж інші підходи. Однак, забезпечуючи безпечніші місця для сидіння або легкий доступ до громадського транспорту, для літніх людей буде створено більше соціальних можливостей .

Вищий рівень профілактики, такий як структурні втручання, спрямований на заохочення літніх людей до більшої взаємодії у своєму суспільстві. Наприклад, одним із випадків може бути працевлаштування старших на попередній роботі наставником або нова робота з можливістю навчання для всіх працівників із гнучким графіком роботи. Крім того, завдяки останнім досягненням у машинному навчанні та робототехніці, роботи здатні моделювати більшість діяльності людей.

Автори зосереджені на тому, як зробити життя старого населення більш керованим за допомогою самокерованих роботів. Хоча роботи, що займаються самообслуговуванням, зроблять літніх людей більш ізольованими через менше спілкування з іншими, роботи допомагають їм виконувати щоденні завдання без потреби в опікунах або членах сім’ї.

1.5 Виявлення соціальної ізоляції на основі опитувань

У роботі [7] автори зібрали дані дослідження охорони здоров’я та пенсійного забезпечення. Суб’єктивна соціальна ізоляція (самотність) вимірювалася за допомогою анкети з трьох пунктів, запровадженої HRS, під назвою «Психосоціальна анкета та опитувальник стилю життя». Для виявлення об’єктивної соціальної ізоляції вони використовували опитування, орієнтовані на соціальні взаємодії. Вони визначили ізольованих людей з двома порогами,

середнім і стандартним відхиленням. Для оцінки соціальної згуртованості менше ніж на одне стандартне відхилення нижче середнього (віднімання середнього та стандартного відхилення) ці особи були ідентифіковані як ізольовані, особи з балом у межах віднімання середнього та стандартного відхилення визначені як не ізольовані, а люди з оцінкою вище, ніж результат віднімання, визначений як добре пов'язаний.

У роботі автори використали алгоритм ройового інтелекту під назвою Firework Algorithm (FA), щоб пояснити концепцію соціальної ізоляції. FA — це алгоритм, який досліджує рішення, вибираючи випадкові точки, обмежені метрикою відстані, щоб знайти багатообіцяючі результати, які дають сконцентрований пошук поблизу. Ця робота зосереджена на описі поведінки соціально ізольованих осіб, які часто стикаються з відсутністю рівних можливостей та адекватних соціальних програм. У роботі автори запропонували новий і точний підхід обробки природної мови (NLP) для ідентифікації ізольованих осіб. Вони використовували дані пацієнтів з раком передміхурової залози, зібрані сховищем даних досліджень Медичного університету Південної Кароліни (MUSC). Автори використовували інформацію деяких пацієнтів, щоб виявити ізольованих пацієнтів, наприклад, їхні фізичні проблеми, історію хвороби, виписки та променеву онкологію.

Автори зібрали дані онлайн-опитування за участю понад 200 учасників у віці від 18 до 56 років. Автори використовували два показники, екстраверсію та шкали добробуту, щоб виявити ізольованих осіб.

У роботі автори запропонували нову модель прогнозування, що включає чотири кроки, засновану на спілкуванні людей похилого віку та їхньої мобільності. Автори заповнили анкету для 144 літніх людей у віці від 60 до 89 років. Це дослідження було проведено на основі фіксації соціальних взаємодій літніх людей протягом місяця. Після першого кроку, збору даних, було застосовано вибір атрибутів, щоб видалити зайві та некорисні атрибути. Наступним кроком була класифікація для влаштування пенсіонерів за рівнем

їхньої соціальної ізоляції. Останнім кроком було збалансувати отриманий набір даних для отримання точного прогнозу з класифікованими категоріями. Результатом цього дослідження стали дві програми: додаток для дорослих і додаток для опікунів, які хочуть добровільно підтримувати ізольованих літніх людей.

1.6 Виявлення викидів на основі графіка

Концепція виявлення викидів або аномалій у графіках тісно пов'язана з цим дослідженням. У цьому дослідженні основна увага зосереджена на виявленні соціально ізольованих людей у мережі, які можна вважати винятковими. У роботі [5] автори пропонують новий метод виявлення зовнішніх ребер на основі віднімання двох наборів ребер. Іншими словами, кількість фактичних і існуючих ребер порівнюється з кількістю очікуваних ребер у графі. Кількість очікуваних ребер розраховується на основі двох алгоритмів генерації випадкових графів, Erdos-Reyni та Preferential attachment. Віднімання двох наборів ребер у цій моделі називається автентичною оцінкою.

У роботі автори запропонували нову модель для виявлення аномалій у великомасштабних мережах. Автори використали поняття спільних сусідів. Вузли будуть призначатися до кластеру на основі того, як сусіди поділяють один з одним. Вузли без кластерів будуть класифікуватися як викиди або концентратори. Якщо ізольований вузол має два або більше ребер, це буде концентратор; в іншому випадку він вважається викидом.

Автори в [54] представили нову структуру даних під назвою DTAR, що означає пряму доступність через два стрибка, щоб оптимізувати алгоритм SCAN для його застосування у великомасштабних мережах. Автори стверджують, що SCAN++ має ті самі результати у виявленні викидів, концентраторів і кластерів із набагато меншим витратою часу та складністю.

На завершення ми розглянули деякі з попередніх робіт, пов'язаних із соціальною ізоляцією. Основна увага деяких досліджень приділяється

профілактичному методу для зменшення страждань ізоляції серед членів суспільства. Інші дослідження базуються на онлайнових або особистих анкетах. Проте майже всі ці роботи виконуються вручну на основі анкет та опитувань без використання будь-яких технологічних підходів. Тому, наскільки нам відомо, не існує жодної ефективної існуючої обчислювальної моделі та алгоритму для вирішення проблеми виявлення соціальної ізоляції.

1.7 Базові поняття для побудови соціальної мережі

Соціальна мережа — це будь яка соціальна взаємодія, що може бути представлена множиною соціальних одиниць та відношенням між ними. В останнє десятиліття у житі суспільства все більшу роль відіграють онлайнові соціальні мережі, в яких соціальна взаємодія здійснюється через веб-технології. Онлайновими соціальними мережами охоплено понад 70 % користувачів Інтернету. В Європі та США один користувач входить в середньому у 2 соціальні мережі. Деякі люди звертаються до соціальної мережі лише епізодично, а деякі – практично «живуть» в них, що інколи межує із патологією. Для когось соціальна мережа — це відпочинок, але дехто не приймає рішень без обговорення в мережі. Достовірно встановлено, що онлайнові соціальні мережі сьогодні реально впливають на поведінку їх учасників поза мережею. Віднедавна онлайнові соціальні мережі перетворилися в елемент прихованого управління соціально-економічними системами. Масштаб такого управління варіюється від просування товарів та брендів за технологією вірусного маркетингу до організації акцій масового суспільного протесту в Тунісі, Єгипті, Великобританії, Україні, Гонконзі тощо. Переваги такого управління полягають в оперативності досягнення мети, невеликих фінансових витратах, в можливості дистанційного маніпулювання свідомістю користувачів в автоматизованому режимі та майже повним протоколюванням результатів

діяльності. Мати у своїх руках такі важелі управління зацікавлені як приватні корпорації, так і політичні сили та органи державної влади. Щоб успішно управляти соціумом, спочатку треба його знати. Тому актуальною є задача аналізу соціальних мереж з метою ідентифікації багатофакторних залежностей для прогнозування поширення і сприйняття нової інформації учасниками мережі, з подальшим виявленням біфуркаційних процесів як всередині мережі, так і зовні. Розв'язання цієї задачі створить передумови для формалізованого управління поведінкою соціально-економічних систем шляхом впливу на характеристики окремих ключових елементів соціальних мереж.

Актор.

Актор це одна персона, корпоративна або колективна соціальна одиниця. Прикладами акторів можуть бути люди у групі, департаменти в корпорації, державні установи у місті, держави у світі. Якщо всі актори одного типу, наприклад, люди у групі, тоді вони утворюють однодомну мережу. Дводомною називається мережа з двома типами акторів, наприклад, учні та вчителі у школі. Зв'язок. Актори пов'язані один з одним соціальними зв'язками, типів яких може бути велика кількість. Кожен зв'язок визначає відношення певного типу між двома акторами, наприклад:

- оцінка однієї людини іншою (прояв дружби, бажання спілкування тощо);
- передача матеріальних ресурсів (бізнес-транзакції, кредитування чи запозичення тощо);
- спільна належність до певного явища, об'єкта чи події (участь у соціальній події, належність до одного соціального об'єднання);
- поведінкова взаємодія (спільні розмови, листування);
- переміщення між місцями чи статусами (міграція, соціальна мобільність);
- формальні відношення, наприклад, субординація;
- біологічні відношення, наприклад, батьки — діти.

Діада (двійка).

В найпростішому випадку відношення є бінарним. Воно описує зв’язок між парою акторів. Сукупність двох акторів та зв’язку між ними отримала назву діада. Аналіз діад направлений на вивчення властивостей бінарних відношень, як то визначення його симетричності чи закономірностей встановлення певних типів зв’язків.

Тріада (трійка).

Підмножина з трьох акторів та всі можливі зв’язки між ними утворюють ще одну одиницю аналізу — тріаду. На основі тріад розглядаються поняття транзитивності і транзитивного балансу. Також на основі транзитивного балансу здійснюється бінарна кластеризація мережі.

Підгрупа.

Діади та тріади утворюють найпростіші підгрупи. Підгрупа — це будь яка підмножина акторів та всі зв’язки між ними.

Група.

Аналіз соціальних мереж може розглядати на тільки діади, тріади та підгрупи, а й цілі системи акторів. Група складається зі всіх акторів та зв’язків між ними, які були отримані з концептуальних, теоретичних чи емпіричних міркувань, і розглядається як скінченна множина соціальних одиниць, які підлягають аналізу.

Соціальна мережа.

Соціальна мережа складається зі скінченої множини акторів і відношень, визначених ними.

Структурні та композиційні змінні.

Структурні змінні описують зв’язки певного типу між парами акторів. Наприклад, структурною змінною може бути міра бізнес-транзакцій між корпораціями, дружби між людьми, торгівлі між країнами. Композиційні змінні описують атрибути акторів, наприклад, стать, вік, етнічне походження, географічне розташування актора тощо.

Мережі належності.

Мережа що містить акторів одного типу називаються однодольними. Інший тип мереж виникає під час збору структурних даних з двох (або навіть більше) наборів сутностей. Наприклад розгляд двох наборів акторів, один з яких складається з корпорацій, а інший — з неприбуткових організацій. Можна розглянути потік фінансової допомоги з корпорацій до неприбуткових організацій. Така мережа є дводольною; вона містить два набори акторів, один з яких є донором, а інший — реципієнтом. Окремим типом дводольних мереж є мережа належності, яку утворюють множина акторів та множина подій (клуби, організації, громадські організації), до якої може відноситись актор або підгрупа акторів.

Соціальна позиція — це мітка акторів, що мають еквівалентні зв'язки з іншими членами групи. Соціальна роль — це система зв'язків між акторами чи соціальними позиціями.

В загальному випадку вершини графа відповідають акторам соціальної мережі, а ребра графа — комунікаціям між акторами (рис. 1.4).

X	n_1	n_2	n_3	n_4	n_5	n_6
n_1	0	1	1	1	0	0
n_2	1	0	1	0	0	0
n_3	1	1	0	1	0	0
n_4	1	0	1	0	0	0
n_5	0	0	0	0	0	1
n_6	0	0	0	0	1	0

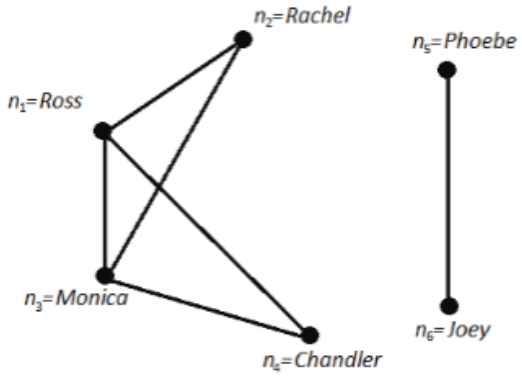


Рисунок 1.4 - Граф соціальної мережі

В аналізі соціальних мереж використовуються такі характеристики графів. Ступінь вершини ненаправленого графа, який дорівнює кількості ребер, що їй інциденті. Для направленого графа ступінь вершини розраховують окремо для вхідних та вихідних дуг. Ступінь вершини характеризує активність актора.

Інколи престиж актора визначають за принципами PageRank-алгоритму з пошукової системи Google.

Однією з основних цілей застосування теорії графів є кластеризація соціальної мережі, тобто формалізоване виявлення в ній груп. Група — це фрагмент графа, щільність внутрішніх зв’язків в якому домінує над щільністю зовнішніх зв’язків. Малозв’язні групи, що мають розріджені зв’язки всередині, але водночас об’єднують декілька підгруп — є дуже критичними.

Інформація в щільних групах, зазвичай, є схожою та надлишковою, тому для групи вона має меншу цінність, ніж та що проходить через слабкі зв’язки з інших підгруп.

2 АЛГОРИТМІЧНІ ТА МАТЕМАТИЧНІ РІШЕННЯ ДЛЯ ЗАДАЧ МОДЕЛЮВАННЯ СОЦІАЛЬНОЇ МЕРЕЖІ

2.1 Пропонована схема реалізації алгоритму

Як пояснювалося раніше, наша модель складається з двох алгоритмів, структурного визначення соціальної ізоляції, який має справу з кількістю активних з'єднань кожного вузла та якістю зв'язків між парою вузлів. З іншого боку, визначення соціальної ізоляції на основі функцій зосереджується на наборі функцій окремих людей. У цьому розділі ми детально опишемо компоненти цих алгоритмів.

