3-10-2021

A Novel Computational Model for Social Isolation Detection in Social Networks

Saghi Khani
University of Windsor

A Novel Computational Model for Social Isolation Detection in Social Networks

By

Saghi Khani

A Thesis

Submitted to the Faculty of Graduate Studies

through the School of Computer Science

in Partial Fulfillment of the Requirements for

the Degree of Master of Science

at the University of Windsor

Windsor, Ontario, Canada

2021

A Novel Computational Model for Social Isolation Detection in Social Networks

by

Saghi Khani

APPROVED BY:

K. Pfaff

Faculty of Nursing

Z. Kobti

School of Computer Science

P. Moradian Zadeh, Co-Advisor

School of Computer Science

S. Samet, Co-Advisor

School of Computer Science

January 21, 2021

Declaration of Co-Authorship / Previous Publication

I. Co-Authorship

I hereby declare that this thesis incorporates material that is the result of research conducted under the supervision of Dr. Pooya Moradian Zadeh (Co-Advisor), Dr. Saeed Samet (Co-Advisor) and Dr. Kathryn Pfaff. In all cases, the key ideas, primary contributions, experimental designs, data analysis, and interpretation were performed by the author, and the contribution of the co-author was primarily through the proofreading of the published manuscripts.

I am aware of the University of Windsor Senate Policy on Authorship, and I certify that I have properly acknowledged the contribution of other researchers to my thesis and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

II. Previous Publication

This thesis includes two original papers that have been previously submitted in peer-reviewed conferences, as follows:

Chapters	Full Citation	Publication Status
1, 2, and 3	Zadeh, P. M., Khani, S., Pfaff, K., & Samet, S. (2020, July). A Computational Model and Algorithm to Identify Social Isolation in Elderly Population. In 2020 IEEE Symposium on Computers and Communications (ISCC) (pp. 1-6). IEEE	Published

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

III. General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis. I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

The human being is a social creature and needs to communicate with others to share information, emotions, and fulfill its basic needs. Social isolation can be considered as a serious health risk issue which not only has unignorably negative impacts on the well-being and quality of life of individuals, but also it is harmful to healthy human development. In this research, a computational model and a couple of novel algorithms are proposed to address social isolation detection in social networks. In our model, a given community is represented by a weighted-directed social graph. An algorithm, SBSID (Structure-based Social Isolation Detection), is proposed to detect socially isolated individuals based on the graph's structure by finding the number of each individual's active friends and their influence on each other. On the other hand, each individual's demographic characteristics in our model are represented by a set of binary attributes. Consequently, another algorithm is proposed, FBSID (Feature-based Social Isolation Detection), to address social isolation based on the nodes' features in the social graph. We propose a couple of metrics and formulas to calculate society's norms based on the overall structure and attributes of the social graph. Structural characteristics and attributes of each individual are compared with the norm of society to identify socially isolated individuals. We have evaluated the performance of our proposed model and algorithms on a set of synthetic networks. The results show that our model is capable of finding socially isolated nodes in various sizes of graphs with high accuracy and efficiency.

Dedication

I Am Dedicating This Thesis to Three Beloved People

Who Have Meant and Continue to Mean So Much to Me:

First and Foremost to

My Beloved and Supportive Parents

Alireza and Mitra

and

My Dearest Sister
Kowsar

Acknowledgements

I want to pay my special regards to my beloved family because they are always supporting me in all possible ways when we are miles and miles away from each other. I would like to express my deep and sincere gratitude to my supervisors, Dr. Pooya Moradian Zadeh and Dr. Saeed Samet, for their patience, motivation, enthusiasm, and immense knowledge. Without their guidance and helpful feedbacks, this dissertation would not have been possible.

I would like to thank my thesis committee members, Dr. Kathryn Pfaff and Dr. Ziad Kobti, for taking the time to review my thesis and attending my thesis proposal and defense sessions. Thanks to their guidance and valuable suggestions for improving this thesis.

Table of Contents

List of Figures

Chapter 1

Introduction

Humans are inherently sociable as members of this society, and interaction and cooperation are essential to keep a society alive. With the help of communications, individuals can achieve a vast amount of information to learn valuable lessons from others' experiences or develop new ideas to function better in different ways. Despite this fact, individuals, especially seniors, would feel lonely and isolated from time to time, and this feeling will be a real challenge when it becomes permanent. According to Statistics Canada, 1.4 million elderly Canadians report feeling lonely [1]. Almost 30% of Canadians live alone, and up to 50% of people over the age of 60 are at risk of social isolation due to factors such as: outliving family and friends, disability, life-threatening illness, caregiving, and low-income [2]. Social isolation is the next major public health epidemic that has remarkable effects on seniors' life [3]. With the rapid growth of geriatric populations around the world, solutions for this problem gained increasing attention as a public health priority because isolation has a destructive effect on physical health as well as psychological wellness.

1.1   Social Isolation

Social isolation has been conceptualized in a number of different ways, such as unwillingness to participate in social events, inability to make eye contact with others [4], the absence of social interaction among small groups (family and friends), or a large group of people in a society [5]. Moreover, there are some similar concepts to social isolation that may interchangeably be used while they have different meanings, such as loneliness, social anxiety, introversion, social anhedonia, and social exclusion.

- **Loneliness** is an unpleasant subjective feeling when there is a quality or quantity reduction in an individual's relationships with others. It can be said that loneliness is common among adults and associated with poor mental and physical well being and unhealthy lifestyle [6].

- **Social anxiety** is a disorder that mostly appears at puberty and usually happens when a person is nervous and uncomfortable in a wild variety of social situations such as interviewing with a recruiter, public speaking, or getting along with strangers [7, 8].

- **Introversion** is identified as a personal trait that an individual prefers to spend more time and energy on their inner world instead of communicating with others that would have adverse effects on socializing [9].

- **Social anhedonia** includes unpleasurable actions or thoughts and is known as a reduction in the desire to seek and participate in social events that we could enjoy before [10].

- **Social exclusion** is an unpleasant experience that is a result of childhood to adulthood. Social exclusion involves being overlooked by social groups such as family and friends. In other terms, social exclusion is the disability to create a united group of society members [11, 12].

As mentioned above, there are various definitions of social isolation based on diverse points of view; however, in this thesis research, we define social isolation as a measure reflecting the lack of interaction between an individual and a society.

1.2 The Importance of Social Isolation

Based on statistics, in the next 30 years, in 2050, about %16 of people around the world will be over 65 years old [13]. Due to insufficient economic and emotional resources such as retirement or losing loving ones, the elderly population is more at the risk of social isolation than others [14]. Meanwhile, research has shown that being socially isolated has irreparable consequences such as psychiatric disorders [15], physical disability due to chronic diseases [16], and premature death [17]. Therefore, social isolation should be considered a major health concern. Social isolation can happen at four different levels of interaction with others, such as individual, relationship, community, and societal.

- Personality traits and characteristics are associated with the individual level. Some of these features are like being over 65 years old or more [18], belonging to certain minority groups [19], mental disorders [15], or financial problems [20].

- At the next level, relationships, the duration of interactions, and communications with family and friends are considered.

- The main focus at the community level is on some social relations that exacerbate isolation, for example, living in neighborhoods with high crime rates, the lack of enough social activities, or living with insufficient amenities and public transportation [21].

- At the societal level, the positive or negative effect of interactions and connections on social participation are considered [22].

Moreover, due to the spread of the COVID-19 (the coronavirus disease 2019) and measures to restrict unnecessary activities, to reduce the number of people infected with COVID-19 applied in various countries, social isolation gained more attention. The COVID-19 pandemic caused a worldwide lockdown and forced many people to stay at their homes and change their daily routines to survive. To remain safe from the virus, individuals should strictly limit their contact with others while this separation has been linked to poor health, depression, high blood pressure, and mortality. On the other hand, quarantine and social distancing may also lead to mental health problems [23, 24].

Therefore, by early identifying people who suffer from social isolation and implementing suitable interventions, it is possible to reduce the serious related mental and physical health side effects.

1.3   Social Network

A network can be represented by a graph, including nodes and edges as links between a pair of nodes [25]. Consequently, a social network is a complex subset of a network that is a collection of social and personal interactions with different purposes representing a real world's relationships, in which people can share their thoughts, experiments, and feelings [26]. A social network consists of social actors with various cultures, professions, characteristics, and educational levels. The connection between these actors is different in terms of the degree of strength and importance. Some of these actors are more influential than others and significantly impact their friends' circle. To put it in another way, these prominent individuals act as leaders to control the network and motivate other members to obey them and establish more connections [27].

A social network is a complicated structure with some special characteristics such as community structure, small-world effect, and power-law degree distribution.

- **Community structure** is an indispensable part of a social network that helps

to have a better understanding of user behavior and network construction. In a graph representing a social network, the clustering coefficient helps to detect communities by measuring a node's tendency to join a cluster [28].

- The **small-world phenomenon** is also known as the six degrees of separation. The main idea behind this characteristic is that a short chain of acquaintances can link strangers. For instance, if we have a graph with three nodes, node #1 and node #3 may not know each other through a direct link, but node #2 can be the bridge between these two because it links to nodes #1 and #3 [25].

- A network is a **power-law degree distributed** or a scale-free network when a small fraction of nodes have many connections while a large number of noes have few [29].

1.4   Social Network Analysis

Social network analysis (SNA) can be defined as a practice of investigating the interactions among actors in a social network in order to study its characteristics and behaviors. It can be said **structural** and **composition** are two variables in SNA. In many real scenarios, structural variables are more important than the composition ones since the focus is on the various ties among social actors such as trust, friendship, etc. Meanwhile, composition variables are mostly about an actor's attributes [30]. Using SNA, researchers can study how individuals establish connections into a broader social structure. Moreover, SNA can be used to show how individuals try to create relationships and how they affect each other [31].

1.5   Research Motivation

The connection between mental and physical health is undeniable, and the quality of individuals' life depends on this association. Social isolation could have adverse effects on physical, social, and psychological well-being at different levels. For instance, at the individual level, mental and physical wellness is considered. Some of the destructive effects on individuals are premature death, depression, dementia, and disability as the result of chronic disease [16]. At the community level, higher crime rates, low income, limited availability of public transportation, and inadequate opportunities for social interaction are some examples [22]. Furthermore, social isolation can lead to an extended hospitalization's period and inappropriate occupation of hospitals' beds and medical resources [32].

Meanwhile, despite the importance and urgency of this problem, there are very few research works in the field that uses technology to deal with the social isolation related issues in communities, and to the best of our knowledge, there exist no computational algorithms to deal with this problem efficiently. Therefore, developing an efficient and feasible computational approach to detect social isolation in a large population, can be used as a decision support system which leads to an enhancement in both individual and community levels. At the same time, it will reduce the chance of associated disorders among people who suffer from isolation.

1.6   Problem Statement

As mentioned before, research has shown that early preventative intervention methods are more beneficial and valuable in terms of approaching the problem of social isolation [22]. Therefore, our primary goal is to develop a computational model and a set of algorithms using social network analysis techniques for detecting isolated individuals in a network.

In this research, the problem of finding socially isolated nodes are defined from two

different perspectives based on the network's structure and the individuals' features.