Суть запропонованої нами моделі полягає у визначенні кількості активних зв'язків кожного індивіда в мережі та ступеня їх впливу на коло друзів. Для вирішення цього питання ми пропонуємо модель з використанням концепції поширення інформації в соціальних мережах. Фактично, у нашій моделі всім особам мережі надсилається певна кількість повідомлень, і їх просять повторно надіслати отримані повідомлення своїм безпосереднім сусідам. Потім записується кількість друзів кожного, хто отримав ці повідомлення. Крім того, фіксується кількість спроб повторної відправки повідомлень. Таким чином ми мали б уявлення про розмір активного кола друзів для кожної людини в мережі. Крім того, відстежуючи кількість повторно надісланих повідомлень, ми отримаємо інформацію щодо а 16 вплив і рівень прийняття людини.

У цьому алгоритмі ми представляємо спільноту, зіставляючи її на зважений орієнтований граф, як ми пояснили раніше. Метод складається з трьох основних етапів:

1. Певна кількість повідомлень надсилається кожному вузлу в графі, і ці вузли будуть передавати ці повідомлення своїм безпосереднім сусідам.

2. Для визначення кількості активних друзів буде підраховано кількість повідомлень, отриманих безпосередніми сусідами. Активний друг – це друг, який отримує принаймні повідомлення.

3. Відстежується кількість спроб, які одержувачі зроблять, щоб надіслати отримані повідомлення колу своїх друзів. На цьому кроці ми визначаємо вплив кожного вузла на його коло друзів.

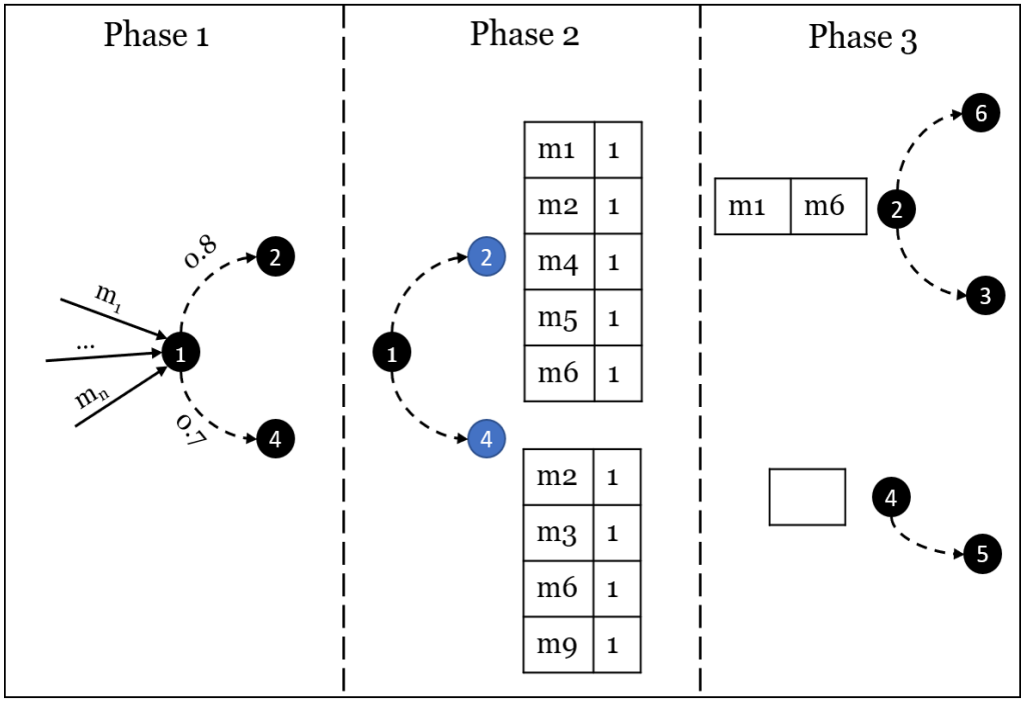


Рисунок 2.1 - Надсилання m повідомлень до вузла №1

Припустимо, що m=10. Наприклад, як показано на рисунку, у нас є мережа, що включає 6 вузлів. Ми припускаємо, що кількість повідомлень дорівнює 10, і наш фокус зосереджується на вузлі 1. Залежно від якості зв'язку пар, сусіди вузла отримають частину з 10 повідомлень. Вузол 1 має двох безпосередніх сусідів, вузли 2 і 4; Він надсилає 10 повідомлень цим сусідам і просить їх повторно надіслати ту частину повідомлень, яку вони отримали, своїм сусідам. На другому етапі, як видно, вузол 2 отримав п'ять, а вузол 4 отримав чотири повідомлення. Ми використовуємо двовимірний масив для запису отриманих повідомлень та ідентифікаторів їх вузлів відправника. На третьому етапі один із сусідів, вузол 2, надсилає лише два з отриманих повідомлень своїм сусідам, а вузол 4 не надсилає жодних повідомлень на всі.

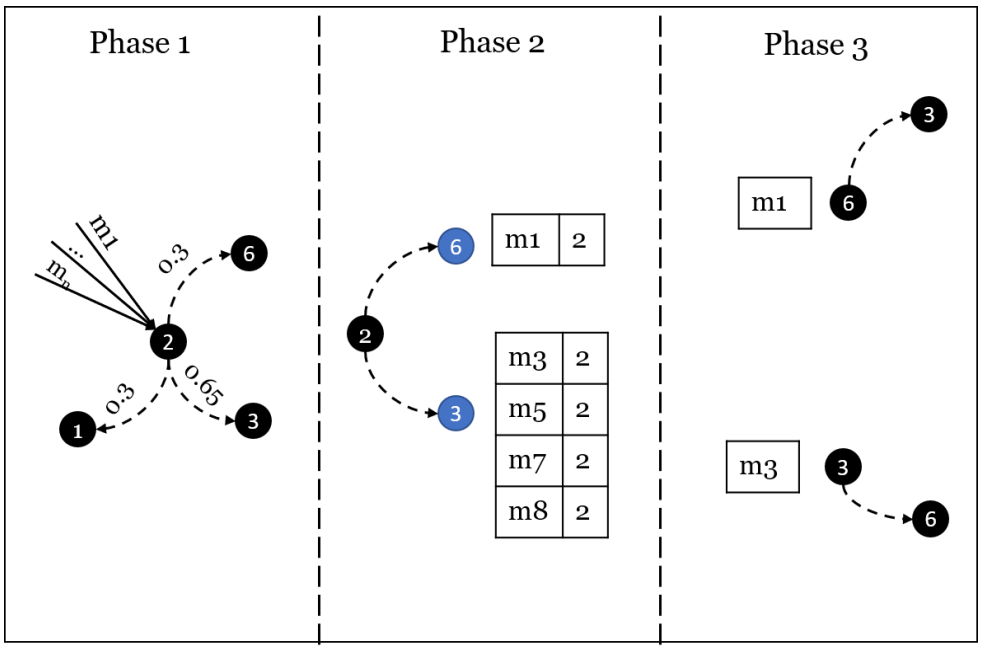


Рисунок 2.2 - Надсилання m повідомлень до вузла №2

Припустимо, що m=10. Аналогічно, на рисунку фокус зосереджений на вузлі 2, тому ми повторюємо процес. Цей вузол має трьох сусідів, але лише два з них, вузли 3 і 6, отримали частину повідомлень. Вузол 6, повторно надіслати єдине повідомлення, яке він отримав від вузла 2, своїм сусідам. З іншого боку, вузол № 3 повторно надсилає лише одне з чотирьох отриманих повідомлень своєму колу друзів. Цей процес повторюється для всіх вузлів мережі. Зрештою, на індивідуальному рівні ми можемо визначити кількість активних друзів (ті друзі, які принаймні отримали повідомлення) і вплив вузла на його коло друзів.

На рівні населення ми можемо визначити середню кількість активних друзів і соціальний вплив у цьому суспільстві, що можна інтерпретувати як соціальні норми спільноти. Отже, якщо вузол має менше ніж на одне стандартне відхилення нижче середнього значення для спільноти, ми позначаємо його як соціально ізольований вузол.

Відповідно, нехай c index(vi) позначає індекс для вимірювання передбачуваної кількості отриманих повідомлень сусідами вузла vi ∈ V (фаза 2). отже,

$$c_index(v_i) = m * \sum_{v_j \in N_G[v_i]} w(v_i, v_j).w(v_j, v_i \quad) \qquad (2.1)$$

де NG[vi] — набір сусідів вузла vi, а w(vi,vj) повертає вагу краю між вузлами vi та vj, а m — кількість повідомлень. Насправді, наявність більшої кількості зв'язків і міцніших зв'язків позитивно впливає на цей індекс. Тим часом на рівні населення середнє значення індексу с можна розрахувати як:

$$avg_c_index = \frac{\sum_{v_i \in V} c_index(v_i)}{|V|} \qquad (2.2)$$

З іншого боку, нехай index(vi) позначає індекс для вимірювання приблизної кількості повторень надсилання повідомлення сусідами вузла vi (фаза 3). Тамвперед,

$$a_index(v_i) = \frac{\sum_{v_j \in N'_G[v_i]} a_inx(v_i, v_j)}{|N'_G[v_i]|} \qquad (2.3)$$

де NG0 [vi] – це набір сусідніх вузлів вузла vi, які вже отримали m0 кількість повідомлень у фазі 2. Між тим, inx(vi,vj) тут вимірює приблизну кількість повідомлень, які будуть повторно надіслані вузлом vj і визначається як:

$$a_inx(v_i, v_j) = m' * w(v_j, v_i) * (1 - \frac{\sum_{v_k \in N_G[v_j]} w(v_j, v_k)}{|N_G[v_j]|} \quad) \qquad (2.4)$$

Алгоритм 1 представляє псевдокод нашого підходу.

Алгоритм 1: SBSID

Вхідні дані: зважений орієнтований графік, G(V,E,w)

Вихід: список соціально ізольованих вузлів

1: ініціалізація;

2: для всіх vi ∈ V до

3: для j = 1 до m

4: надіслати повідомлення всім vj ∈ NG(vi);

5: підрахувати кількість отриманих повідомлень;

6: створити);

7: для всіх

8: підрахувати кількість спроб повторно надіслати повідомлення;

9: кінець для

10: кінець для

11: кінець для

12: для всіх vi ∈ V до

13: якщо (c index(vi) < avg c index − σ)||

(a index(vi) < avg a index − σ) тоді

14: ізольований список ← vi; 15: кінець якщо

16: Повернути ізольований список;

2.2 Пропонована модель соціальної мережі

Фундаментальною частиною цієї моделі є спадкові та екологічні характеристики індивіда. Деякі з цих особливостей – проживання в сільській місцевості , куріння та фінансові проблеми . Ця частина алгоритму зосереджена на зборі деяких із цих особливих функцій та визначенні ізольованих вузлів у даній мережі на їх основі. Кожен вузол має набір двійкового вектора фіксованого розміру з назвою набір ознак, і нам потрібен датчик як набір ознак здорового суспільства, щоб порівняти вектор ознак кожного вузла з ним і виявити відхилення.

Отже, ми визначаємо інший вектор, який називається вектором індексу, двійковий вектор з такою ж довжиною, як вектори ознак, унікальний для даної мережі, і показує набір нормальних ознак для здорової популяції. Щоб розпізнати різницю між вектором ознак та вектором індексу, ми обчислюємо

відстань Хеммінга, яка є однією з найкращих метрик для порівняння двох бінарних векторів. Отже, цей алгоритм, як і попередній, має три основні фази:

1. Створити вектор підсумовування, який є вектором у результаті підсумовування значень клітинок векторів ознак індивідів. Формування індексного вектора. Якщо значення кожної клітинки вектора підсумовування принаймні на одиницю більше, ніж розмір сукупності, поділений на два, значення комірки буде 1, інакше 0.

1. Обчислити відстань Хеммінга між кожною особою та індексним вектором.

Наприклад, ми розглядаємо Рисунок як нашу гіпотетичну мережу. Як бачимо, ми маємо шість вузлів з різними наборами характеристик. На першому кроці, на основі алгоритму, ми повинні обчислити вектор підсумовування, показаний на рисунку 2.4.

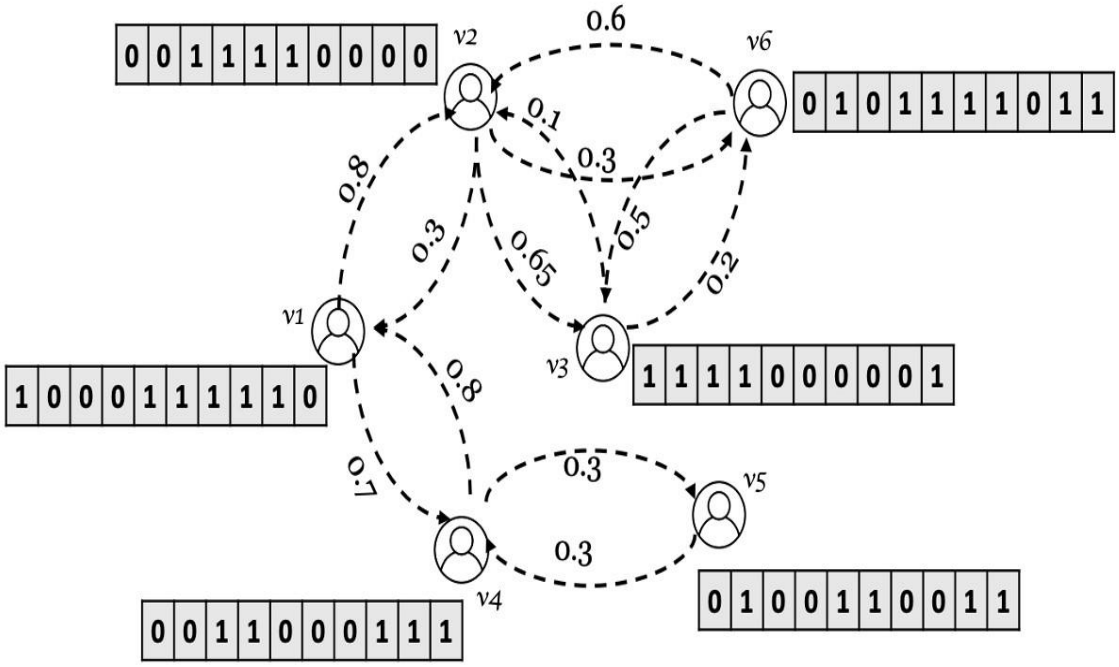


Рисунок 2.3 - Соціальний графік на основі пропозицій

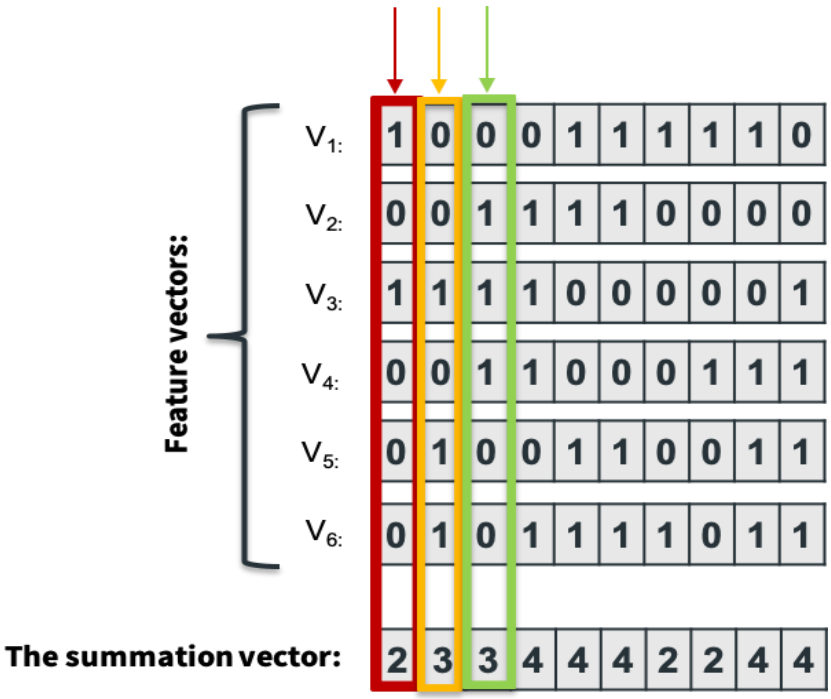


Рисунок 2.4 - Зразок обчислення вектора підсумовування

Після вектора підсумовування нам потрібно згенерувати вектор індексу. У цьому дослідженні, якщо більше половини населення має особливість, ми вважаємо це нормальною характеристикою. Ми обчислюємо підсумовування одиниці та значення для половини розміру популяції через це визначення.

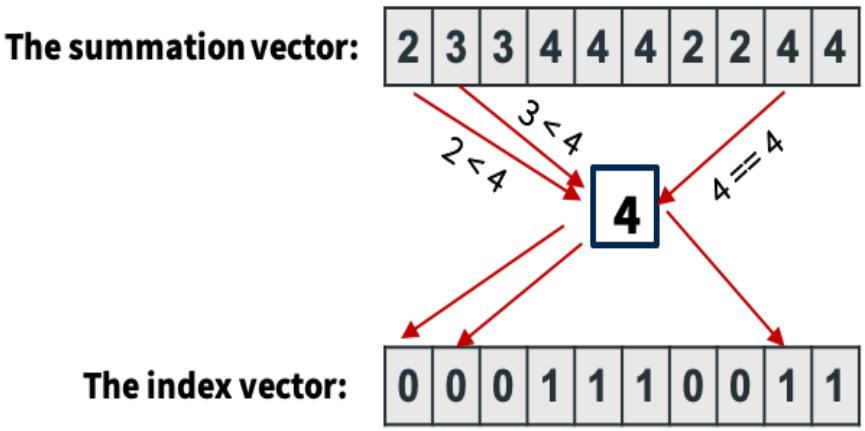


Рисунок 2.5 - Зразок генерування індексного вектора

Як показано на рисунку 2.5, якщо значення кожної клітинки у векторі підсумовування менше 4, відповідне значення у векторі індексу буде 0, інакше 1.

На наступному кроці, маючи вектори ознак та індексів, ми обчислюємо відстань Хеммінга для кожного вузла Рисунок 2.6. Якщо відстань Хеммінга кожного вузла перевищує суму середнього та стандартного відхилення, це ідентифікується як вузол з набором характеристик, відмінним від норми суспільства. Фактично, підсумовування середньої відстані Хеммінга та стандартного відхилення (5,33) вказує на максимальну допустиму відстань від норми суспільства. Тому ми можемо розглядати вузол 5 як вузол, відмінний від інших.

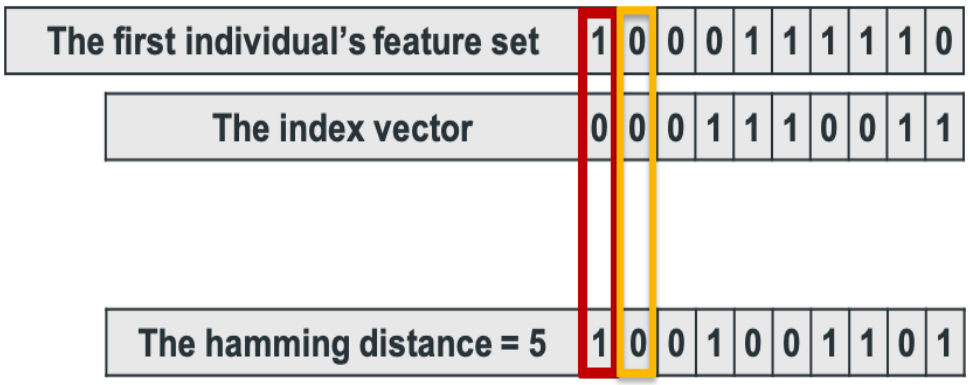


Рисунок 2.6 - Зразок розрахунку відстані Хеммінга

Hamming Distance	V_1	V_2	V_3	V_4	V_5	V_6
	5	3	2	2	6	4

Рисунок 2.7 - Зразок порівняння відстані Хеммінга

Середня дистанція Хеммінга = 3,666 ≈ 3,7

Стандартне відхилення відстані Хеммінга = 1,632 ≈ 1,63

Середнє + Стандартне значення = 5,33

Алгоритм 2 є псевдокодом алгоритму FBSID.

Для реалізації алгоритмів ми можемо розглянути три різні стратегії. У першій стратегії обидва алгоритми можуть виконуватися одночасно, але

незалежно. Як показано на рисунку 3.8, FBSID і SBSID запускаються одночасно на одному наборі даних, щоб виявити набір вузлів зі слабкими та низькоякісними зв’язками та ідентифікувати ті вузли з підозрілими характеристиками, які посилюють соціальну ізоляцію.

Алгоритм 2: FBSID

Вхідні дані: зважений орієнтований граф G(V,E,w) з набором бінарних характеристик, який називається вектором ознак

Вихід: список вузлів з характеристиками, відмінними від більшості населення

1: ініціалізація;

2: для всіх vi ∈ V до

3: Обчисліть вектор підсумовування

4: вектор підсумовування ← Σfi;

5: для всього елемента ∈ вектор підсумовування до

6: якщо елемент ≥ (половина населення + 1), то

7: індексний вектор ← 1;

8: ще

9: індексний вектор ← 0;

10: кінець якщо

11: кінець для

12: відстань Хеммінга ← Обчисліть відстань Хеммінга для кожного fi та індексного вектора;

13: avg = Обчисліть середнє значення для предметів на відстані Хеммінга;

14: std = Обчисліть стандартне відхилення для елементів на відстані Хеммінга;

15: норма = avg + std;

16: для всього пункту ∈ відстань Хеммінга до

17: якщо пункт > норма, то

18: ізольований список ← пункт;

19: ще

20: предмет – звичайна особа;

21: кінець якщо

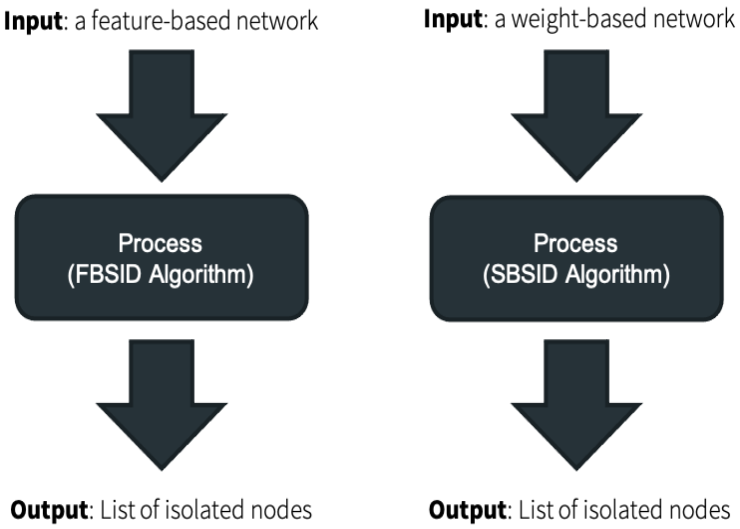


Рисунок 2.8 - Стратегія впровадження 1: обидва алгоритми працюють паралельно

У другій стратегії основна увага зосереджена на виявленні вузлів зі слабкими та низькоякісними зв'язками. На наступному кроці алгоритм FBSID перевіряє характеристики цих виявлених вузлів. Фактично, у цьому підході ми аналізуємо алгоритми на макрорівні, тобто розглядаємо їх зв'язок з іншими, а потім зосереджуємось на характеристиках кожного вузла, (рисунок 2.9.)

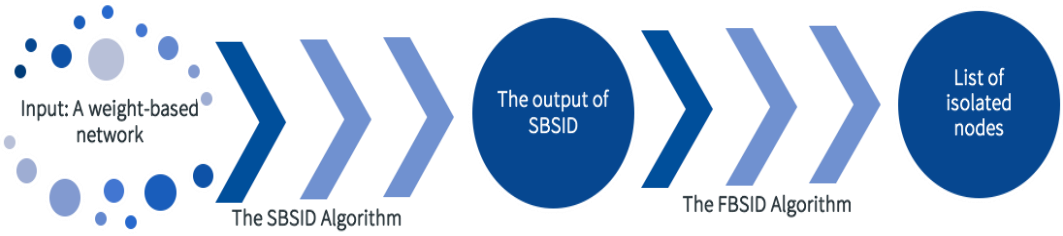


Рисунок 2.9 - Стратегія впровадження 2: послідовність алгоритму SBSID та алгоритму FBSID

Однак третя стратегія є повною протилежністю другої. Концентрація знаходиться на мікрорівні, що означає, що вона починається з виявлення вузлів з підозрілими атрибутами, які можуть спричинити ізоляцію, а потім вимірювання якості взаємодії вузлів з іншими (рисунок 2.10.)

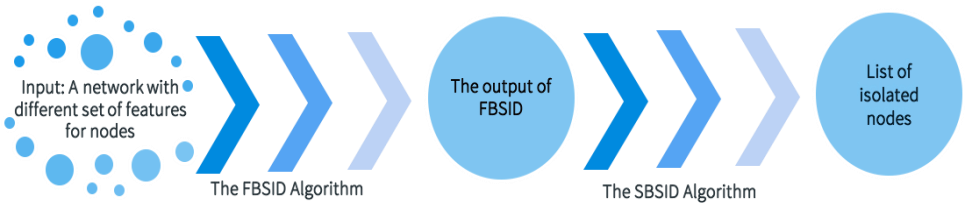


Рисунок 2.10 - Стратегія впровадження 3: послідовність алгоритму FBSID та алгоритму SBSID

3 АЛГОРИТМІЧНЕ ТА ПРОГРАМНЕ РІШЕННЯ ЗАДАЧІ МОДЕЛЮВАННЯ СОЦІАЛЬНОЇ АКТИВНОСТІ В МЕРЕЖІ

3.1 Складові програмного рішення

У цьому розділі ми оцінюємо продуктивність, ефективність і точність запропонованої нами моделі. Щоб оцінити наші алгоритми (SBSID і FBSID), ми створимо різні синтетичні мережі за допомогою тесту Ланчікінетті-Фортунато-Радіччі (LFR). Тест LFR створює мережі з реальними функціями соціальних мереж, такими як структури спільноти та розподіл ступенів за степенями. Ми також порівнюємо продуктивність нашої моделі з точним підходом.

3.2 Налаштування моделі

Ми створили чотири набори зважених орієнтованих соціальних графіків з 200, 500, 1000 і 10000 вузлів. Для кожного набору було створено три різні графіки різного рівня складності. Вагові коефіцієнти були згенеровані випадковим чином за допомогою еталону LFR і були нормалізовані в діапазоні від нуля до одиниці. Всього було згенеровано 12 мереж з такими параметрами, як зазначено в таблиці 4.1. Крім того, наступні параметри були використані для створення цих синтетичних мереж:

• k = 15, k являє собою середній ступінь для кожного вузла в графі.

• maxk = 50, встановлює максимальний розмір ступеня для кожного вузла в графі.

26

• μ = 0,2, 0,3, 0,4, μ – параметр змішування, який показує частку кількості ребер у кожній спільноті до їх загальної кількості.