1.6.1 Structural-based Social Isolation

Assume a community is represented by a weighted directed graph G(V, E,w). V is a set of nodes representing our individuals in a community, and E is a set of edges between each pair of them, $e = (v_i, v_j) \in E$, where $v_i, v_j \in V$. Furthermore, the weight function $w : E \to R$ assigns a real value between 0 and 1 to every edge. The weight is used to measure the quality and the strength of the connections between the pairs. Therefore, a higher weight represents a stronger tie. Meanwhile, if the weight of the edge between v_i and v_j is x, the weight of the edge between v_j and v_i will be x'. Hence, the value of x is not necessarily equal to the value of x'.
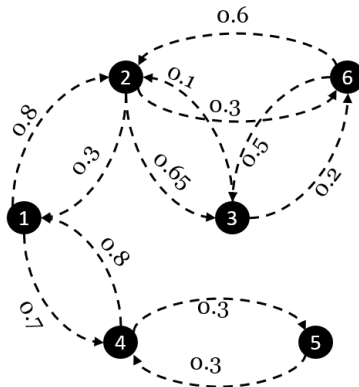
FIGURE 1.1: A weighted directed social graph

For example, in Figure 1.1, we have six nodes representing six individuals in our community. The weight of the link between nodes one and two is 0.8, but the weight of the link from nodes two to one is 0.3. The weight can be calculated based on the frequency of the connections between a pair of nodes, duration of the connection, number of common friends, similar group memberships, level of dependency, degree of influence on each other, or other metrics. Therefore, the degree of a node can denote that if an individual does not have a tendency to interact with others, the edge's weights can be used to calculate the degree of influence and a node's social acceptance by its neighbors.

Suppose the number of a node's active connections (considering the edge weights) is less than one standard deviation below the community's mean value (the subtraction of the mean and the standard deviation for the given community); in that case, we classify that node as a socially isolated node. For example, assuming the average number of active connections in a community is 2 with the standard deviation of 0.63, any node with less than 1.37 active connections can be classified as an isolated node. For instance, in Figure1.1, node five can be identified as an isolated node.

1.6.2   Feature-based Social Isolation

This algorithm uses the same weighted directed graph G(V, E, w) in the previous algorithm, representing a human being community. However in this part, each individual, a_i, has a set of n features that can be defined as $F_{a_i} = \{f_1, f_2, ..., f_n\}$. These features represent a fixed-size binary vector where a cell's value is 1 if the person has the corresponding feature and is 0 otherwise. For instance, Figure 1.2 shows a feature vector if ten characteristics considered for each individual. Table 1.1, shows a list of various features. This set of features are not basic characteristics. In other words, research has shown some of our characteristics or behaviors can exacerbate isolation. Some of these features are like: being over 65 years old [13], smoking[33], gender minority [34], physical disability [35], poverty [20], living at countrysides [36], living alone [37], obesity [38], marital status [39]. For example, assume the first and second cells of this vector represent features of being over 65 years and the smoking status based on Table 1.1, respectively. Consequently, this person is under 65 years old and a smoker. The rest of the cells can be interpreted similarly.

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}

FIGURE 1.2: Sample of a binary vector

A sample of our hypothetical network that has both weights and feature sets is

Features	Value	Features	Value
Over 65	1	Under 65	0
Smoker	1	Non-Smoker	0
Gender Minority	1	Cisgender	0
Physical Disabilities	1	Being Healthy	0
Poverty	1	Enough Wealth	0
Rural	1	Urbanize	0
Living alone	1	Not alone	0
Obese	1	Non-obese	0
Unemployment	1	Having a job	0
Single	1	Not single	0

TABLE 1.1: Binary values for features

shown in Figure1.3. Therefore, we define the problem of social isolation detection as finding a set of nodes in the graph, which either:

(a) have no tendency to communicate with other nodes in the graph; or

(b) are not able to communicate in an efficient way while they may have many connections; or

(c) can not establish a connection with others due to the characteristics they have.
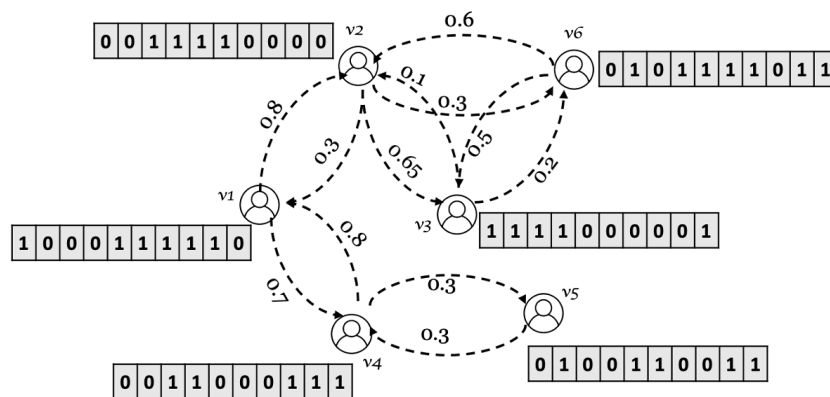
FIGURE 1.3: A featured-based social graph

1.7   Research Objectives

The main objective of this research is to propose a computational model to tackle the problem of social isolation detection from both structural and feature-based perspectives. In order to achieve this objective, we have three main steps. The first step is to represent a community using a social graph. The second one, is to define a set of metrics to measure the quality and quantity of connections between individuals. Finally, the last step is to develop a couple of algorithms to efficiently identify socially isolated individuals based on the topology of the network and their own personal features.

1.8   Research Contribution

The main contribution of this thesis can be summarized in following items:

- Modeling social isolation in social networks using weighted-directed social graphs

- Defining a couple of mathematics formula to measure social isolation in a network

- Developing a messaged-based algorithm to identify isolated nodes in weighted-directed graph

1.9   Thesis Outline

The rest of this thesis research structured as follows:
Literature review and related works are reviewed briefly in section 2. Our proposed model and algorithms are presented and describe in section 3, and section 4 is devoted to the algorithms' evaluation and analysis. Finally, the last chapter is our conclusion and future works.

Chapter 2

Related Works

In this chapter, related research works in this field are reviewed. Previous works can be divided into three major fields, such as graph-based outlier detection methods, social isolation prevention among the aging population, and social isolation detection based on questionnaires and surveys.

2.1   Social Isolation Interventions

Due to the destructive effect of social isolation, with the help of preventional methods, isolation's adverse consequences will be decreased to lower levels. Based on [22], there are six types of interventions to fight social isolation, such as individual-based, group-based, service provision-based, technology-based, neighborhood-based, and structural-based interventions.

The first type, one-to-one interventions, is the connection between older adults and a professional or a volunteer to help them. The connection is base on common interests, and its positive effects on aging adults are proved while it is a cost-effective process [40, 41].

In group interventions, a group of people with mutual interests gather together to have social, physical, or educational activities. This method is more effective for

individuals who belong to minority groups. Hence, they can share their concerns with people of the same language or culture [40, 42].

Service provision interventions offer long-term or medium-term support for seniors. For instance, eleven local Age UK services in England tried to involve older adults who suffer from health issues in exercise groups [43]. Based on reports, seniors felt less isolated after exercising. Another service can include students trying to grow seniors' technology skills. Therefore, isolated seniors would feel less lonely and depressed by contacting their family and friends [41].

Interventions in the context of technology would encourage aging people to use more social networking sites and smartphones to be able to improve their social connections through internet [44].

Due to walking limitations in older adults, neighborhood intervention is less popular than other approaches. However, by providing safer seating areas or easy access to public transportation, more social opportunities would be created for seniors [45].

A higher level of prevention, such as structural interventions is about encouraging older adults to have more interactions within their society. For example, one case could be seniors employment in their previous job as a mentor or a new job with a training option for all employees with flexible working hours [45]. Moreover, with recent advancements in machine learning and robotics, robots are capable of simulating most humans activities.

In [46], the authors' focus is on how to make the old population's life more manageable with the help of self-directed robots. Although self-care robots would make seniors more isolated due to less communication with others, robots help them to do daily tasks without any need for caregiver workers or family members.

2.2   Social Isolation Detection Based on Medical Surveys

In [47], the authors collected data from the University of Michigan Health and Retirement Study, (HRS). Subjective social isolation (loneliness) was measured by a three-item scale questionnaire introduced by the HRS called the Psychosocial and Lifestyle Questionnaire. In order to detect objective social isolation, they used surveys focused on social interactions. They identified isolated people with two thresholds, mean and standard deviation. For the score of social cohesion less than one standard deviation below the mean (the subtraction of mean and standard deviation), those individuals were identified as *isolated*, individuals with a score within the subtraction of mean and standard deviation identified as *not isolated*, and people with a score higher than the subtraction result identified as *well-connected*.

In [48], the authors used a swarm intelligence algorithm named Firework Algorithm (FA) to explain the concept of social isolation. FA is an algorithm that explores solutions by choosing random points confined in a distance metric to find promising results that yield a concentrated search nearby. This work focuses on describing the behavior of socially isolated individuals who are often faced with a lack of equal opportunities and adequate social programs.   In [49], the authors proposed a novel and accurate natural language processing (NLP) approach for identifying isolated individuals. They used data from patients with prostate cancer collected by the Medical University of South Carolina (MUSC) Research Data Warehouse. The authors used some patients' information to detect isolated patients, such as their physical issues, illness history, discharge notes, and radiation oncology.

In [50], the authors collected data from an online survey with more than 200 Indian participants from the age of 18 to 56. The authors used two measures, extraversion and well-being scales, to discover isolated individuals.

In [51], the authors proposed a new predictive model, including four steps, based on older adults' communications and their mobility activities. The authors carried

out a questionnaire applied to 144 older adults, age range from 60 to 89. This study
was conducted in Cuernavaca, Mexico, based on capturing older adults' social inter-
actions over a month. After the first step, data collection, attribute selection has
been applied to remove redundant and unhelpful attributes. The next step was clas-
sification to arrange seniors based on the level of their social isolation. The final step
was to balance the obtained dataset to have an accurate prediction with the classified
categories. The result of this research was two applications: the adults' app and the
caregivers' app who want to support isolated seniors voluntarily.

2.3   Graph-based Outlier Detection

The concept of outlier or anomaly detection in graphs is closely related to this
thesis research. In this research, the main focus is on detecting socially isolated peo-
ple in a network that can be considered outliers. In [52], the authors suggest a novel
method to detect outlier edges based on the subtraction of two sets of edges. In other
words, the number of actual and existing edges are compared with the number of
expected edges in a graph. The number of expected edges is calculated based on two
random graph generation algorithms, Erdos-Reyni, and Preferential attachment. The
subtraction of two sets of edges is called *authentic score* in this model.

In [53], the authors proposed a new model to discover anomalies in large scale
networks. The authors used the concept of common neighbors. Nodes would be as-
signed to a cluster based on the way the shared neighbors with each other. Nodes
without clusters would be classified as outliers or hubs. If the isolated node has two
or more edges, it would be hub; otherwise, it is considered an outlier.

Authors in [54] introduced a new data structure called $DTAR$ stands for directly
two-hop-away reachable, to optimize the SCAN algorithm to apply it on large scale
networks. The authors claimed that the SCAN++ has the same results in detecting
outliers, hubs, and clusters with much less time consumption and complexity.

In a conclusion, we have reviewed some of the previous works related to social

isolation. Some studies' primary focus is on the preventional method to reduce the suffering of isolation among society members. Other research are based on online or in-person questionnaires. However, almost all of these works are done manually based on questionnaires and surveys without using any technologically-based approaches. Therefore, to the best of our knowledge, there is not any efficient existing computational model and algorithm to deal with the problem of social isolation detection.

Chapter 3

Methodology

As explained before, our model consists of two algorithms, Structural-based Social Isolation Detection (SBSID), that deals with the number of active connections of each node and the quality of the relationship between a pair of nodes. On the other hand, Feature-based Social Isolation Detection (FBSID) concentrates on the individuals' set of features. In this chapter, we describe these algorithms' components in detail.

3.1   Proposed Scheme - SBSID

The core of our proposed model is to identify the number of active connections of each individual in a network and their degree of influence on their circle of friends. To tackle this issue, we propose a model using the concept of information propagation in social networks. In fact, in our model, a certain number of messages are sent to all the network's individuals and they are asked to resend the received messages to their immediate neighbors. Then, the number of friends of each individuals who have received those messages are recorded. Furthermore, the number of attempts of resending messages are also captured. In this way, we would have an idea about the size of the active circle of friends for each person in the network. Furthermore, by tracking the number of resending messages, we will gain information regarding a

person's influence and acceptance level.

In this algorithm, we represent a community by mapping it to a weighted-directed graph, as we explain it before. The method is consists of three main phases:

1. A certain number of messages is sent to each node in a graph, and those nodes will broadcast these messages to their immediate neighbors.

2. The number of received messages by immediate neighbors will be counted to determine the number of active friends. An active friend is a friend who receives at least a message.

3. The number of attempts the receivers will perform to send the received messages to their circle of friends is monitored. In this step, we determine the influence of each node on its circle of friends.
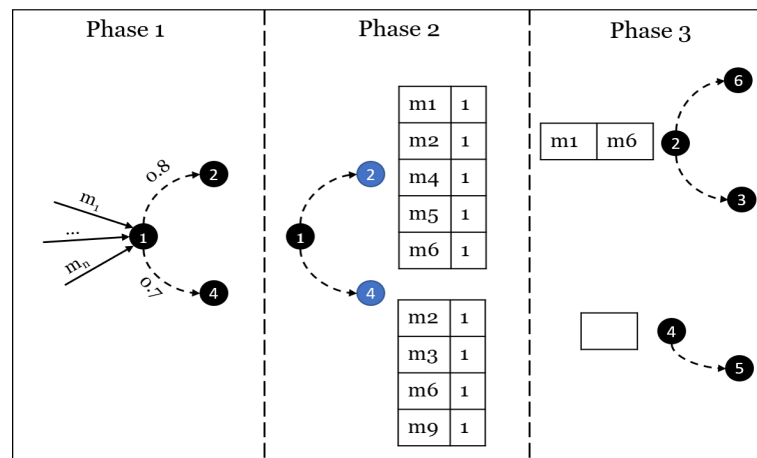
FIGURE 3.1: Sending m messages to node #1, assume m=10

For example, as shown in Figure 3.1, we have a network, including 6 nodes. We assume the number of messages is 10, and our focus is node 1. Based on the quality of a pairs' relationship, the node's neighbors will receive a portion of 10 messages. Node 1 has two immediate neighbors, nodes 2 and 4; It sends 10 messages to these neighbors and asks them to resend the portion of messages they received to their neighbors. In the second phase, as can be seen, node 2 received five, and node 4 received four messages. We use a two-dimensional array to record the received messages and the

id of their sender nodes. In phase three, one of the neighbors, node 2, sends just two of its received messages to its neighbors, and node 4 does not send any messages at all.
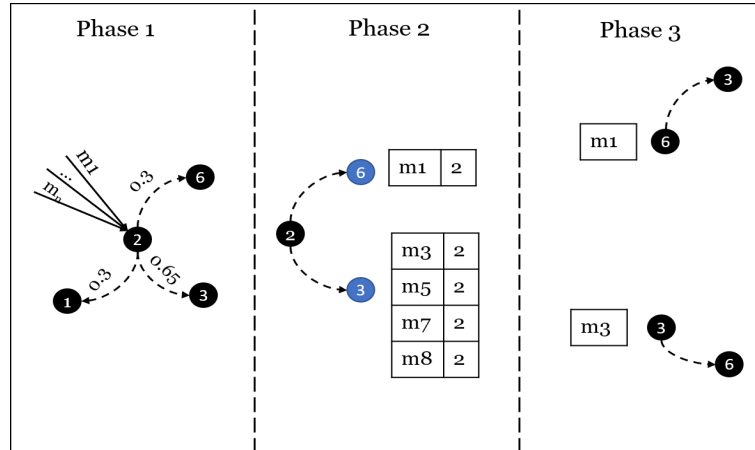
FIGURE 3.2: Sending m messages to node #2, assume m=10

Similarly, in Figure 3.2, the focus is on node 2, so we repeat the process. This node has three neighbors, but just two of them, nodes 3 and 6, received a portion of messages. Node 6, resend the only message that it has received from node 2 to its neighbors. On the other hand, node number 3, resend just one of the four received messages to its circle of friends. This process repeats for all the nodes in the network. In the end, on an individual level, we can identify the number of active friends (those friends who at least received a message) and the influence of the node on its circle of friends.

At the population level, we can identify the average number of active friends and social influence in this society, which can be interpreted as the community's social norms. So if a node has less than one standard deviation below the community's mean, we label it as a socially isolated node. Accordingly, let $c\_index(v_i)$ denotes an index to measure the estimated number of received messages by the neighbors of node $v_i \in V$ (phase 2). Hence,

$$c\_index(v_i) = m * \sum_{v_j \in N_G[v_i]} w(v_i, v_j).w(v_j, v_i) \tag{3.1}$$

where $N_G[v_i]$ is the set of neighbors of node v_i, and $w(v_i, v_j)$ returns the weight of the edge between nodes v_i and v_j, and m is the number of messages. In fact, having more connections and stronger links have positive impacts on this index. Meanwhile, in the population level, the average value for c_index can be calculated as:

$$avg\_c\_index = \frac{\sum_{v_i \in V} c\_index(v_i)}{|V|} \qquad (3.2)$$

On the other hand, let a_index(v_i) denotes an index to measure the estimated number of times that a message resend by the neighbors of node v_i (phase 3). Therefore,

$$a\_index(v_i) = \frac{\sum_{v_j \in N'_G[v_i]} a\_inx(v_i, v_j)}{|N'_G[v_i]|} \qquad (3.3)$$

where $N'_G[v_i]$ is a set of neighbor nodes of node v_i that have already received m' number of messages in phase 2. Meanwhile, $a\_inx(v_i, v_j)$ here measures the estimated number of messages which will be resent by node v_j and is defined as:

$$a\_inx(v_i, v_j) = m' * w(v_j, v_i) * (1 - \frac{\sum_{v_k \in N_G[v_j]} w(v_j, v_k)}{|N_G[v_j]|}) \qquad (3.4)$$

Algorithm 1 represents the pseudocode of our approach.

Algorithm 1: SBSID

Input: A weighted directed graph, G(V,E,w)

Output: A list of socially isolated nodes

1: initialization;

2: **for** all $v_i \in V$ **do**

3:     **for** $j = 1$ to m **do**

4:         send a message to all $v_j \in N_G(v_i)$;

5:         count the number of received messages;

6:         create $N'_G(v_i) \subseteq N_G(v_i)$;

7:         **for** all $v_j \in N'_G(v_i)$ **do**

8:             count the number of attempts to resend the message;

9:         **end for**

10:     **end for**

11: **end for**

12: **for** all $v_i \in V$ **do**

13:     **if** $(c\_index(v_i) < avg\_c\_index - \sigma)||$

        $(a\_index(v_i) < avg\_a\_index - \sigma)$ **then**

14:         $isolated\_list \leftarrow v_i$;

15:     **end if**

16:     Return $isolated\_list$;

17: **end for**

3.2   Proposed Scheme - FBSID

This model's fundamental part is the individual's inherited and environmental characteristics [55]. Some of these features are living in rural areas [36], smoking [33], and having financial problems [20]. This part of the algorithm concentrates on collecting some of these special features and identifying isolated nodes in a given network based on them. Each node has a set of a fixed-size binary vector named feature set, and we need a gauge as a healthy society's feature set to compare each node's feature

vector with it and detecting abnormals. Hence, we define another vector called the index vector, a binary vector with the same length as feature vectors, unique for the given network, and shows a set of normal features for a healthy population. In order to recognize the difference between the vector of features and the index vector, we compute the Hamming distance, which is one of the best metrics to compare two binary vectors. Therefore, this algorithm, the same as the previous one, has three main phases:

1. Create the summation vector that is a vector as a result of summing the cell values of the individuals' feature vectors.

2. Forming the index vector. If the value of each cell of the summation vector is at least one unit greater than the size of the population divided by two, the cell's value will be 1, otherwise 0.

3. Compute the Hamming distance between each individual and the index vector.

For instance, we consider the Figure 3.3 as our hypothetical network. As can be seen, we have six nodes with different sets of characteristics. In the first step, based on the algorithm, we have to compute the summation vector shown in the Figure 3.4.
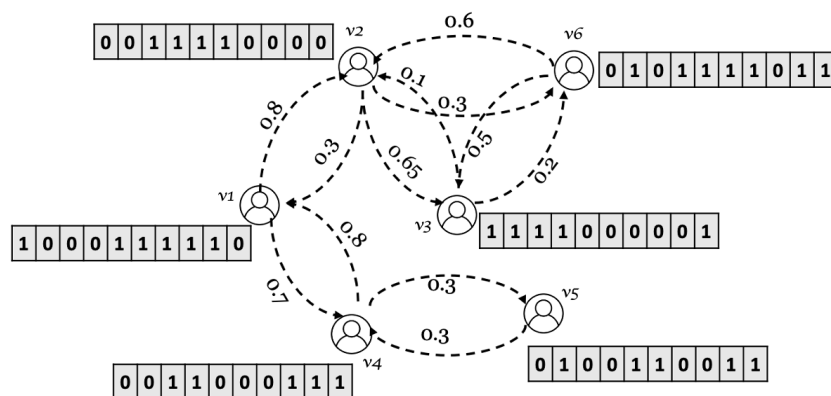
FIGURE 3.3: A featured-based social graph
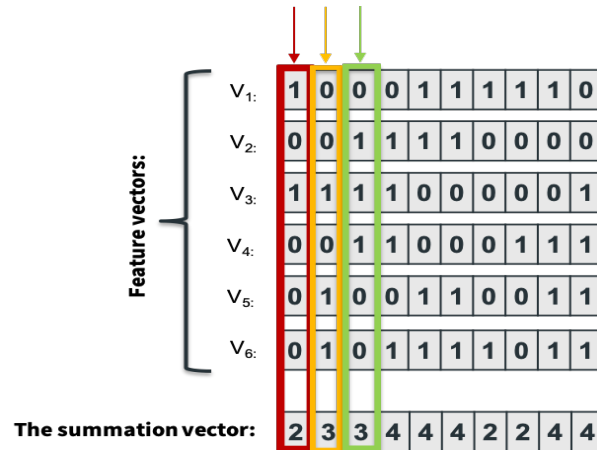
FIGURE 3.4: A sample of computing the summation vector

After the summation vector, we need to generate the index vector. In this research, if more than half of the population has a feature, we consider it a normal characteristic. We calculate the summation of one and the value for half the population size due to this definition.
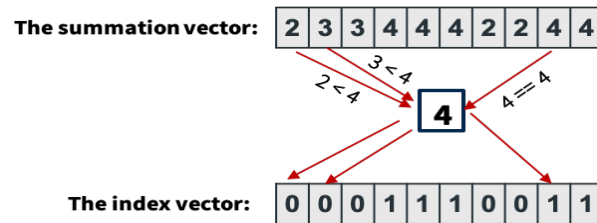
FIGURE 3.5: A sample of generating an index vector

As indicated in Figure 3.5, if the value of each cell in the summation vector is less than 4, the corresponding value in the index vector would be 0, otherwise 1. In the next step, with having the feature and index vectors, we compute the hamming distance for each node Figure 3.6. If each node's hamming distance is bigger than the summation of the average and standard deviation, that is identified as a node with a different set of features from the society's norm. In fact, the summation of the hamming distance's average and standard deviation (5.33) indicates the maximum valid distance from society's norm. Therefore, we can consider node 5 as a different node from others.
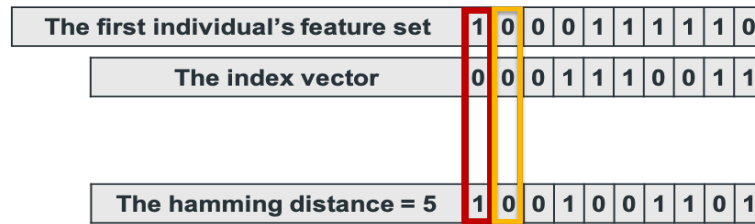
FIGURE 3.6: A sample of calculating the Hamming Distance

FIGURE 3.7: A sample of the comparing the Hamming Distance

Average of Hamming Distance $= 3.666 \approx 3.7$

Standard Deviation of Hamming Distance $= 1.632 \approx 1.63$

Avg + Std $= 5.33$

Algorithm 2 is the pseudocode of the FBSID algorithm.