• minc = 20, мінімальна кількість спільнот у графі.

• maxc = 50 – максимальна кількість спільнот у графі.

Таблиця 3.1 - Список синтетичних мереж, створених за допомогою тесту LFR

Етикетка	# Nodes	# Edges	μ	Максимальний ступінь
Мережа №1	200	5154	0.2	20
Мережа №2	200	5434	0.3	20
Мережа №3	200	5618	0.4	20
Мережа №4	500	13484	0.2	50
Мережа №5	500	13600	0.3	50
Мережа №6	500	14650	0.4	50
Мережа №7	1000	28800	0.2	100
Мережа №8	1000	26394	0.3	100
Мережа №9	1000	27892	0.4	100
Мережа №10	10000	312988	0.2	150
Мережа №11	10000	303700	0.3	150
Мережа №12	10000	311888	0.4	150

3.3 Експериментальні дослідження та імітаційне моделювання

Ми провели кілька експериментів з алгоритмами FBSID і SBSID, щоб виміряти час їх виконання та оцінити результати. Усі експерименти проводилися на MacBook Pro, macOS Cataline v10.15.6 з чотирьохядерним процесором Intel Core i7 2,2 ГГц та 16 ГБ оперативної пам’яті. Для розробки експериментальної моделі на фреймворку Pycharm була використана мова програмування Python.

Алгоритм виявлення соціальної ізоляції на основі функцій

Ми провели 12 незалежних експериментів, кожен з яких ми повторили 5 разів, щоб отримати середні значення результатів. Отримані результати для алгоритму FBSID наведено в таблиці. У цій таблиці представлено кількість ізольованих вузлів у кожному наборі даних та середній час роботи алгоритму для кожного графіка. Тим часом результати також проілюстровані на рисунках 3.1 та 3.2.

Таблиця 3.2 - Результати, отримані за допомогою алгоритму FBSID

Dataset	Тривалість (с)	# ізольовані вузли
200-0.2	0,025 ± 0,005	11
200-0.3	0,025 ± 0,005	12
200-0.4	0.025 ± 0.006	8
500-0.2	0.152 ± 0.083	21
500-0.3	0.157 ± 0.086	16
500-0.4	0.158 ± 0.090	24
1000-0.2	0.630 ± 0.339	37
1000-0.3	0.634 ± 0.339	48
1000-0.4	0.638 ± 0.342	44
10000-0.2	66.929 ± 33.733	500
10000-0.3	71.810 ± 31.971	503
10000-0.4	72.688 ± 37.897	506

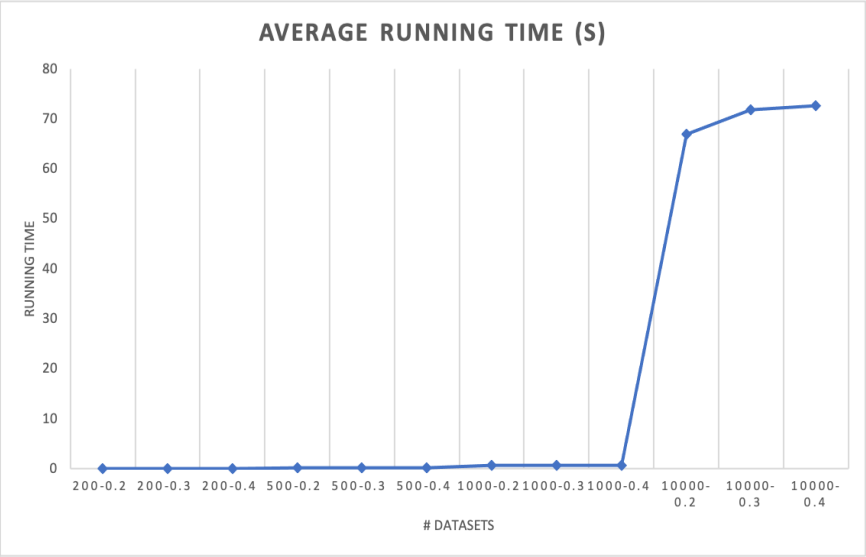


Рисунок 3.1 - Середній час роботи алгоритму FBSID

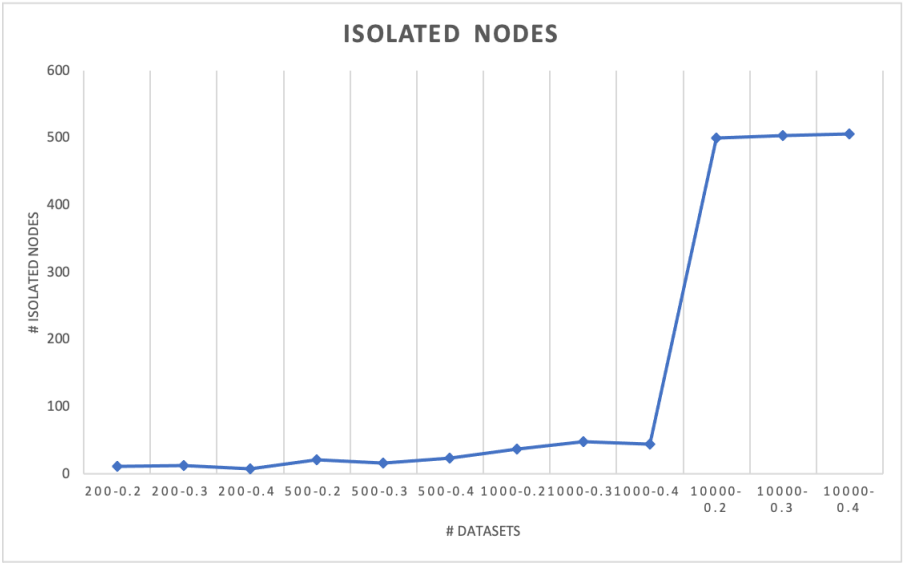


Рисунок 3.2 - Кількість ізольованих вузлів в алгоритмі FBSID

Виходячи з результатів, можна помітити, що результати роботи запропонованого нами алгоритму FBSID мають меншу залежність від складності, ніж кількість вузлів у мережі. Наприклад, як видно з таблиці 4.1, кількість ізольованих вузлів для мережі з 200 вузлами різної складності (μ = 0,2, 0,3, 0,4) знаходиться майже в одному діапазоні.

Однак, змінюючи розмір кожної мережі, кількість ізольованих вузлів збільшується. Насправді результати цього алгоритму мають пряме відношення до кількості вузлів у графі і не залежать від кількості ребер. Крім того, виходячи з рисунка 4.1, час роботи та розмір кожного графіка мають пряму залежність.

Отже, графіки більшого розміру потребують більше часу для обробки алгоритмом.

Виявлення соціальної ізоляції на основі структури

Ми віримо, що зі збільшенням кількості повідомлень результати будуть точнішими. Однак збільшення кількості повідомлень збільшить час роботи нашого алгоритму. Щоб знайти компроміс між кількістю повідомлень і необхідним часом обробки, ми провели серію експериментів, щоб порівняти час виконання та результати нашого алгоритму, змінивши кількість повідомлень зі

100 до 1000. Отже, ми маємо провели 48 незалежних експериментів, і кожен з них повторили 5 разів для отримання середніх значень результатів. Для проведення цих експериментів було використано 4 різні налаштування повідомлень (тобто 100, 250, 500 і 1000), щоб перевірити масштабованість та надійність алгоритму SBSID.

Отримані результати наведено в таблицях 3.3, 3.4, 3.5, 3.6. Ми можемо помітити, що в алгоритмі SBSID кількість ізольованих вузлів не залежить від складності мережі. Однак це має пряме відношення до розміру графіка. Крім того, збільшення кількості вузлів у графіках та кількості повідомлень в алгоритмі призводило до збільшення часу виконання кожного експерименту.

Таблиця 3.3 - Результати, отримані за алгоритмом SBSID - #Msg = 100

Message Number = 100		
Dataset	Running Time (s)	# Isolated Nodes
200-0.2	1.411 ± 0.102	55 ± 4
200-0.3	1.454 ± 0.133	54 ± 2
200-0.4	1.931 ± 0.086	56 ± 4
500-0.2	3.086 ± 0.270	133 ± 3
500-0.3	3.709 ± 0.254	127 ± 2
500-0.4	4.042 ± 0.194	139 ± 7
1000-0.2	9.379 ± 0.839	227 ± 10
1000-0.3	8.857 ± 0.283	258 ± 6
1000-0.4	10.052 ± 0.636	251 ± 20
10000-0.2	161.716 ± 9.845	2405 ± 33
10000-0.3	177.281 ± 9.976	2086 ± 189
10000-0.4	192.228 ± 4.482	1978 ± 24

Таблиця 3.4 - Результати, отримані за алгоритмом SBSID - #Msg = 250

Message Number = 250		
Dataset	Running Time (s)	# Isolated Nodes
200-0.2	3.395 ± 0.277	40 ± 3
200-0.3	3.768 ± 0.175	42 ± 3
200-0.4	4.978 ± 0.226	44 ± 2
500-0.2	7.385 ± 0.298	113 ± 4
500-0.3	9.657 ± 0.307	93 ± 3
500-0.4	11.439 ± 0.181	100 ± 2
1000-0.2	25.667 ± 0.319	182 ± 6
1000-0.3	22.996 ± 0.453	197 ± 5
1000-0.4	25.551 ± 0.607	198 ± 3
10000-0.2	414.253 ± 20.852	1930 ± 25
10000-0.3	392.439 ± 17.023	1966 ± 22
10000-0.4	403.995 ± 11.844	1956 ± 19

Таблиця 3.5 - Результати, отримані за алгоритмом SBSID - #Msg = 500

Message Number = 500		
Dataset	Running Time (s)	# Isolated Nodes
200-0.2	6.406 ± 0.352	39 ± 3
200-0.3	6.598 ± 0.213	40 ± 3
200-0.4	9.389 ± 0.461	41 ± 3
500-0.2	13.542 ± 0.451	115 ± 3
500-0.3	16.502 ± 0.454	91 ± 3
500-0.4	20.746 ± 0.825	101 ± 3

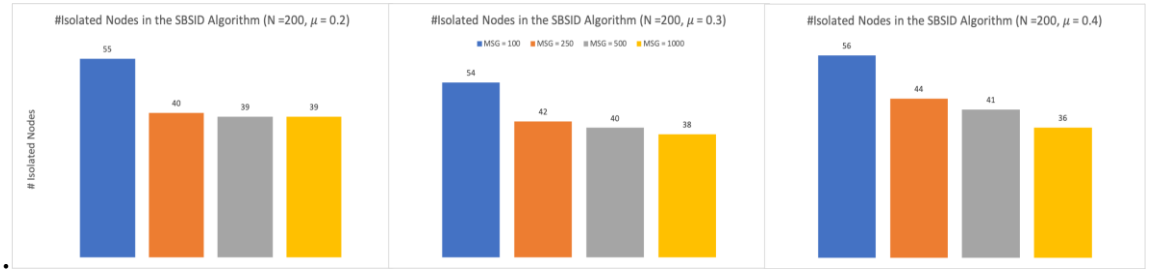
1000-0.2	47.161 ± 1.762	178 ± 4
1000-0.3	42.855 ± 1.333	193 ± 7
1000-0.4	45.756 ± 1.301	197 ± 5
10000-0.2	719.410 ± 59.896	1939 ± 15
10000-0.3	791.463 ± 14.623	1928 ± 17
10000-0.4	824.010 ± 50.555	1960 ± 18

Таблиця 3.6 - Результати, отримані за алгоритмом SBSID - #Msg = 1000

Message Number = 1000		
Dataset	Running Time (s)	# Isolated Nodes
200-0.2	14.182 ± 0.575	39 ± 1
200-0.3	14.918 ± 0.125	38 ± 1
200-0.4	21.275 ± 0.926	36 ± 1
500-0.2	31.145 ± 0.841	116 ± 3
500-0.3	41.808 ± 1.478	87 ± 3
500-0.4	47.415 ± 1.045	100 ± 2
1000-0.2	110.910 ± 6.765	180 ± 4
1000-0.3	89.421 ± 1.303	196 ± 5
1000-0.4	102.305 ± 5.437	194 ± 3
10000-0.2	1410.300 ± 83.49	1930 ± 8
10000-0.3	1253.374 ± 82.51	1916 ± 16
10000-0.4	1463.000 ± 75.71	1940 ± 16

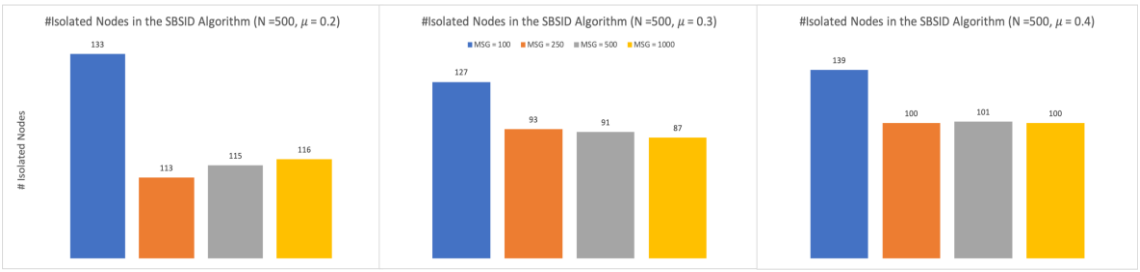
Як показано на рисунках, ми маємо менше ізольованих вузлів, надсилаючи більше повідомлень у мережі. Наприклад, на рисунку 3.3, коли N = 200 і μ = 0,2, після відправлення 100 повідомлень було виявлено 55 ізольованих вузлів, тоді як після збільшення кількості повідомлень до 1000, лише 39 вузлів було ідентифіковано як ізольовані. Відповідно до результатів, розмір повідомлення

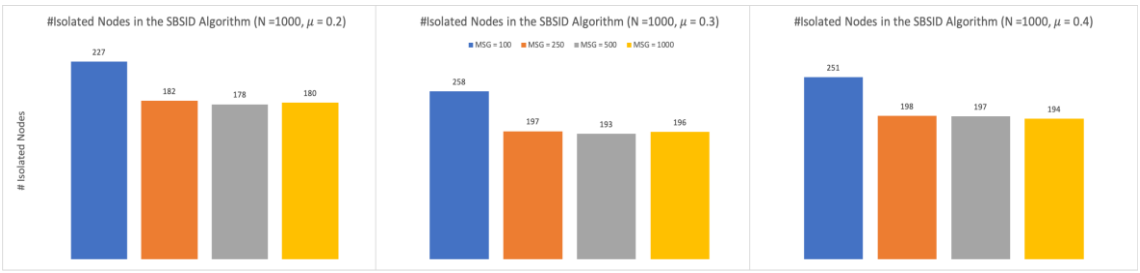
250 можна вибрати як компромісне значення, оскільки є незначні зміни в кількості ізольованих вузлів для номерів повідомлень.