3.3   Implementation Strategies

In order to implement the algorithms, we can consider three different strategies. In the first strategy, both algorithms can be executed simultaneously but independently. As shown in Figure 3.8, FBSID and SBSID are run simultaneously on the same dataset to detect a set of nodes with weak and low-quality relationships and identify those nodes with suspicious characteristics that intensify social isolation.

Algorithm 2: FBSID

Input: A weighted directed graph, G(V,E,w), with a set of binary
characteristics called feature vector

Output: A list of nodes with different characteristics from the majority of
the population

 1: initialization;
 2: **for** all $v_i \in V$ **do**
 3:    Compute the summation vector
 4:    $summation\_vector \leftarrow \Sigma f_i$;
 5:    **for** all item $\in$ summation_vector **do**
 6:       **if** item $\geq$ (half_of_population + 1) **then**
 7:          $index\_vector \leftarrow 1$;
 8:       **else**
 9:          $index\_vector \leftarrow 0$;
10:       **end if**
11:    **end for**
12:    hamming_distance $\leftarrow$ Compute the hamming distance for each f_i and the
       index_vector;
13:    $avg =$ Compute the average for items in the hamming distance;
14:    $std =$ Compute the standard deviation for the items in the hamming distance;

15:    norm $= avg + std$;
16:    **for** all item $\in$ hamming_distance **do**
17:       **if** item $>$ norm **then**
18:          $isolated\_list \leftarrow item$;
19:       **else**
20:          item is a normal individual;
21:       **end if**
22:    **end for**
23: **end for**
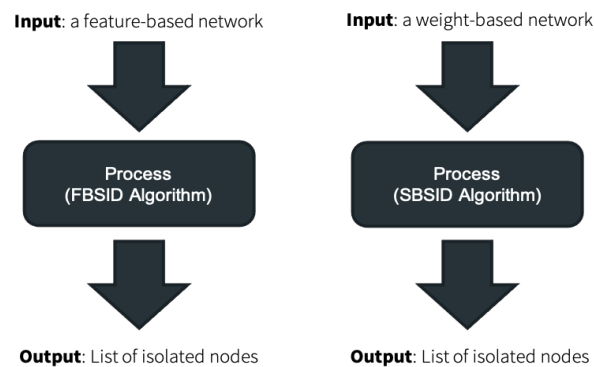
FIGURE 3.8: Implementation strategy 1: both algorithms run in parallel

In the second strategy, the main focus is on identifying nodes with weak and low-quality relationships. In the next step, the FBSID algorithm checks those detected nodes' characteristics. In fact, in this approach, we analyze algorithms at a macro level, which means we consider their connection with others, then we focus on each node's characteristics, Figure 3.9.

However, the third strategy is the exact opposite of the second one. The concentration is on a micro-level, which means it starts by detecting nodes with suspicious attributes that may cause isolation and then measure the quality of nodes' interactions with others, Figure 3.10.
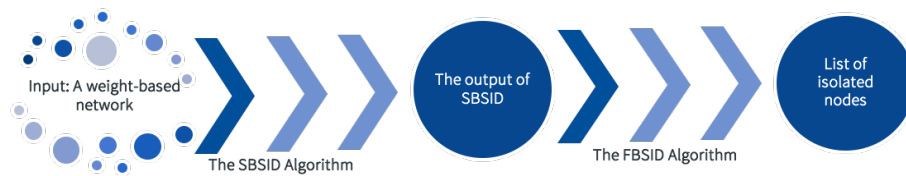
FIGURE 3.9: Implementation strategy 2: a sequential of the SBSID algorithm and the FBSID algorithm
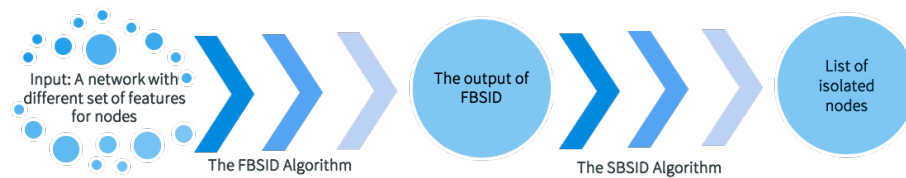
FIGURE 3.10: Implementation strategy 3: a sequential of FBSID algorithm and the SBSID algorithm

Chapter 4

Evaluation

In this chapter, we evaluate the performance, efficiency, and accuracy of our proposed model. To evaluate our algorithms (SBSID and FBSID), we will generate various synthetic networks using the Lancichinetti-Fortunato-Radicchi (LFR) Benchmark. LFR benchmark creates networks with real-world social network features such as community structures and power-law degree distribution [56, 57]. We also compare the performance of our model with the Exact approach.

4.1   Setup

We have generated four sets of weighted-directed social graphs with 200, 500, 1000, and 10000 nodes. For each set, three different graphs were generated with different complexity levels. The weights were generated randomly by the LFR benchmark and have been normalized to be in the range of zero to one. In total, 12 networks with the following parameters have been generated as indicated in Table 4.1. In addition, the following parameters, have been used to generate these synthetic networks:

- $k = 15$, k represents the average degree for each node in the graph.

- $maxk = 50$, sets the maximum degree size for each node in the graph.

- $\mu = 0.2,\ 0.3,\ 0.4$, μ is the mixing parameter, which shows the fraction of the number of edges among each community to the total number of them.

- $minc = 20$, is the minimum number of communities in a graph.

- $maxc = 50$, is the maximum number of communities in a graph.

Label	# Nodes	# Edges	μ	Maximum Degree
Network #1	200	5154	0.2	20
Network #2	200	5434	0.3	20
Network #3	200	5618	0.4	20
Network #4	500	13484	0.2	50
Network #5	500	13600	0.3	50
Network #6	500	14650	0.4	50
Network #7	1000	28800	0.2	100
Network #8	1000	26394	0.3	100
Network #9	1000	27892	0.4	100
Network #10	10000	312988	0.2	150
Network #11	10000	303700	0.3	150
Network #12	10000	311888	0.4	150

TABLE 4.1: The list of synthetic networks generated by the LFR benchmark

4.2 Experiments

We have conducted several experiments on both FBSID and SBSID algorithms to measure their run-time and evaluate their outputs. All the experiments were conducted on a MacBook Pro, macOS Cataline v10.15.6 with a 2.2 GHz Quad-Core Intel Core i7 CPU and 16 GB RAM. Python programming language was used to develop the experimental model on the Pycharm framework.

4.2.1 Feature-based Social Isolation Detection Algorithm

We have conducted 12 independent experiments, and we have repeated each of them 5 times in order to obtain the average values of the results. The obtained results

for the FBSID algorithm are shown in Table 4.2. In this table, the number of isolated nodes in each dataset and the average running time of the algorithm for each graph are presented. Meanwhile, the results are also illustrated in Figures 4.1, and 4.2.

Dataset	Running Time (s)	# Isolated Nodes
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

TABLE 4.2: The results obtained from FBSID algorithm
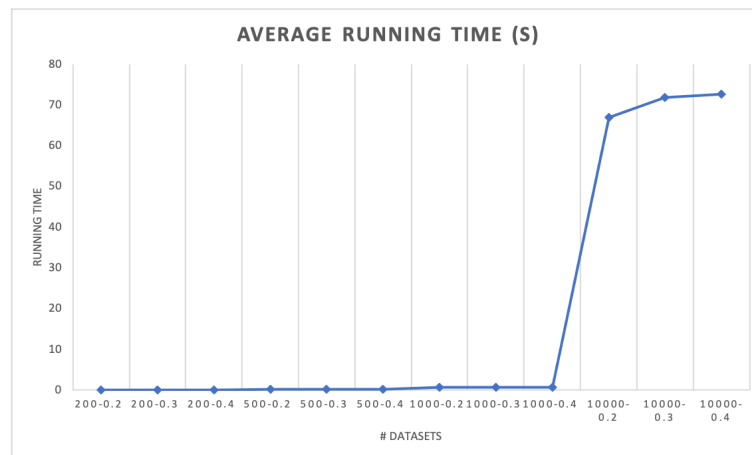
FIGURE 4.1: Average running time of the FBSID algorithm
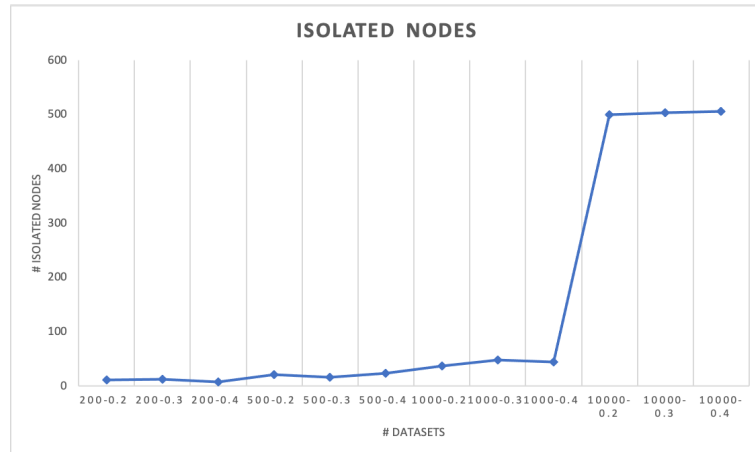
FIGURE 4.2: The number of isolated nodes in FBSID Algorithm

Based on the results, it can be observed that the results of running our proposed FBSID algorithm have less dependency on the complexity than the number of nodes in a network. For example, as can be seen in Table 4.1, the number of isolated nodes for a network with 200 nodes in various complexities ($\mu = 0.2$, 0.3, 0.4) are almost in the same range.

However, by changing the size of each network, the number of isolated nodes are increasing. In fact, the results of this algorithm has a direct relation to the number of nodes in a graph, and it does not depend on the number of edges. In addition, based on the Figure 4.1, the running time and the size of each graph has a direct relation. Consequently, larger size graphs need more time to be processed by the algorithm.

4.2.2   Structure-based Social Isolation Detection

We believe that, by increasing the number of messages, the results will be more accurate. However, increasing the number of message will increase the run-time of our algorithm. In order to find a trade-off between the number of messages and the required processing time, we conducted a series of experiments to compare the run-time and the results of our algorithm by changing the number of messages from 100 to 1000. Consequently, we have conducted 48 independent experiments, and repeated each of them 5 times in order to obtain the average values of the results. To conduct

these experiments, 4 different message settings (i.e., 100, 250, 500, and 1000) were used to examine the scalability and the reliability of the SBSID algorithm.

The obtained results are shown in Tables 4.3, 4.4, 4.5, and 4.6. We can observe that in the SBSID algorithm, the number of isolated nodes is independent from the network's complexity. However, it has a direct relation with the size of a graph. Furthermore, increasing the number of nodes in the graphs and the number of messages in the algorithm, led to longer run-times in each experiments.

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

TABLE 4.3: The results obtained from SBSID algorithm - #Msg = 100

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

TABLE 4.4: The results obtained from SBSID algorithm - #Msg = 250

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

TABLE 4.5: The results obtained from SBSID algorithm - #Msg = 500

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 ± 83.494	1930 ± 8
10000-0.3	1253.374 ± 82.519	1916 ± 16
10000-0.4	1463.000 ± 75.718	1940 ± 16

TABLE 4.6: The results obtained from SBSID algorithm - #Msg = 1000

As shown in the Figures 4.3, 4.4, 4.5, 4.6, we have fewer isolated nodes by sending more messages in a network. For example, in Figure 4.3, when N = 200 and $\mu = 0.2$, after sending 100 messages, 55 isolated nodes were detected, while after increasing the number of messages to 1000, only 39 nodes were identified as isolated ones. According to the results, the message size of 250 can be selected as a trade-off value, because there are negligible changes in the number of isolated nodes for the message numbers 500 and 1000.

FIGURE 4.3: The number of isolated nodes in various message settings (nodes = 200)

FIGURE 4.4: me
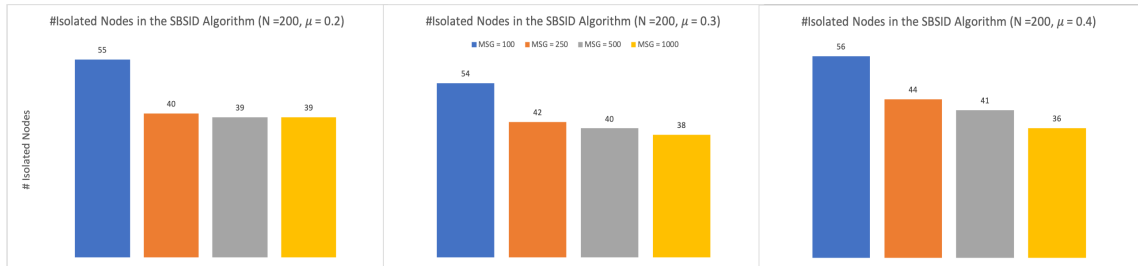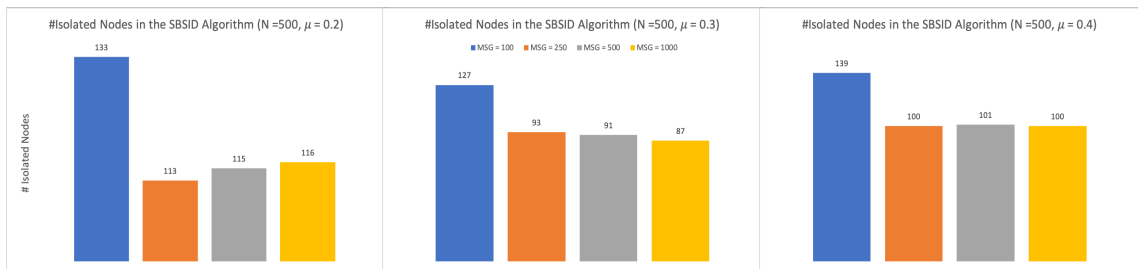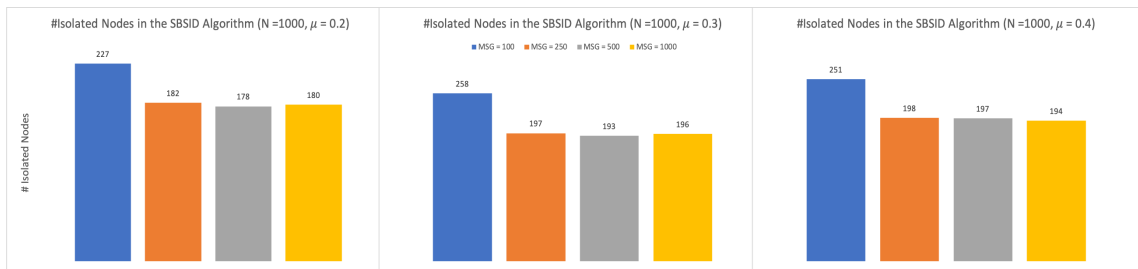

FIGURE 4.5: The number of isolated nodes in various message settings (nodes = 1000)
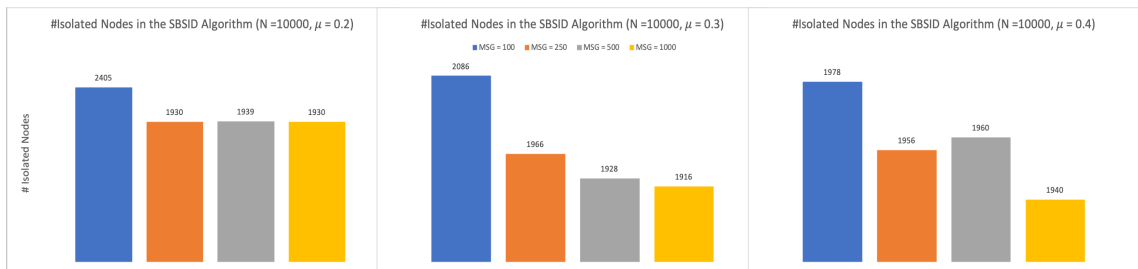
FIGURE 4.6: The number of isolated nodes in various message settings (nodes = 10000)

Similarity Analysis

As mentioned before, we have repeated each experiment 5 times to get the average values for the result. Therefore, we considered two similarity indexes, Jaccard and

the Overlap coefficient, to measure the similarities among the answer sets. We have randomly chosen a pairs of obtained results and calculated their similarities based on the Jaccard and the Overlap coefficient similarity indexes. Tables 4.7, 4.8, 4.9, and 4.10 show the results from similarity comparison among answer sets for different message settings. Figures 4.10, and 4.11 also illustrate the Jaccard and overlap coefficient similarity measurements.

The results confirm that by increasing the number of messages, the reliability of the results will be increased. For example, when the number of messages is 100, the Jaccard similarity is around 50% and the overlap coefficient is around 65%. However, when the message number is 1000, these values increased to around 80% and 90%, accordingly. Overall, as can be seen in the figures, by increasing the number of messages, the similarity among answers has been increased.

For example, as shown in Figure 4.7 ,in a network with 200 nodes, $\mu = 0.2$, after sending 100 messages, S_1 and S_2 are two answer sets which selected randomly to compare with each other with the size of 56 and 60. The list of mutual nodes shows the 39 common nodes between S_1, and S_2. The Jaccard and the overlap coefficient similarities between these two sets are almost %50 and %70, respectively.

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

FIGURE 4.7: A sample output from the SBSID algorithm - MSG = 100

In the network with the same structure (200 nodes, $\mu = 0.2$),as shown in Figure 4.8, after sending 250 messages, S_3 with 44 nodes and S_4 with 40 nodes are the isolated nodes selected to compare with each other, randomly. The list of mutual

nodes shows the 36 common nodes between S_3 and S_4. Accordingly, the Jaccard and the overlap coefficient similarity indexes for these two sets are near %70 and %80. It is notable that similarity indexes are increased by sending 250 messages instead of 100 messages.

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

FIGURE 4.8: A sample output from the SBSID algorithm - MSG = 250

With considering the same network (N= 200, $\mu = 0.2$), as shown in Figure 4.9, after sending 1000 messages, S_5 with 40 and S_6 with 39 isolated nodes are selected to compare with each other, randomly. The list of mutual nodes shows the 36 common nodes between S_5 and S_6. Consequently, after increasing the number of messages, the Jaccard and the overlap coefficient similarities are also increased and reached %85 and %90.

The obtained results show that the overlap coefficient index has higher values than the Jaccard index. The Jaccard similarity is a very reliable and powerful index, but it has a significant disadvantage when we compare sets with different sizes [58]. It means that the use of the overlap coefficient similarity is more relevant than the Jaccard index in this study.

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

FIGURE 4.9: A sample output from the SBSID algorithm - MSG = 1000

Message Number = 100				
Dataset	Jaccard Max	Jaccard Mean	Overlap Coefficient Max	Overlap Coefficient Mean
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

TABLE 4.7: The SBSID Algorithm Comparison Results (Msg = 100)

In addition, the figure 4.12 represents the average running time of experiments on 12 datasets with various message sizes. We observed that the run-time for each experiment directly relates to the number of messages in each experiment. It means that by increasing the number of messages(100, 250, 500, 1000), the response time for detecting isolated nodes is also longer.

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

TABLE 4.8: The SBSID Algorithm Comparison Results (Msg = 250)

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

TABLE 4.9: The SBSID Algorithm Comparison Results (Msg = 500)

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

TABLE 4.10: The SBSID Algorithm Comparison Results (Msg = 1000)
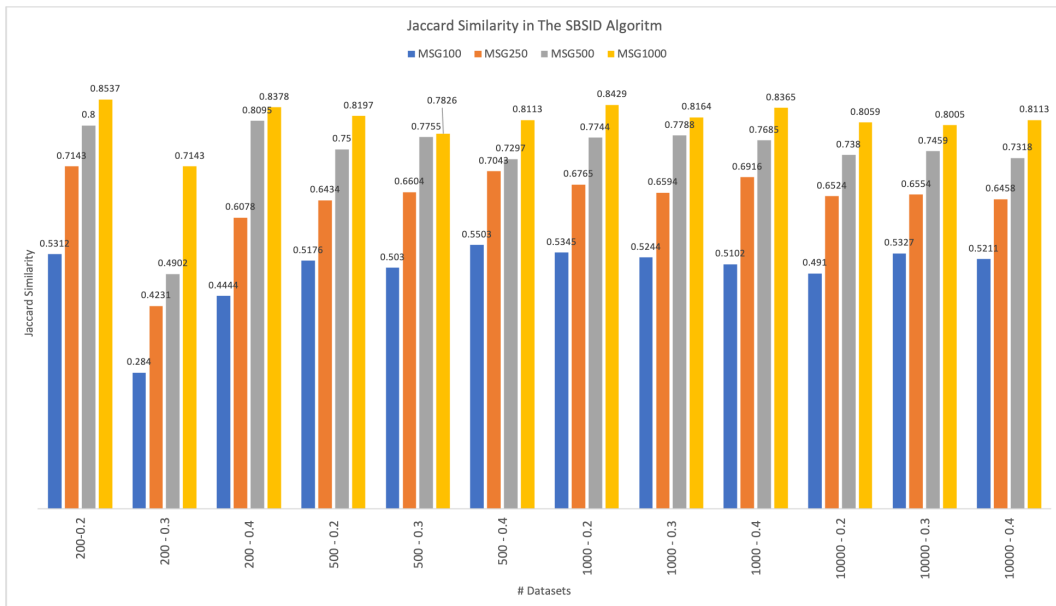