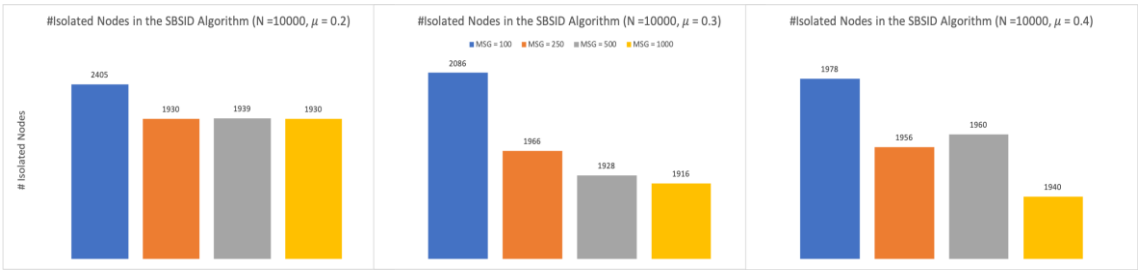
.

Кількість ізольованих вузлів у різних налаштуваннях повідомлень (вузлів = 200)





Кількість ізольованих вузлів у різних налаштуваннях повідомлень (вузлів = 1000)



Кількість ізольованих вузлів у різних налаштуваннях повідомлень (вузлів = 10000)

Рисунок 3.5 - Кількість ізольованих вузлів у різних налаштуваннях повідомлень

Аналіз подібності

Як згадувалося раніше, ми повторювали кожен експеримент 5 разів, щоб отримати середні значення для результату. Тому ми розглянули два індекси подібності, Жаккара і коефіцієнт перекриття, щоб виміряти подібність між наборами відповідей. Ми випадковим чином вибрали пари отриманих результатів і розрахували їх схожість на основі індексів подібності Жаккара та коефіцієнта перекриття. У таблицях наведено результати порівняння подібності між наборами відповідей для різних налаштувань повідомлень. Рисунки також ілюструють вимірювання подібності Жаккара та коефіцієнта перекриття.

Результати підтверджують, що за рахунок збільшення кількості повідомлень підвищиться достовірність результатів. Наприклад, коли кількість повідомлень дорівнює 100, подібність Жаккарда становить близько 50%, а коефіцієнт перекриття — близько 65%. Однак, коли номер повідомлення становить 1000, ці значення збільшуються приблизно до 80% і 90% відповідно. Загалом, як видно на рисунках, за рахунок збільшення кількості повідомлень подібність відповідей збільшилася.

S_1 = 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 16, 19, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 33, 34, 35, 36, 46, 48, 53, 54, 58, 59, 60, 64, 65, 67, 68, 69, 71, 73, 76, 81, 82, 83, 91, 95, 101, 109, 110, 118, 135, 151, 153, 166, 169, 178

S_2 = 2, 3, 4, 5, 7, 8, 11, 12, 16, 19, 20, 22, 23, 24, 25, 30, 31, 32, 34, 35, 36, 39, 40, 44, 46, 47, 48, 53, 56, 57, 58, 59, 60, 63, 64, 65, 67, 68, 69, 71, 72, 76, 77, 81, 82, 86, 91, 101, 102, 109, 118, 133, 139, 140, 147, 151, 155, 156, 169, 178

Mutual_Nodes = 3, 4, 5, 7, 8, 11, 16, 19, 22, 23, 24, 25, 30, 31, 34, 35, 36, 46, 48, 53, 58, 59, 60, 64, 65, 67, 68, 69, 71, 76, 81, 82, 91, 101, 109, 118, 151, 169, 178

Рисунок 3.7 - Зразок виводу з алгоритму SBSID – MSG = 100

Наприклад, як показано на рисунку, у мережі з 200 вузлами, μ = 0,2, після відправлення 100 повідомлень S1 і S2 являють собою два набори відповідей, які випадковим чином вибираються для порівняння один з одним розміром 56 і 60. Список взаємних вузлів показує 39 загальних вузлів між S1 і S2. Подібність Жаккара і коефіцієнта перекриття між цими двома наборами становить майже %50 і %70 відповідно.

У мережі з такою ж структурою (200 вузлів, μ = 0,2), як показано на рисунку 3.8, після відправки 250 повідомлень S3 з 44 вузлами і S4 з 40 вузлами є ізольованими вузлами, вибраними для порівняння один з одним випадковим чином. Список взаємних вузлів показує 36 загальних вузлів між S3 і S4. Відповідно, індекси подібності Жаккара та коефіцієнта перекриття для цих двох наборів знаходяться біля %70 та %80. Примітно, що індекси схожості збільшуються за рахунок надсилання 250 повідомлень замість 100 повідомлень.

S_3 = 2, 3, 4, 5, 7, 11, 12, 16, 19, 21, 23, 24, 25, 26, 30, 31, 32, 34, 35, 36, 40, 44, 46, 48, 53, 58, 59, 69, 76, 77, 80, 81, 82, 86, 91, 109, 133, 143, 153, 156, 169, 178, 197, 199

S_4 = 2, 3, 4, 5, 7, 11, 12, 16, 21, 23, 24, 25, 26, 30 ,31, 34, 35, 36, 37, 40, 44, 46, 53, 58, 59, 61, 68, 69, 76, 81, 82, 86, 102, 109 133, 143, 153, 169, 178, 199

Mutual_Nodes = 2, 3, 4, 5, 7, 11, 12, 16, 19, 23, 24, 25, 30, 32, 34, 35, 36, 44, 46, 47, 48, 53, 58, 59, 68, 69, 76, 81, 82, 86, 91, 109, 133, 151, 156, 169

Рисунок 3.8 - Зразок виводу з алгоритму SBSID – MSG = 250

Враховуючи ту саму мережу (N= 200, μ = 0,2), як показано на рисунку 4.9, після відправки 1000 повідомлень, S5 з 40 і S6 з 39 ізольованими вузлами вибираються для порівняння один з одним, випадковим чином. Список взаємних вузлів показує 36 загальних вузлів між S5 і S6. Отже, після збільшення кількості

повідомлень подібність Жаккарда та коефіцієнта перекриття також збільшується і досягає %85 і %90.

Отримані результати показують, що індекс коефіцієнта перекриття має вищі значення, ніж індекс Жакара. Подібність Жаккара є дуже надійним і потужним індексом, але він має суттєвий недолік, коли ми порівнюємо набори різних розмірів.

Це означає, що використання подібності коефіцієнта перекриття є більш доречним, ніж подібності. Індекс Жакара в цьому дослідженні.

S_5 = 2, 3, 4, 5, 7, 11, 12, 16, 19, 23, 24, 25, 26, 30, 31, 32, 34, 35, 36, 37, 40, 46, 51, 53, 58, 59, 61, 69, 76, 81, 82, 86, 91, 109, 133, 143, 153, 169, 178, 199

S_6 = 2, 3, 4, 5, 7, 11, 12, 16, 19, 21, 23, 24, 25, 26, 30, 31, 32, 34, 35, 36, 37, 40, 46, 53, 58, 59, 68, 69, 76, 80, 81, 82, 86, 91, 109, 133, 143, 153, 169

Mutual_Nodes Mutual_Nodes = 2, 3, 4, 5, 7, 11, 12, 16, 19, 23, 24, 25, 26, 30, 31, 32, 34, 35, 36, 37, 40, 46, 53, 58, 59, 69, 76, 81, 82, 86, 91, 109, 133, 143, 153, 169

Рисунок 3.9 - Зразок виводу з алгоритму SBSID – MSG = 1000

Таблиця 3.7 - Результати порівняння алгоритму SBSID (Msg = 100)

Номер повідомлення = 100				
Набір даних	Жаккард Макс	Жаккард Середній	Коефіцієнт перекриття Макс	Середній коефіцієнт перекриття
200-0.2	0.531	0.460 ± 0.049	0.693	0.628 ± 0.047
200-0.3	0.284	0.231 ± 0.046	0.442	0.373 ± 0.062
200-0.4	0.444	0.394 ± 0.032	0.615	0.565 ± 0.033
500-0.2	0.517	0.457 ± 0.038	0.682	0.626 ± 0.036

500-0.3	0.503	0.476 ± 0.017	0.669	0.645 ± 0.016
500-0.4	0.550	0.474 ± 0.043	0.709	0.642 ± 0.039
1000-0.2	0.534	0.485 ± 0.030	0.696	0.653 ± 0.027
1000-0.3	0.524	0.502 ± 0.013	0.688	0.668 ± 0.011
1000-0.4	0.510	0.477 ± 0.021	0.675	0.645 ± 0.019
10000-0.2	0.491	0.483 ± 0.005	0.658	0.652 ± 0.004
10000-0.3	0.532	0.524 ± 0.004	0.695	0.688 ± 0.003
10000-0.4	0.521	0.542 ± 0.004	0.685	0.674 ± 0.004

Крім того, на рисунку представлено середній час проведення експериментів на 12 наборах даних з різними розмірами повідомлень. Ми помітили, що час виконання кожного експерименту безпосередньо пов’язане з кількістю повідомлень у кожному експерименті. Це означає, що за рахунок збільшення кількості повідомлень (100, 250, 500, 1000) час відповіді на виявлення ізольованих вузлів також збільшується.

Таблиця 3.8 - Результати порівняння алгоритму SBSID (Msg = 250)

Message Number = 250				
Dataset	Jaccard Max	Jaccard Mean	Overlap Coefficient Max	Overlap Coefficient Mean
200-0.2	0.714	0.646 ± 0.060	0.833	0.783 ± 0.044
200-0.3	0.423	0.371 ± 0.035	0.594	0.540 ± 0.038
200-0.4	0.607	0.555 ± 0.062	0.756	0.712 ± 0.054
500-0.2	0.643	0.585 ± 0.033	0.783	0.737 ± 0.026
500-0.3	0.660	0.630 ± 0.027	0.795	0.772 ± 0.020
500-0.4	0.704	0.662 ± 0.039	0.8265	0.795 ± 0.028
1000-0.2	0.676	0.655 ± 0.019	0.807	0.791 ± 0.014
1000-0.3	0.659	0.642 ± 0.019	0.794	0.782 ± 0.014
1000-0.4	0.691	0.665 ± 0.018	0.817	0.799 ± 0.013

10000-0.2	0.652	0.643 ± 0.007	0.789	0.783 ± 0.005
10000-0.3	0.655	0.645 ± 0.005	0.791	0.784 ± 0.004
10000-0.4	0.645	0.637 ± 0.006	0.784	0.778 ± 0.004

Таблиця 3.9 - Результати порівняння алгоритму SBSID (Msg = 500)

Message Number = 500				
Dataset	Jaccard Max	Jaccard Mean	Overlap Coefficient Max	Overlap Coefficient Mean
200-0.2	0.800	0.716 ± 0.057	0.888	0.833 ± 0.039
200-0.3	0.520	0.462 ± 0.044	0.684	0.631 ± 0.041
200-0.4	0.809	0.678 ± 0.089	0.894	0.805 ± 0.063
500-0.2	0.750	0.707 ± 0.027	0.857	0.828 ± 0.019
500-0.3	0.775	0.713 ± 0.035	0.873	0.832 ± 0.023
500-0.4	0.729	0.688 ± 0.040	0.843	0.814 ± 0.029
1000-0.2	0.774	0.649 ± 0.019	0.872	0.787 ± 0.014
1000-0.3	0.778	0.699 ± 0.040	0.875	0.822 ± 0.027
1000-0.4	0.768	0.728 ± 0.023	0.869	0.842 ± 0.015
10000-0.2	0.739	0.009 ± 0.009	0.850	0.843 ± 0.006
10000-0.3	0.745	0.728 ± 0.011	0.854	0.842 ± 0.007
10000-0.4	0.731	0.726 ± 0.006	0.845	0.841 ± 0.004

Таблиця 3.10 - Результати порівняння алгоритму SBSID (Msg = 1000)

Message Number = 1000				
Dataset	Jaccard Max	Jaccard Mean	Overlap Coefficient Max	Overlap Coefficient Mean
200-0.2	0.853	0.801 ± 0.032	0.921	0.889 ± 0.019
200-0.3	0.714	0.624 ± 0.067	0.833	0.766 ± 0.051
200-0.4	0.837	0.747 ± 0.083	0.911	0.852 ± 0.055
500-0.2	0.819	0.765 ± 0.032	0.900	0.866 ± 0.020