FIGURE 4.10: Jaccard Similarity Comparison in SBSID ALgorithm
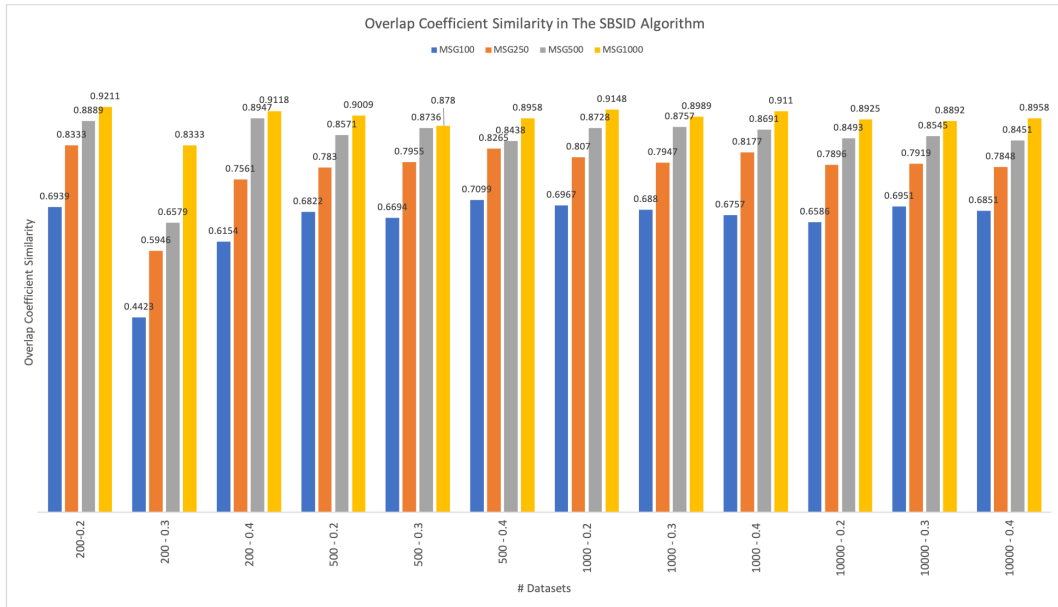
FIGURE 4.11: Overlap Coefficient Similarity Comparison in SBSID Algorithm
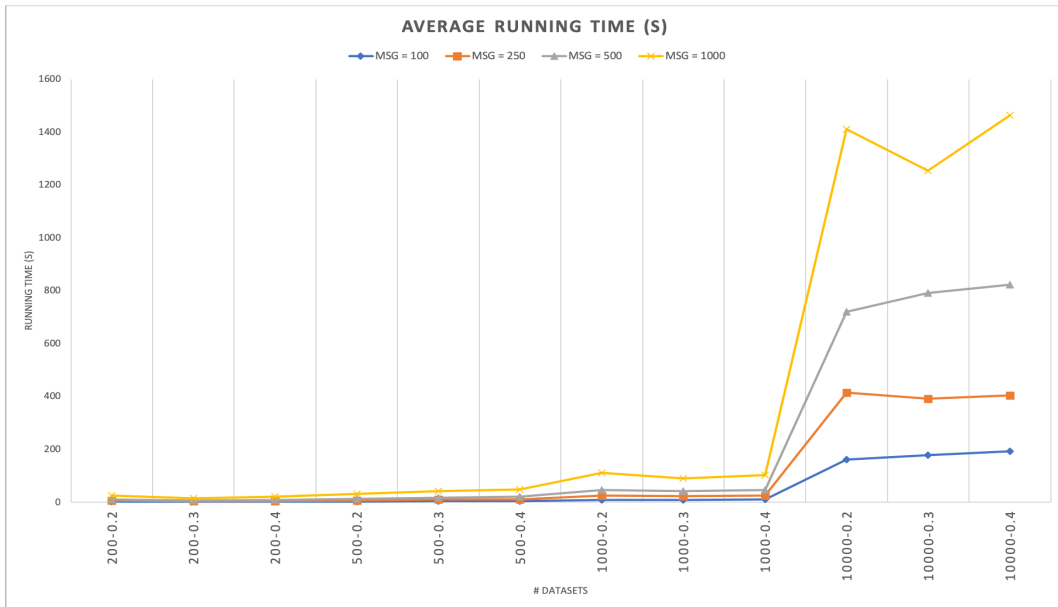
FIGURE 4.12: The results for the average Running Time based on the number of messages in the SBSID algorithm

4.3 Comparison with the Exact Method

In this part, we have implemented another algorithm called Exact algorithm. We run the Exact algorithm on 48 independent synthetic datasets and repeated each of them 10 times to compare the achieved results with the obtained results from the SBSID algorithm. In the Exact algorithm, a certain portion of messages (100, 250, 500, 1000) is sent to all the nodes in the graph without considering the probability of receiving messages. Therefore, all the existing nodes in a network will be receiving the same amount of messages. In this way, the exact algorithm detects all the potential socially isolated nodes.

The obtained results of the Exact algorithm are illustrated in Tables 4.11, 4.12, 4.13, and 4.14, which show that the number of isolated nodes in the Exact algorithm has a direct relation with the size of the network. It is also notable that by increasing the size of networks, the running time for each experiment, and the number of detected nodes as isolated are increased.

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

TABLE 4.11: The Exact Algorithm Final Result - #Msg = 100

As shown in Figures 4.13, 4.14, 4.15, and 4.16, the number of isolated nodes and various message settings are not related in the exact algorithm. Moreover, the number of identified nodes as isolated ones in the exact algorithm is almost identical to each

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

TABLE 4.12: The Exact Algorithm Final Result - #Msg = 250

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

TABLE 4.13: The Exact Algorithm Final Result - #Msg = 500

network. For instance, according to Figure 4.13, in networks with $\mu = 0.2$, $\mu = 0.3$, and $\mu = 0.4$, 46, 44, and about 38 isolated nodes were identified in various message sizes, respectively.

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

TABLE 4.14: The Exact Algorithm Final Result - #Msg = 1000

FIGURE 4.13: The number of isolated nodes in various message settings (nodes = 200)
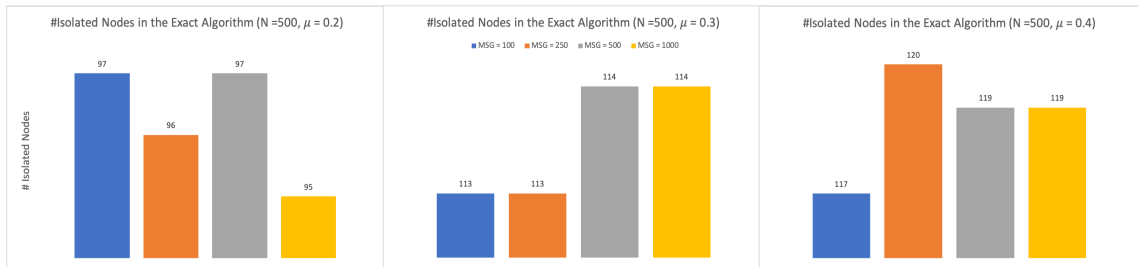
FIGURE 4.14: The number of isolated nodes in various message settings (nodes = 500)
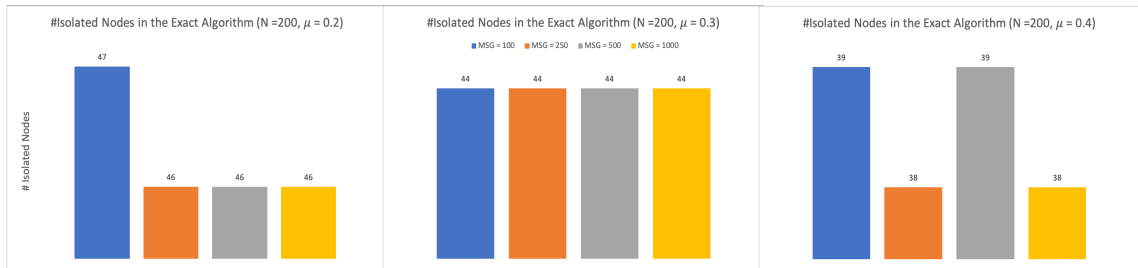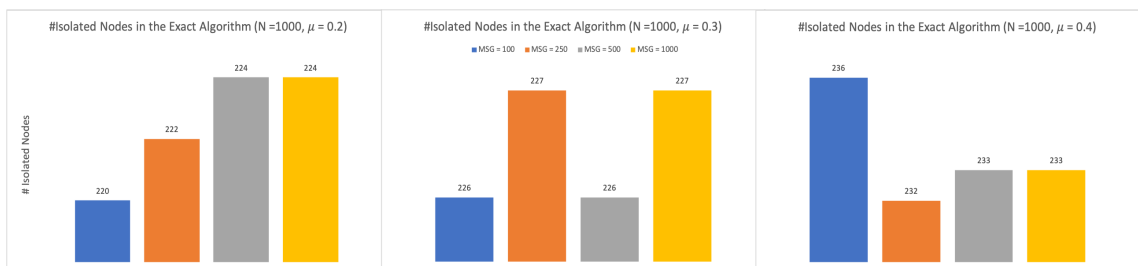

FIGURE 4.15: The number of isolated nodes in various message settings (nodes = 1000)
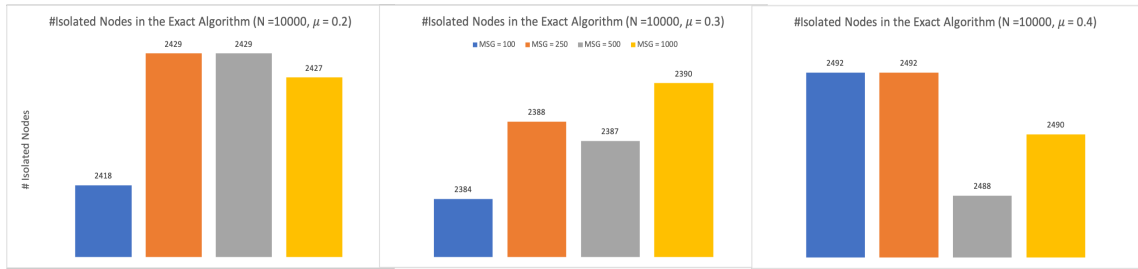
FIGURE 4.16: The number of isolated nodes in various message settings (nodes = 10000)

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

TABLE 4.15: The Exact Algorithm Comparison Results - #Msg = 100

In order to compare the results of our algorithm with the Exact method, we have analyzed their results using the following procedure. The output of the exact algorithm has been compared to all the obtained results from the SBSID algorithm. In the nest step, the best answer with maximum value and also the average value for the similarity indexes have been reported. For instance, for a network with 200 nodes, $\mu = 0.2$, and 100 messages, we obtained 5 answer sets, S_1 to S_5, as shown in Figure 4.17.

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

TABLE 4.16: The Exact Algorithm Comparison Results - #Msg = 250

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

TABLE 4.17: The Exact Algorithm Comparison Results - #Msg = 500

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

TABLE 4.18: The Exact Algorithm Comparison Results - #Msg = 1000

S_1 = 2,3, 4, 5, 7,10, 11, 12, 15, 16, 19, 21, 23, 24, 25, 26, 30, 32, 34, 35, 36, 38, 43, 44, 46, 47, 48, 52, 53, 54, 58, 59, 61, 68, 69, 76, 79, 80, 81, 82, 83, 86, 89, 91, 95, 109, 133, 135, 138, 143, 146, 151, 152, 153, 156, 169

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

S_3 = 3, 4, 5, 7, 9, 11, 15, 16, 23, 24, 25, 28, 30, 31, 32, 34, 35, 36, 37, 40, 44, 46, 48, 51, 53, 64, 67, 68, 69, 71, 72, 73, 76, 80, 81, 82, 89, 95, 101, 109, 118, 122, 133, 138, 143, 151, 153, 169, 178

S_4 = 2, 4, 5, 6, 7, 9, 10, 11, 15, 16, 17, 21, 23, 24, 25, 26, 28, 29, 30, 31, 33, 35, 36, 40, 44, 46, 51, 53, 54, 58, 59, 61, 67, 69, 76, 80, 81, 82, 86, 87, 89, 91, 92, 101, 109, 114, 144, 151, 152, 153, 156, 159, 169, 178, 199

S_5 = 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

Exact_Set = 2, 3, 4, 6, 7, 10, 11, 12, 15, 16, 21, 23, 24, 25, 26, 27, 30, 31, 32, 34, 35, 36, 37, 40, 46, 53, 58, 59, 60, 61, 62, 68, 69, 76, 81, 82, 86, 91, 102, 109, 133, 143, 151, 153, 169, 178, 199

FIGURE 4.17: A sample output from the Exact algorithm

The average of Jaccard and the overlap coefficient similarity, as can be seen in table 4.16, were about %47 and %64. The maximum similarity was between S_1 with

56 nodes and the Exact Set where the Jaccard index was about %55, and the overlap coefficient index was about %70.

Figures 4.18, and 4.19 demonstrate the similarity indexes (Jaccard, Overlap coefficient) for the obtained results from the Exact algorithm. Although due to the use of the synthetic data, at some points, a sharp difference among results can be seen, most of the points are in the same range.

Figure 4.20 shows the running time for the Exact algorithm. The running time has a direct relation to the size of each network. Therefore, as indicated, a bigger network needs a longer time to detect isolated nodes. In addition, the running time for networks of the same size is almost the same because it is independent of the number of messages.
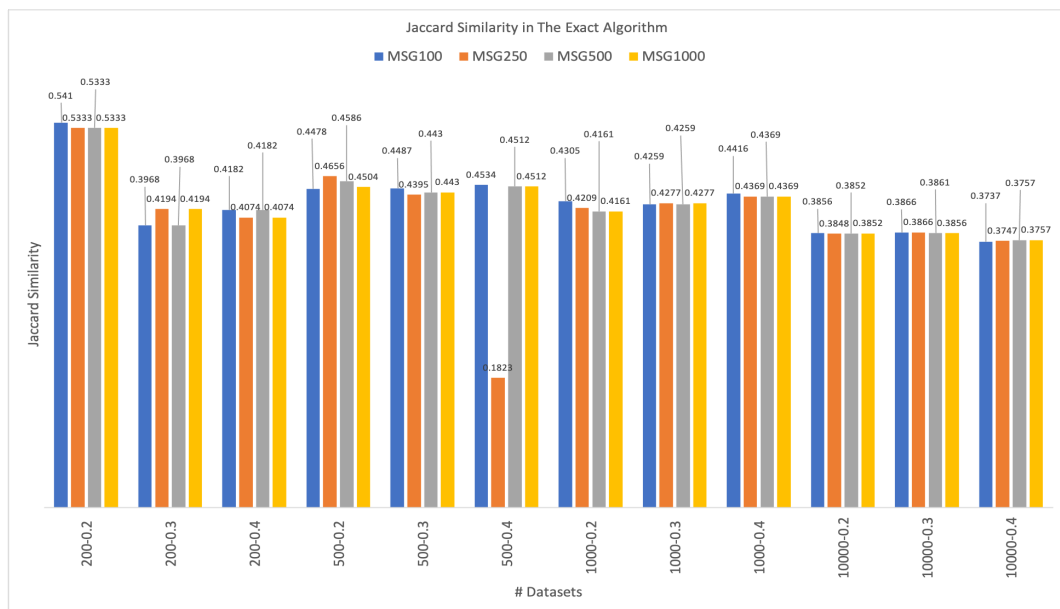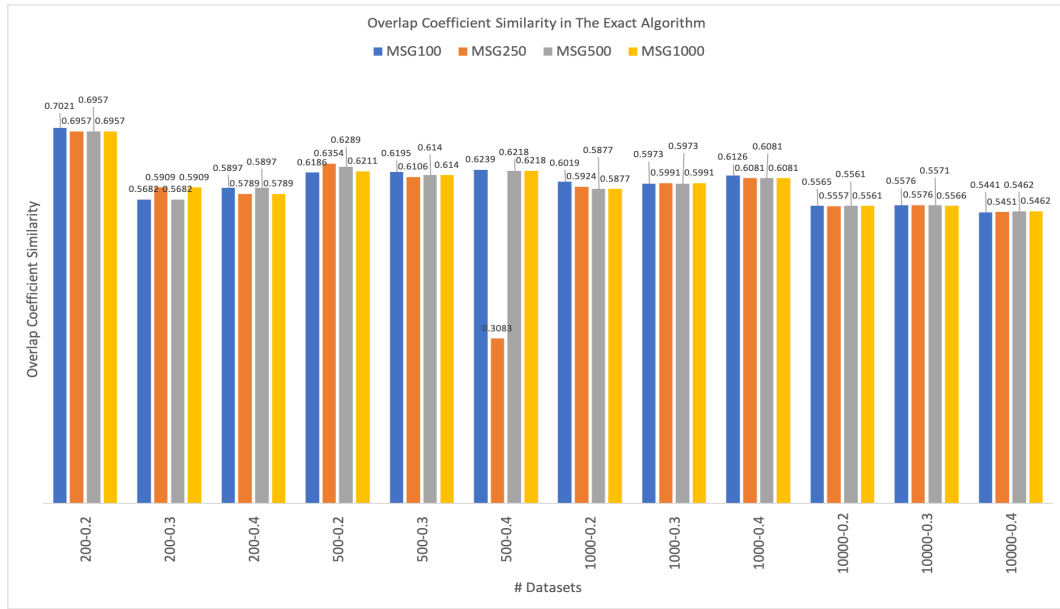

FIGURE 4.18: Jaccard Similarity Comparison

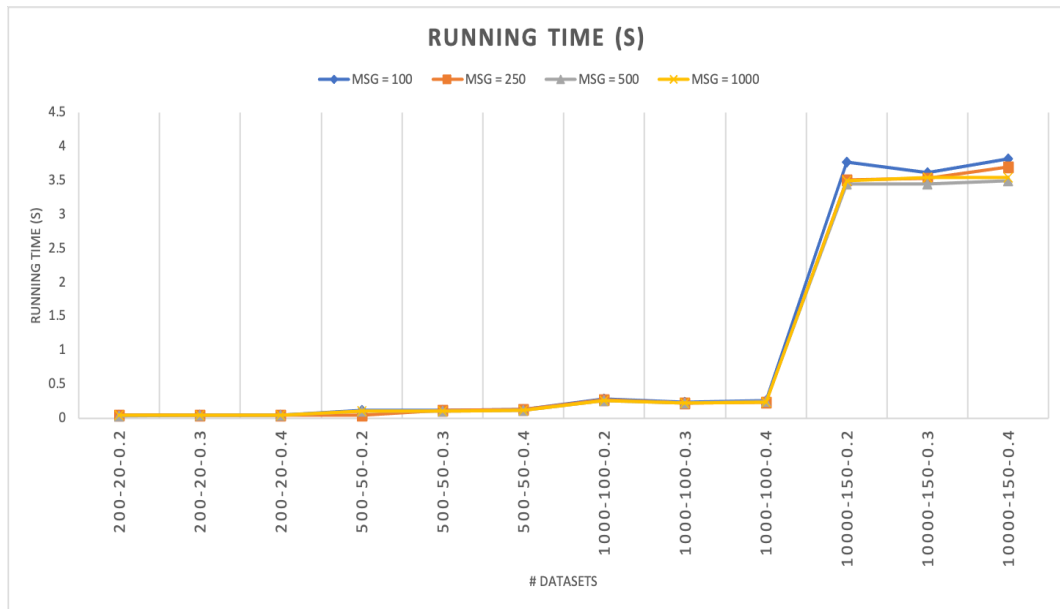FIGURE 4.19: Overlap Coefficient Similarity Comparison

FIGURE 4.20: Response Time based on the number of messages

4.4 Discussion

In this research study, we proposed a novel computational model to detect socially isolated individuals in a social network. Two algorithms have been proposed in this dissertation.

The FBSID algorithm deals with individuals' characteristics, and the SBSID algorithm is a message-passing algorithm that focuses on the relationship among individuals in a social network. To the best of our knowledge, there are no other computational models; therefore, we proposed another algorithm to compare the SBSID algorithm with it named Exact algorithm.

We have done several experiments with various sizes and complexities to test our model's performance, reliability, and accuracy. Based on the obtained results and similarity comparison among outputs, it can be concluded that the exact algorithm detects potentially isolated nodes, which means it identifies both real isolated nodes and nodes at the risk of isolation. On the other hand, the SBSID algorithm filters the Exact algorithm and detects fewer isolated nodes. It means the SBSID algorithm only detects real isolated nodes. Although the SBSID algorithm has a longer running time than the Exact algorithm, the results obtained from it have higher accuracy. The obtained accuracy for the SBSID algorithm was about %75, while the Exact algorithm reached almost %70 for the accuracy.

In addition, we tested the SBSID on a various number of messages (i.e., 100, 250, 500, 1000). Based on achieved results, the run-time for the SBSID algorithm directly relates to the number of messages. It means that by increasing the number of messages, the run-time will be longer. However, after analyzing the obtained results, it can be stated that message size 250 is the trade-off between the number of messages and the required processing time. In this way, the run-time for the SBSID algorithm would be shorter while it has reliable performance.

Chapter 5

Conclusion

Social Isolation is harmful to healthy human development. Socially isolated people in this research are considered as those individuals who have no tendency to communicate with other individuals in the networks or, they are not able to communicate in an efficient way while they may have many connections, or can not establish a connection with others due to the characteristics they have. In this thesis, we have proposed a novel computational model and two algorithms (FBSID and SBSID) to identify socially isolated individuals in social networks.

Our approach is to map a given social network to a feature-based weighted directed social graph, where each node represents an individual with a set of characteristics. Edges represent existence of relationships between individuals, and the weight of each edge shows the quality and the strength of these relationships. Therefore, we have proposed a computational model based on the message-passing technique to extract information from the structure of the social graph to identify the norms of society and calculate the distance of each node with the society's norm to identify the isolated nodes. Consequently, we proposed a couple of metrics to measure the quantity and quality of relationships among the individuals in a given network by analyzing the characteristics of a node in a graph, the number of connections for each node, and each node's tendency to communicate with others. As a result, the isolated individual

can be seen as an outlier node that behaves differently compared to the rest of the nodes.

We have also presented an algorithm to identify outliers in attribute-based social graphs. To capture each individual's demographic characteristics, we used an array of binary attributes and developed a method to identify the outlier nodes. These arrays then are used to extract the norm of the society. The attributes of each node is then compared with the norm in order to detect the outlier nodes.

We have evaluated our proposed model on various types of synthetic social networks with different sizes and configurations. We have conducted multiple experiments to analyze the accuracy, efficiency and reliability of our proposed model. The obtained results have shown that our model can identify the isolated nodes with around %75 accuracy where compared to the exact method. The results also show that our model is scalable and capable to identify isolated people in large scale social networks in a linear time complexity.

5.1 Future Works

The performance of our method was tested on synthetic social networks generated by the LFR benchmark. However, in order to precisely and comprehensively analyze the accuracy and efficiency of this method, it should be tested on real data. As a part of our future work, we aim to test our proposed algorithms on real data in more details.

Moreover, as we deal with real data in a real-world implementation, the privacy risks for this model should be analyzed. The input data for the FBSID algorithm is a list of individuals' demographic characteristics; for the SBSID algorithm, information about each person's social relationships is essential, and the output for both algorithms is a list of socially isolated individuals. Therefore, risk factors in terms of sensitive data exposure should be analyzed and considered in the future because this algorithm will be implemented on health data.

Appendices

Evaluation With Real Data

This chapter contains an additional way we used to evaluate the performance of the proposed model. In this part, we used real data to evaluate our model.

Data Collection

This research was conducted on ten patients ranging from age 50 to 90 years. The population sample was drawn from the *Hospice of Windsor and Essex County Inc.* We analyzed ten booklets that belong to each patient to extract the required information we needed for this study. To collect the essential data for the SBSID algorithm, at first, we created a list of all individuals who had a connection with the patient. In this way, we had ten separate networks for each patient. To assign a weight to each link that connects patients to their relatives or friends, we defined three categories to analyze the frequency, quality, and variety of relationships.

- Frequency represents the amount of time two people are spending with each other.

- Quality includes three classes of family members, distant relatives together with friends, and medical supports to compute the quality of each connection based on the status of the relation and prioritize the closest individuals, Table 1.