500-0.3	0.782	0.748 ± 0.021	0.878	0.856 ± 0.014
500-0.4	0.811	0.762 ± 0.032	0.895	0.864 ± 0.020
1000-0.2	0.842	0.807 ± 0.025	0.914	0.893 ± 0.015
1000-0.3	0.816	0.777 ± 0.024	0.898	0.874 ± 0.015
1000-0.4	0.836	0.803 ± 0.019	0.911	0.891 ± 0.012
10000-0.2	0.805	0.797 ± 0.004	0.892	0.887 ± 0.003
10000-0.3	0.800	0.795 ± 0.005	0.889	0.885 ± 0.003
10000-0.4	0.811	0.796 ± 0.008	0.895	0.886 ± 0.005

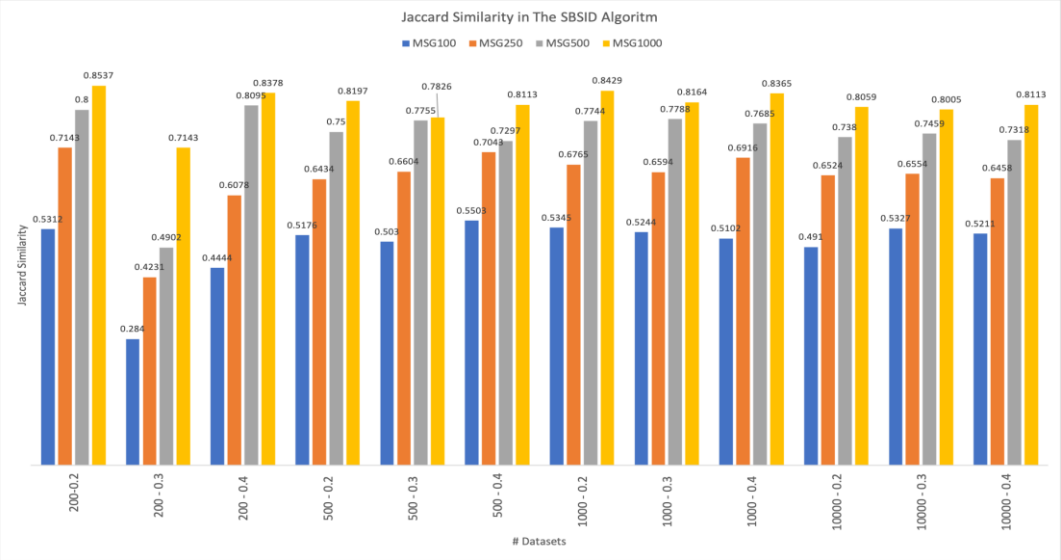


Рисунок 3.10 - Порівняння подібності Жаккара в алгоритмі SBSID

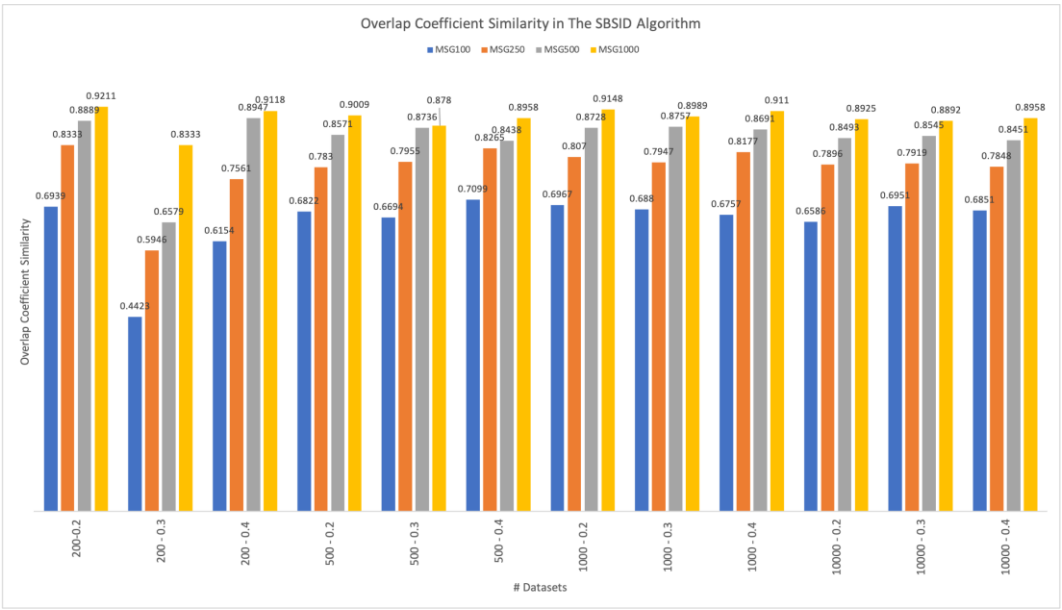


Рисунок 3.11 - Порівняння подібності коефіцієнтів перекриття в SBSID

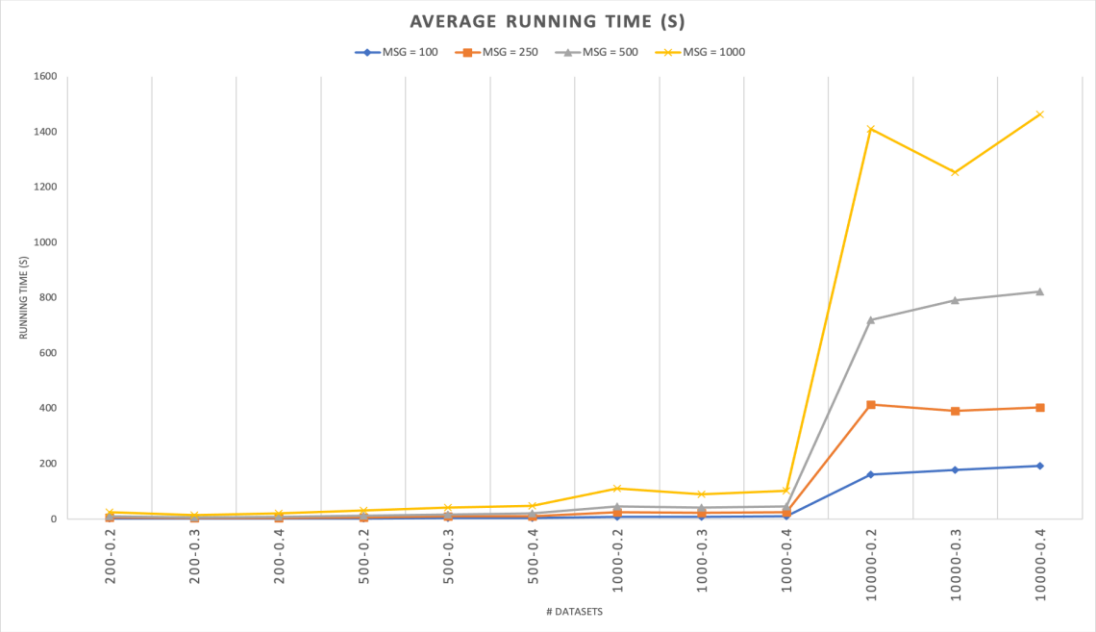


Рисунок 3.12 - Результати для середнього часу виконання на основі кількості повідомлень в алгоритмі SBSID

3.4 Порівняння результатів з точним методом

У цій частині ми реалізували інший алгоритм під назвою Точний алгоритм. Ми запускаємо алгоритм Exact на 48 незалежних синтетичних наборах даних і повторюємо кожен з них 10 разів, щоб порівняти досягнуті результати з результатами, отриманими від алгоритму SBSID. У алгоритмі Exact певна частина повідомлень (100, 250, 500, 1000) надсилається всім вузлам на графі без урахування ймовірності отримання повідомлень. Таким чином, усі існуючі вузли мережі отримають однакову кількість повідомлень. Таким чином, точний алгоритм виявляє всі потенційні соціально ізольовані вузли.

Отримані результати точного алгоритму проілюстровано в таблицях 3.11, 3.12, 3.13 і 3.14, які показують, що кількість ізольованих вузлів в алгоритмі Exact має прямий зв’язок з розміром мережі. Також примітно, що за рахунок збільшення розміру мереж збільшуються час виконання кожного експерименту та кількість виявлених ізольованих вузлів.

Таблиця 3.11 - Точний підсумковий результат алгоритму - #Msg = 100

Message Number = 100		
Dataset	Running Time (s)	# Isolated Nodes
200-0.2	0.038 ± 0.003	47
200-0.3	0.040 ± 0.004	44
200-0.4	0.043 ± 0.009	39
500-0.2	0.110 ± 0.019	97
500-0.3	0.112 ± 0.017	113
500-0.4	0.121 ± 0.007	117
1000-0.2	0.275 ± 0.021	220
1000-0.3	0.238 ± 0.021	226
1000-0.4	0.26 ± 0.012	236
10000-0.2	3.769 ± 0.172	2418
10000-0.3	3.61 ± 0.239	2384
10000-0.4	3.818 ± 0.273	2492

Як показано на рисунках 3.13, 3.14, 3.15 і 3.16, кількість ізольованих вузлів і різні параметри повідомлень не пов’язані в точному алгоритмі. Більше того, кількість ідентифікованих вузлів як ізольованих у точному алгоритмі майже ідентична кожному.

Таблиця 3.12 - Точний підсумковий результат алгоритму - #Msg = 250

Message Number = 250		
Dataset	Running Time (s)	# Isolated Nodes
200-0.2	0.037 ± 0.008	46
200-0.3	0.038 ± 0.001	44
200-0.4	0.039 ± 0.002	38
500-0.2	0.040 ± 0.003	96
500-0.3	0.113 ± 0.014	113
500-0.4	0.130 ± 0.016	120

1000-0.2	0.269 ± 0.009	222
1000-0.3	0.220 ± 0.010	227
1000-0.4	0.235 ± 0.008	232
10000-0.2	3.507 ± 0.216	2429
10000-0.3	3.534 ± 0.214	2388
10000-0.4	3.696 ± 0.158	2492

Таблиця 3.13 - Точний підсумковий результат алгоритму

Message Number = 500		
Dataset	Running Time (s)	# Isolated Nodes
200-0.2	0.034 ± 0.001	46
200-0.3	0.042 ± 0.008	44
200-0.4	0.039 ± 0.001	39
500-0.2	0.101 ± 0.007	97
500-0.3	0.106 ± 0.011	114
500-0.4	0.115 ± 0.005	119
1000-0.2	0.262 ± 0.008	224
1000-0.3	0.224 ± 0.015	226
1000-0.4	0.243 ± 0.014	233
10000-0.2	3.442 ± 0.175	2429
10000-0.3	3.448 ± 0.197	2387
10000-0.4	3.498 ± 0.126	2488

- #Msg = 500 мережі. Наприклад, згідно з рисунком 4.13, у мережах з μ = 0,2, μ = 0,3 і μ = 0,4, 46, 44 і близько 38 ізольованих вузлів було ідентифіковано в різних повідомленнях розміри відповідно.

Таблиця 3.14 - Точний кінцевий результат алгоритму - #Msg = 1000

Message Number = 1000		
Dataset	Running Time (s)	# Isolated Nodes

200-0.2	0.036 ± 0.002	46
200-0.3	0.040 ± 0.004	44
200-0.4	0.041 ± 0.001	38
500-0.2	0.100 ± 0.006	95
500-0.3	0.104 ± 0.010	114
500-0.4	0.118 ± 0.007	119
1000-0.2	0.261 ± 0.015	224
1000-0.3	0.223 ± 0.014	227
1000-0.4	0.238 ± 0.010	233
10000-0.2	3.500 ± 0.199	2427
10000-0.3	3.538 ± 0.204	2390
10000-0.4	3.545 ± 0.143	2490

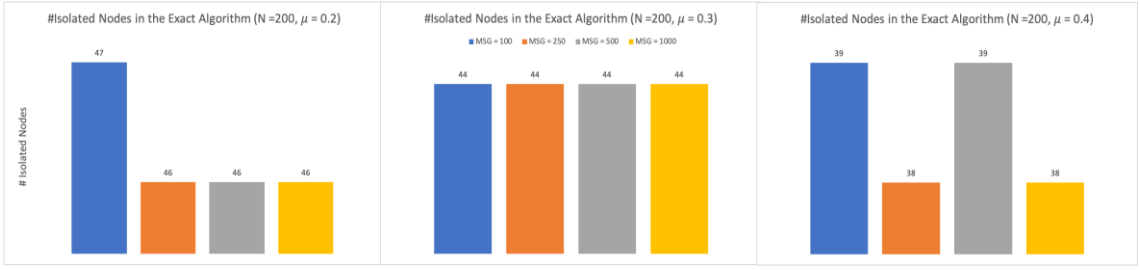


Рисунок 3.13 - Кількість ізольованих вузлів у різних налаштуваннях повідомлень (вузлів = 200)

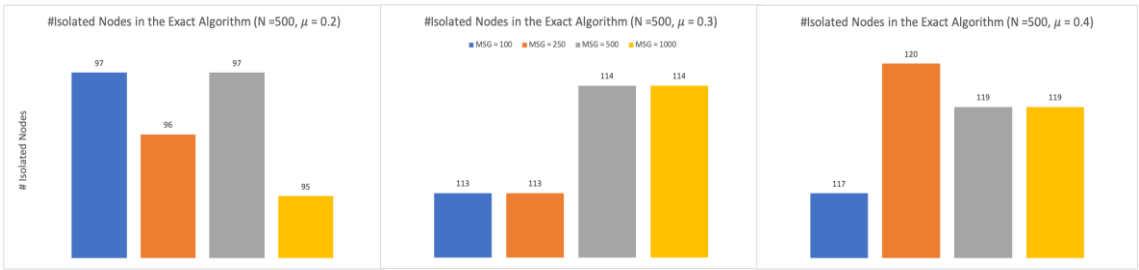


Рисунок 3.14 - Кількість ізольованих вузлів у різних налаштуваннях повідомлень (вузлів = 500)

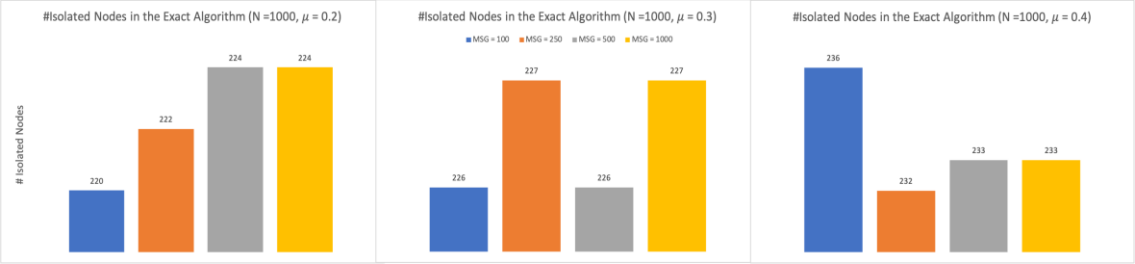


Рисунок 3.15 - Кількість ізольованих вузлів у різних налаштуваннях повідомлень (вузлів = 500)

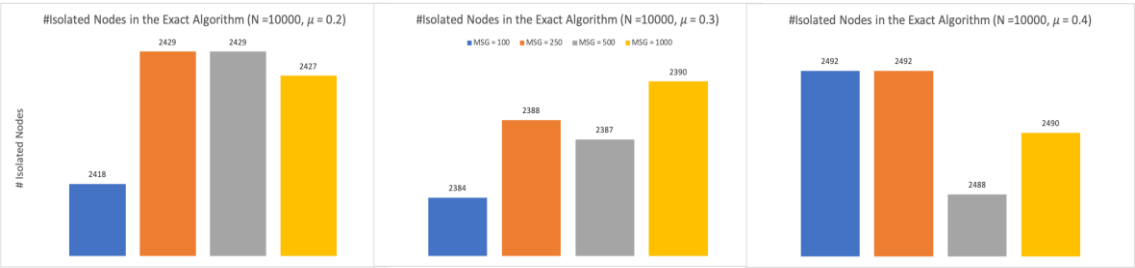


Рисунок 3.16 - Кількість ізольованих вузлів у різних налаштуваннях повідомлень (вузлів = 10000)

Таблиця 3.15 - Результати порівняння точних алгоритмів - #Msg = 100

Message Number = 100				
Dataset	Jaccard Max	Jaccard Mean	Overlap Coefficient Max	Overlap Coefficient Mean
200-0.2	0.541	0.469 ± 0.036	0.702	0.638 ± 0.33
200-0.3	0.396	0.305 ± 0.68	0.568	0.463 ± 0.79
200-0.4	0.418	0.396 ± 0.028	0.569	0.533 ± 0.029
500-0.2	0.447	0.426 ± 0.018	0.597	0.597 ± 0.018
500-0.3	0.448	0.421 ± 0.018	0.619	0.592 ± 0.018
500-0.4	0.453	0.421 ± 0.017	0.623	0.593 ± 0.017
1000-0.2	0.430	0.405 ± 0.023	0.601	0.576 ± 0.023
1000-0.3	0.425	0.410 ± 0.008	0.597	0.582 ± 0.008
1000-0.4	0.441	0.402 ± 0.031	0.612	0.573 ± 0.032
10000-0.2	0.385	0.381 ± 0.002	0.556	0.551 ± 0.002

10000-0.3	0.386	0.377 ± 0.005	0.557	0.548 ± 0.005
10000-0.4	0.373	0.368 ± 0.002	0.544	0.538 ± 0.003

Щоб порівняти результати нашого алгоритму з точним методом, ми проаналізували їх результати, використовуючи наступну процедуру. Вихід точного алгоритму порівнюється з усіма результатами, отриманими від алгоритму SBSID.

На кроці вкладення було повідомлено найкращу відповідь з максимальним значенням, а також середнє значення для індексів подібності. Наприклад, для мережі з 200 вузлами, μ = 0,2 і 100 повідомленнями, ми отримали 5 наборів відповідей, від S1 до S5, як показано на рисунку 3.17.

Таблиця 3.16 - Результати порівняння точних алгоритмів - #Msg = 250

Message Number = 250				
Dataset	Jaccard Max	Jaccard Mean	Overlap Coefficient Max	Overlap Coefficient Mean
200-0.2	0.533	0.456 ± 0.043	0.695	0.626 ± 0.039
200-0.3	0.419	0.309 ± 0.074	0.590	0.468 ± 0.085
200-0.4	0.407	0.358 ± 0.034	0.578	0.526 ± 0.037
500-0.2	0.465	0.428 ± 0.022	0.635	0.6 ± 0.021
500-0.3	0.439	0.416 ± 0.016	0.610	0.587 ± 0.016
500-0.4	0.182	0.162 ± 0.013	0.308	0.280 ± 0.020
1000-0.2	0.420	0.397 ± 0.22	0.592	0.568 ± 0.023
1000-0.3	0.427	0.410 ± 0.009	0.599	0.582 ± 0.009
1000-0.4	0.436	0.405 ± 0.028	0.608	0.576 ± 0.028
10000-0.2	0.384	0.378 ± 0.004	0.555	0.549 ± 0.004
10000-0.3	0.386	0.378 ± 0.005	0.557	0.548 ± 0.005
10000-0.4	0.374	0.369 ± 0.003	0.545	0.539 ± 0.003

Таблиця 3.17 - Точні результати порівняння алгоритмів - #Msg = 500

Message Number = 500				
Dataset	Jaccard Max	Jaccard Mean	Overlap Coefficient Max	Overlap Coefficient Mean
200-0.2	0.533	0.456 ± 0.043	0.695	0.626 ± 0.039
200-0.3	0.396	0.305 ± 0.068	0.568	0.436 ± 0.079
200-0.4	0.418	0.364 ± 0.028	0.589	0.533 ± 0.029
500-0.2	0.458	0.426 ± 0.023	0.628	0.597 ± 0.023
500-0.3	0.443	0.418 ± 0.017	0.614	0.589 ± 0.017
500-0.4	0.451	0.421 ± 0.015	0.621	0.593 ± 0.014
1000-0.2	0.416	0.389 ± 0.023	0.587	0.560 ± 0.024
1000-0.3	0.425	0.410 ± 0.009	0.597	0.582 ± 0.009
1000-0.4	0.436	0.403 ± 0.03	0.608	0.573 ± 0.03
10000-0.2	0.385	0.379 ± 0.003	0.556	0.549 ± 0.004
10000-0.3	0.386	0.377 ± 0.005	0.557	0.547 ± 0.005
10000-0.4	0.375	0.369 ± 0.003	0.546	0.539 ± 0.003

Таблиця 3.18 - Точні результати порівняння алгоритмів - #Msg = 1000

Message Number = 1000				
Dataset	Jaccard Max	Jaccard Mean	Overlap Coefficient Max	Overlap Coefficient Mean
200-0.2	0.533	0.456 ± 0.043	0.695	0.626 ± 0.039
200-0.3	0.419	0.309 ± 0.074	0.590	0.468 ± 0.085
200-0.4	0.407	0.358 ± 0.034	0.578	0.526 ± 0.037
500-0.2	0.450	0.422 ± 0.019	0.621	0.593 ± 0.019
500-0.3	0.443	0.418 ± 0.017	0.614	0.589 ± 0.017
500-0.4	0.451	0.421 ± 0.015	0.621	0.593 ± 0.014
1000-0.2	0.416	0.389 ± 0.023	0.587	0.560 ± 0.024
1000-0.3	0.427	0.410 ± 0.009	0.599	0.582 ± 0.009

1000-0.4	0.436	0.403 ± 0.030	0.608	0.573± 0.030
10000-0.2	0.385	0.378 ± 0.004	0.556	0.549 ± 0.004
10000-0.3	0.385	0.376 ± 0.005	0.556	0.546 ± 0.005
10000-0.4	0.375	0.369 ± 0.003	0.546	0.539 ± 0.003

Середнє значення подібності Жаккара і коефіцієнта перекриття, як видно з таблиці 4.16, становили приблизно 47% і 64%. Максимальна подібність була між S1 з 56 вузлами та точним набором, де індекс Жаккара становив близько 55 %, а індекс коефіцієнта перекриття — близько 70 %.

На рисунках 3.18 та 3.19 наведено індекси подібності (Жаккара, коефіцієнт перекриття) для результатів, отриманих за алгоритмом Exact. Хоча завдяки використанню синтетичних даних у деяких моментах можна побачити різку різницю між результатами, більшість балів знаходяться в одному діапазоні.

На рисунку 3.20 показано час роботи для точного алгоритму. Час роботи має пряме відношення до розміру кожної мережі. Тому, як зазначено, більшій мережі потрібен більший час для виявлення ізольованих вузлів. Крім того, час роботи для мереж однакового розміру майже однаковий, оскільки не залежить від кількості повідомлень.

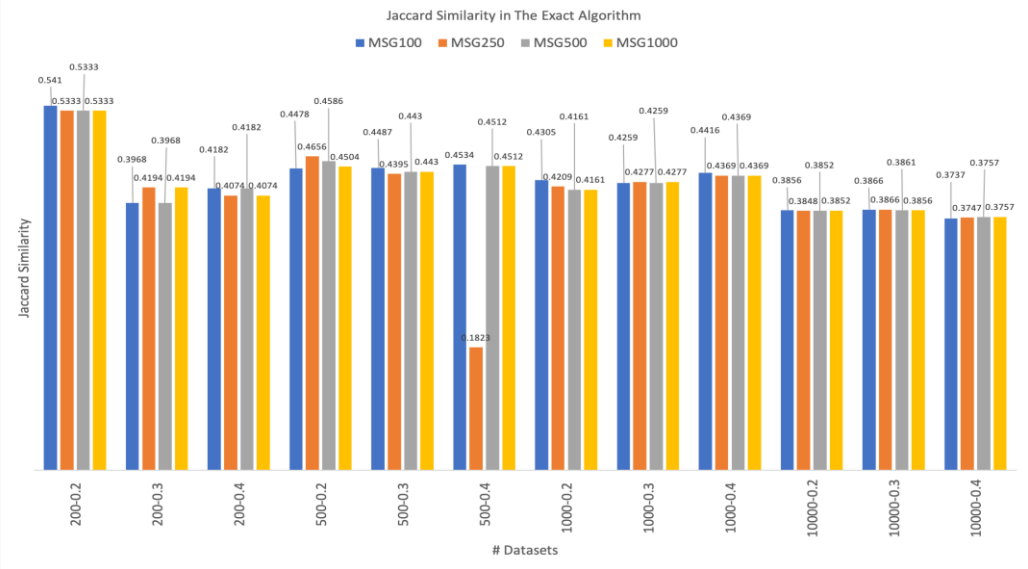


Рисунок 3.18 - Порівняння подібності Жаккара

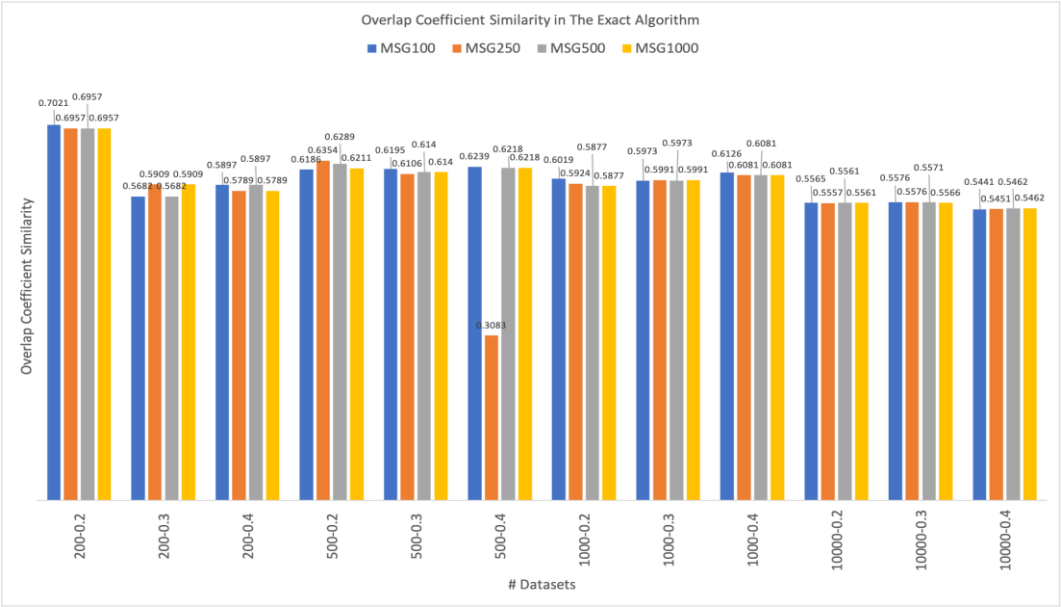


Рисунок 3.19 - Порівняння подібності коефіцієнта перекриття

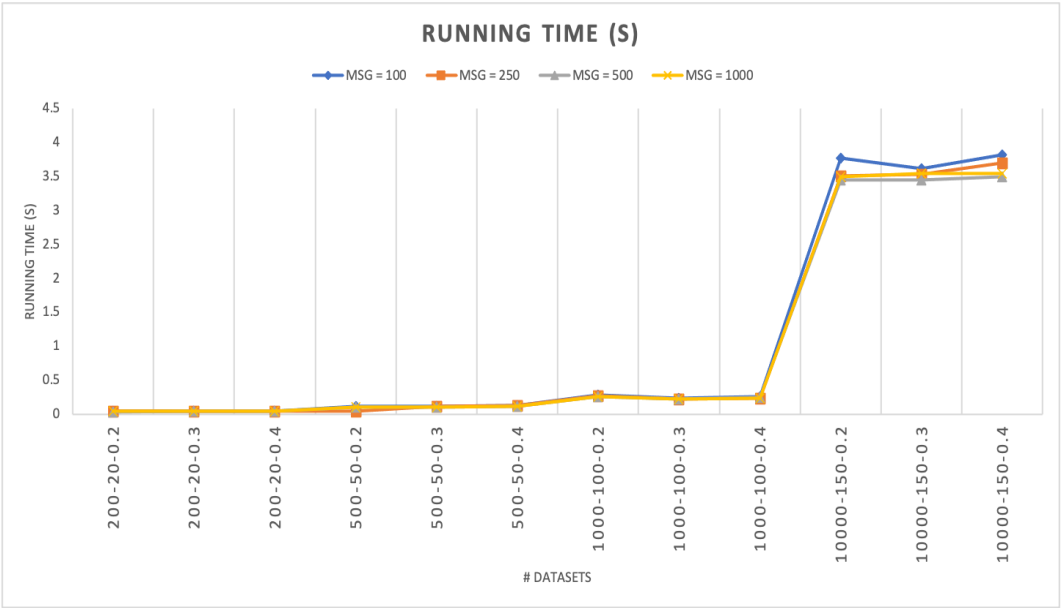


Рисунок 3.20 - Час відповіді на основі кількості повідомлень

3.5 Перспективи вдосконалення запропонованого алгоритму

Ефективність нашого методу була перевірена в синтетичних соціальних мережах, створених за допомогою тесту LFR. Однак, щоб точно і всебічно проаналізувати точність і ефективність цього методу, його слід перевірити на

реальних даних. У рамках нашої майбутньої роботи ми прагнемо більше випробувати запропоновані алгоритми на реальних даних.

Крім того, оскільки ми маємо справу з реальними даними в реальному світі, слід проаналізувати ризики конфіденційності для цієї моделі. Вхідними даними для алгоритму FBSID є список демографічних характеристик осіб; для алгоритму SBSID необхідна інформація про соціальні стосунки кожної людини, а вихідним результатом для обох алгоритмів є список соціально ізольованих осіб. Таким чином, фактори ризику з точки зору впливу конфіденційних даних мають бути проаналізовані та враховані в майбутньому, оскільки цей алгоритм буде реалізовано на даних про здоров'я.

Цей розділ містить додатковий спосіб, який ми використовували для оцінки продуктивності запропонованої моделі. У цій частині ми використовували реальні дані для оцінки нашої моделі.

Ми проаналізували десять буклетів, які належать кожному пацієнту, щоб отримати необхідну інформацію, необхідну для цього дослідження. Щоб зібрати необхідні дані для алгоритму SBSID, спочатку ми створили список усіх осіб, які мали зв'язок з пацієнтом. Таким чином, ми мали десять окремих мереж для кожного пацієнта. Щоб призначити вагу кожній ланці, яка з'єднує пацієнтів із їхніми родичами чи друзями, ми визначили три категорії для аналізу частоти, якості та різноманітності відносин.

• Частота показує кількість часу, який дві людини проводять з кожним інший.

• Якість включає в себе три класи членів сім'ї, далеких родичів разом із друзями та медичну підтримку для обчислення якості кожного зв'язку на основі статусу зв'язку та визначення пріоритетів .

ВИСНОВКИ

В дипломній роботі представлено підхід який полягає в тому, щоб відобразити дану соціальну мережу у зважений орієнтований соціальний граф на основі функцій, де кожен вузол представляє індивідуума із набором характеристик. Грані представляють існування відносин між індивідами, а вага кожного ребра показує якість і силу цих відносин. Тому ми запропонували обчислювальну модель, засновану на техніці передачі повідомлень, щоб отримати інформацію зі структури соціального графа для визначення норм суспільства та обчислити відстань кожного вузла від норми суспільства для ідентифікації ізольованих вузлів. Отже, ми запропонували кілька показників для вимірювання кількості та якості зв’язків між людьми в даній мережі шляхом аналізу характеристик вузла на графіку, кількості з’єднань для кожного вузла та тенденції кожного вузла спілкуватися з іншими. В результаті виходить ізольований вузол можна розглядати як зовнішній вузол, який веде себе інакше порівняно з іншими вузлами.

Ми також представили алгоритм визначення викидів у соціальних графах на основі атрибутів. Щоб охопити демографічні характеристики кожного індивіда, ми використали набір бінарних атрибутів і розробили метод визначення зовнішніх вузлів. Потім ці масиви використовуються для вилучення норм суспільства. Потім атрибути кожного вузла порівнюються з нормою, щоб виявити вузли, що виділяються.

Соціальна ізоляція шкідлива для здорового розвитку людини. Соціально ізольовані люди в цьому дослідженні розглядаються як ті особи, які не мають тенденції спілкуватися з іншими особами в мережах, або вони не в змозі спілкуватися ефективним способом, але можуть мати багато зв'язків, або не можуть встановити зв'язок з іншими завдяки характеристикам, які вони мають. У цій роботі ми запропонували нову обчислювальну модель і два алгоритми для виявлення соціально ізольованих осіб у соціальних мережах.

В роботі ми оцінили запропоновану модель на різних типах синтетичних соціальних мереж різних розмірів і конфігурацій. Ми провели численні експерименти, щоб проаналізувати точність, ефективність та надійність запропонованої нами моделі. Отримані результати показали, що наша модель може ідентифікувати ізольовані вузли з точністю близькою до 80% порівняно з точним методом. Результати також показують, що наша модель масштабована та здатна ідентифікувати ізольованих людей у великомасштабних соціальних мережах у лінійній часовій складності.

СПИСОК ПОСИЛАНЬ НА ДЖЕРЕЛА

1. Antonucci, T. (1985). Social support: Theoretical advances, recent findings, and pressing issues. In I.G. Sarason & B.R. Sarason (Eds.), Social support: Theory, research, and application. Boston: Martinus Nyhoff.© Jones and Bartlett Publishers, LLC. NOT FOR SALE OR DISTRIBUTION. 110 Chapter 5: Social Isolation

2. Cadman, D., Rosenbaum, P., Boyle, M., & Offord, D. (1991). Children with chronic illness: Family and parent demographic characteristics and psychosocial adjustment. Pediatrics, 87(6), 884–889.

3. Carpenito-Moyet, L.J. (2006). Social Isolation. In L.J. Carpenito-Moyet, Nursing diagnosis: Application to clinical practice (11th ed.), 734–736.

4. Catalan, J. (1998). Mental health problems in older adults with HIV referred to a psychological medicine unit. AIDS Care: Psychological and Socio-medical Aspects of AIDS/HIV, 10(2), 105–112.

5. Catlin, A.J. (1998). Editor's choice. When cultures clash, comments on a brilliant new book. The spirit catches you and you fall down. New York: Farrar, Straus, and Giroux.

6. Ceria, C.D., Masaki, K.H., Rodriguez, B.L., Chen, R., Yano, K., & Curb, J.D. (2001). The relationship of psychosocial factors to total mortality among older Japanese-American men: The Honolulu heart program. Journal of the American Geriatrics Society, 49(6), 725–731.

7. Charmaz, K. (1987). Struggling for a self: Identity levels of the chronically ill. In J. Roth & P. Conrad (Eds.), Research in the sociology of health care. Greenwich, CT: JAI Press.

8. Cheng, B.K. (1997). Cultural clash between providers of majority culture and patients of Chinese culture. Journal of Long Term Home Health Care, 16(2), 39–43.

9. Chinman, M.J., Weingarten, R., Stayner, D., & Davidson, L. (2001). Chronicity reconsidered: Improving person-environment fit through a consumer-run service. Community Mental Health Journal, 37(3), 215–229.

10. Christ, G. (1987). Social consequences of the cancer experience. The American Journal of Pediatric Hematology/Oncology, 9(1), 84–88.

11. Clark, D.J. (2002). Older adults living through and with their computers. Computers, Informatics, Nursing, 20(3), 117–124.

12. Cobb, S. (1979). Social support and health through the life course. In M.W. Riley (Ed.), Aging from birth to death. Boulder, CO: Westview Press. Cohen, M. (1993). The unknown and the unknowable— Managing sustained uncertainty. Western Journal of Nursing Research, 15(1), 77–96.

13. Collins, D., Fitzgerald, S., Sachs-Ericsson, N., Scherer, M., Cooper, R., & Boninger, M. (2006). Psychosocial Austin, D. (1989). Becoming immune to loneliness: Helping the elderly fill a void. Journal of Gerontological Nursing, 15(9), 25–28.

14. Barker, J.C. (1994). Recognizing cultural differences: Health care providers and elderly patients. Gerontology & Geriatric Education, 15(1), 9–21.

15. Barnes, L.L., Mendes de Leon, C.F., Wilson, R.S., Bienias, J.L., & Evans, D.A. (2004). Social resources and cognitive decline in a population of older African Americans and whites. Neurology, 63(12), 2322–2326.

16. Bendor, S. (1990). Anxiety and isolation in siblings of pediatric cancer patients: The need for prevention. Social Work in Health Care, 14(3), 17–35.

17. Bennet, R. (1980). Aging, isolation, and resocialization. New York: VanNostrand Reinhold. Berkman, L. (1983). The assessment of social networks and social support in the elderly. Journal of the American Geriatrics Society, 31(12), 743–749.

18. Berkman, L.F., & Syme, S.L. (1979). Social networks, host resistance, and mortality: A nine-year followup study of Alameda County residents. American Journal of Epidemiology, 109(2), 186–204.

19. Biordi, D. (2000). Research agenda: Emerging issues in the management of health and illness. Seminars for Nurse Managers, 8, 205–211.

20. Biordi, D. (1993). [In-home care and respite care as self- care] Unpublished data.

21. Boden-Albala, B., Litwak, E., Elkind, M.S., Rundek, T., & Sacco, R.L. (2005). Social isolation and outcomes post stroke. Neurology, 64(11), 1888–1892.

22. Borman, J., & Biordi, D. (1992). Female nurse executives: Finally, at an advantage. Journal of Nursing Administration, 22(9), 37–41.

23. Bricout, J.C. (2004). Using telework to enhance return to work outcomes for individuals with spinal cord injuries. Neurorehabilitation, 19(2), 147–159.

24. Browne, A.J. (1997). The concept analysis of respect applying the hybrid model in cross-cultural settings. West- ern Journal of Nursing Research, 19(6), 762–780.

25. Brummett, B.H., Barefoot, J.C., Siegler, I.C., Clapp- Channing, N.E., Lytle, B.L., Bosworth, H.B., et al. (2001). Characteristics of socially isolated patients with coronary artery disease who are at elevated risk for mortality. Psychosomatic Medicine, 63(2), 267–272.

26. Burnley, I.H. (1992). Mortality from selected cancers in NSW and Sydney, Australia. Social Science and Medicine, 35(2), 195–208.© Jones and Bartlett Publishers, LLC. NOT FOR SALE OR DISTRIBUTION. References 111 A longitudinal study of rheumatoid arthritis. Journal of the American Medical Association, 283, 2579–2584.

27. Fitzpatrick, R., Newman, R., Archer, R., & Shipley, M. (2000). Inequalities in racial access to health care. Journal of the American Medical Association, 284(16), 2053.

28. Fowles, D., & Greenberg, S. (2003). A profile of older Americans: 2003. Retrieved November 10, 2007, from www.aoa.gov/prof/statistics/ profile/2003/2003profile.pdf

29. Foxall, M., Barron, C., Dollen, K., Shull, K., et al. (1994). Low vision elders: Living arrangements, loneliness, and social support. Journal of Gerontological Nursing, 20, 6–14.

30. Foxall, M., Eckberg, J., & Griffith, N. (1986). Spousal adjustment to chronic illness. Rehabilitation Nursing, 11, 13–16.

31. Fratiglioni, L., Paillard-Borg, S., & Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. Lancet Neurol, 3(6), 343–353.

32. Fratiglioni, L., Wang, H.X., Ericsson, K., Maytan, M., & Winblad, B. (2000). Influence of social network on occurrence of dementia: A community-based longitudinal study. Lancet, 355(9212), 1315–1319.

33. Frierson, R.L. (1991). Suicide attempts by the old and the very old. Archives of Internal Medicine, 151(1), 141–144.

34. Fyrand, L. (2003). The effect of social network intervention for women with rheumatoid arthritis. Family Process, 42(1), 71.

35. Gallo, A.M., Breitmayer, B.J., Knafl, K.A., & Zoeller, L.H. (1991). Stigma in childhood chronic illness:A well sibling perspective. Pediatric Nursing, 17(1), 21–25.

36. Gamba, A., Romano, M., Grosso, I., Tamburini, M., Cantu, G., Molinari, R., et al. (1992). Psychosocial adjustment of patients surgically treated for head and neck cancer. Head and Neck, 14(3), 218–223.

37. Goodman, C. (1984). Natural helping among older adults. Gerontologist, 24(2), 138–143.

38. Greaves, C.J., & Farbus, L. (2006). Effects of creative and social activity on the health and well-being of socially isolated older people: Outcomes from a multi-method observational study. Journal of the Royal Society of Health, 126(3), 134–142.

БІБЛІОГРАФІЧНА ДОВІДКА

Тема дипломної роботи бакалавра: "Інформаційне моделювання соціальної активності в мережі"

Обсяг пояснювальної записки: 65 аркушів

Дата закінчення дипломної роботи 13 червня 2021р.

Підпис студента _____