- Variety of each connection is defined as the division of the total number of connections in quality classes and the total number of links. For example, the information of a patient is shown in Table 2. This patient has 15 connections that are weighted based on their quality. The patient has five connections with immediate family members (quality_id = 1), six connections with distant relatives and friends (quality_id = 2), and four connections with physicians and medical supporters (quality_id = 3), Table 3.

ID	Relation Status	Weight
1	Immediate Family Members	1
2	Distant Relatives + Friends	0.75
3	Medical Supports + Other	0.5

TABLE 1: Quality Table

No.	ID	Relation	Quality
1	100	Husband	1
2	101	Son	1
3	102	Daughter	1
4	103	Sister#1	1
5	104	Sister#2	1
6	105	Friend#1	2
7	106	Friend#2	2
8	107	Friend#3	2
9	108	Friend#4	2
10	109	Priest	2
11	110	Sister-in-law	2
12	111	ALSO(PSW) [1]	3
13	112	Therapist	3
14	113	Care Provider	3
15	114	Pharmacist	3

TABLE 2: A sample of a patient's connections information

The final dataset for the aforementioned patient is shown in Table 4. All three categories have a direct relationship with the final weight value; therefore, the three categories' values multiplication would be the final weight for the graph's links. The values for the final weight were normalized to be in a range of zero and one. One of

---

[1] Assisted Living Southwestern Ontario (Personal Support Worker)

Quality_ID	Variety	Weight
1	5	$5/15 = 0.34$
2	6	$6/15 = 0.4$
3	4	$4/15 = 0.27$

TABLE 3: Variety Table

the main restrictions we had was a lack of enough information to create a weighted directed graph. Therefore, we used the same weights for other people's relations with patients to have a directed graph. For example, the weight of the link between node #3 and node #100 is 1, and we considered the same weight for the link between node #100 and #3.

Patient_ID	Relation_ID	Frequency	Quality	Variety	Final Weight
3	100	1	1	0.34	1
3	101	1	1	0.34	1
3	102	1	1	0.34	1
3	103	0	1	0.34	0
3	104	0.07	1	0.34	0.07
3	105	0	0.75	0.4	0
3	106	0.14	0.75	0.4	0.12
3	107	0.14	0.75	0.4	0.12
3	108	0.07	0.75	0.4	0.06
3	109	0.21	0.5	0.27	0.08
3	110	0.21	0.5	0.27	0.08
3	111	0.14	0.5	0.27	0.06
3	112	0.24	0.5	0.27	0.1
3	113	0.02	0.75	0.4	0.02
3	114	0.26	0.75	0.4	0.23

TABLE 4: A sample of the weighted-directed dataset

Patient_ID	Age	Gender	Living Alone	Marital Status	Social Support	Serious illnesses	Education	Job Status	Physical Disability	Technology Skills
3	0	1	0	1	1	1	1	1	1	0
5	1	1	1	0	1	0	1	1	0	0
6	1	0	1	0	0	0	0	1	1	0
7	1	1	1	0	0	0	0	1	0	0
8	0	1	0	1	1	1	0	0	1	1
9	0	1	1	0	1	1	1	1	1	1
11	1	1	0	0	0	1	0	0	1	0
13	0	1	1	0	1	0	0	0	1	1
14	0	1	0	1	0	1	0	1	0	1
18	1	0	1	0	0	1	1	1	1	1

TABLE 5: The final dataset for the FBSID algorithm

Also, for the FBSID algorithm implementation, we needed an attribute-based dataset. Based on patients' narratives, we selected ten characteristics to generate a dataset shown in Table 5. We considered these characteristics that intensify isolation among individuals. In this way, if the value for these features is one, it means it has a negative effect on a person; otherwise, the value would be zero. To convert age to binary values, ages higher than 65 were considered 1 and less than 65 as 0. For the gender column, we consider females as 1 and males as 0. If the value for living alone, marital status, and social support are 0, it means that patients are not living alone, single, and do not have any social support; otherwise, the value is 1. We considered multiple sclerosis (MS), diabetes, Alzheimer's, and dementia as serious illnesses. Consequently, if they have one of these diseases, their status would be 1 and 0, otherwise. If they were educated or employed in the past, their status is 1 and 0 if they were not. If a patient has any functional or mobility impairment, the status is 1 and 0, otherwise. Technology can be a very effective tool in reducing social isolation and improving connectedness, such as mobile technologies, internet connection, videoconferencing, and digital games. Accordingly, if their status is 1 if they have technology skills and 0 otherwise.

In this research study, we tried to evaluate our model with real data, which was more reliable. Unfortunately, we did not have access to enough information to be able to evaluate the proposed model by using real data. Therefore we used synthetic data for the evaluation part.

Bibliography

[1] A profile of persons with disabilities among canadians aged 15 years or older, 2012. `https://www150.statcan.gc.ca/n1/pub/89-654-x/89-654-x2015001-eng.htm`. (Accessed on 11/15/2020).

[2] Quality of personal networks: Does living alone make a difference? `https://www150.statcan.gc.ca/n1/pub/11-008-x/2011002/article/11591-eng.htm`. (Accessed on 11/15/2020).

[3] Thomas KM Cudjoe, David L Roth, Sarah L Szanton, Jennifer L Wolff, Cynthia M Boyd, and Roland J Thorpe Jr. The epidemiology of social isolation: National health and aging trends study. *The Journals of Gerontology: Series B*, 75(1):107–113, 2020.

[4] Matthew Pantell, David Rehkopf, Douglas Jutte, S Leonard Syme, John Balmes, and Nancy Adler. Social isolation: a predictor of mortality comparable to traditional clinical risk factors. *American journal of public health*, 103(11):2056–2062, 2013.

[5] Brian A Primack, Ariel Shensa, Jaime E Sidani, Erin O Whaite, Liu yi Lin, Daniel Rosen, Jason B Colditz, Ana Radovic, and Elizabeth Miller. Social media use and perceived social isolation among young adults in the us. *American journal of preventive medicine*, 53(1):1–8, 2017.

[6] Razieh Bandari, Hamid Reza Khankeh, Farahnaz Mohammadi Shahboulaghi, Abbas Ebadi, Abbas Ali Keshtkar, and Ali Montazeri. Defining loneliness in

older adults: protocol for a systematic review. *Systematic reviews*, 8(1):1–6, 2019.

[7] Su-Hyun Han, Keun Tae Kim, Han Uk Ryu, Sang-Ahm Lee, Yang-Je Cho, Jee Hyun Kim, Kyung-Wook Kang, Dong Jin Shin, Gha Hyun Lee, Kyoung Jin Hwang, et al. Factors associated with social anxiety in south korean adults with epilepsy. *Epilepsy & Behavior*, 101:106569, 2019.

[8] Yu Mao, Xinian Zuo, Cody Ding, and Jiang Qiu. Ofc and its connectivity with amygdala as predictors for future social anxiety in adolescents. *Developmental Cognitive Neuroscience*, page 100804, 2020.

[9] Samsi Hasan and Nurmi Yulianti. Introversion personality and students' reading comprehension. *Indonesian Journal of Integrated English Language Teaching*, 4 (2):2018–229, 2019.

[10] Verena Enneking, Pia Krüssel, Dario Zaremba, Katharina Dohm, Dominik Grotegerd, Katharina Förster, Susanne Meinert, Christian Bürger, Fanni Dzvonyar, Elisabeth J Leehr, et al. Social anhedonia in major depressive disorder: a symptom-specific neuroimaging approach. *Neuropsychopharmacology*, 44(5): 883–889, 2019.

[11] Patrick O'Donnell, Diarmuid O'Donovan, and Khalifa Elmusharaf. Measuring social exclusion in healthcare settings: a scoping review. *International journal for equity in health*, 17(1):15, 2018.

[12] Gökmen Arslan. Mediating role of the self–esteem and resilience in the association between social exclusion and life satisfaction among adolescents. *Personality and Individual Differences*, 151:109514, 2019.

[13] National Seniors Council. Ageing, 2020. URL https://www.un.org/en/sections/issues-depth/ageing/.

[14] Emilie Courtin and Martin Knapp. Social isolation, loneliness and health in old

age: a scoping review. *Health & social care in the community*, 25(3):799–812, 2017.

[15] Ziggi Ivan Santini, Paul E Jose, Erin York Cornwell, Ai Koyanagi, Line Nielsen, Carsten Hinrichsen, Charlotte Meilstrup, Katrine R Madsen, and Vibeke Koushede. Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older americans (nshap): a longitudinal mediation analysis. *The Lancet Public Health*, 5(1):e62–e70, 2020.

[16] Oliver Hämmig. Health risks associated with social isolation in general and in young, middle and old age. *PLoS One*, 14(7):e0219663, 2019.

[17] National Seniors Council. Report on the social isolation of seniors, 2014.

[18] Roser Beneito-Montagut, Nizaiá Cassián-Yde, and Arantza Begueria. What do we know about the relationship between internet-mediated interaction and social isolation and loneliness in later life? *Quality in Ageing and Older Adults*, 2018.

[19] Christina R Victor, Vanessa Burholt, and Wendy Martin. Loneliness and ethnic minority elders in great britain: an exploratory study. *Journal of cross-cultural gerontology*, 27(1):65–78, 2012.

[20] Alain Topor, Ingemar Ljungqvist, and Eva-Lena Strandberg. The costs of friendship: Severe mental illness, poverty and social isolation. *Psychosis*, 8(4):336–345, 2016.

[21] Debra Umberson and Jennifer Karas Montez. Social relationships and health: A flashpoint for health policy. *Journal of health and social behavior*, 51(1_suppl): S54–S66, 2010.

[22] Natalie Cotterell, Tine Buffel, and Christopher Phillipson. Preventing social isolation in older people. *Maturitas*, 113:80–84, 2018.

[23] Matthew Lee Smith, Lesley E Steinman, and EA Casey. Combatting social isolation among older adults in a time of physical distancing: the covid-19 social connectivity paradox. *Frontiers in public health*, 8:403, 2020.

[24] Takahiro A Kato, Norman Sartorius, and Naotaka Shinfuku. Forced social isolation due to covid-19 and consequent mental health problems: Lessons from hikikomori. *Psychiatry and clinical neurosciences*, 2020.

[25] Lenka Skanderova, Tomas Fabian, and Ivan Zelinka. Small-world hidden in differential evolution. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 3354–3361. IEEE, 2016.

[26] Pooya Moradian Zadeh and Ziad Kobti. A multi-population cultural algorithm for community detection in social networks. In *ANT/SEIT*, pages 342–349, 2015.

[27] Yousra Asim, Ahmad Kamran Malik, Basit Raza, and Ahmad Raza Shahid. A trust model for analysis of trust, influence and their relationship in social network communities. *Telematics and Informatics*, 36:94–116, 2019.

[28] Xiaohui Pan, Guiqiong Xu, Bing Wang, and Tao Zhang. A novel community detection algorithm based on local similarity of clustering coefficient in social networks. *IEEE Access*, 7:121586–121598, 2019.

[29] Andrew T Stephen and Olivier Toubia. Explaining the power-law degree distribution in a social commerce network. *Social Networks*, 31(4):262–270, 2009.

[30] Francesco Martino and Andrea Spoto. Social network analysis: A brief theoretical review and further perspectives in the study of information technology. *PsychNology Journal*, 4(1):53–86, 2006.

[31] Rachel Isba, Katherine Woolf, and Robert Hanneman. Social network analysis in medical education. *Medical education*, 51(1):81–88, 2017.

[32] Filipa Landeiro, José Leal, and Alastair M Gray. The impact of social isolation on delayed hospital discharges of older hip fracture patients and associated costs. *Osteoporosis international*, 27(2):737–745, 2016.

[33] Takaaki Ikeda, Noriko Cable, Masashige Saito, Shihoko Koyama, Taishi Tsuji, Taiji Noguchi, Katsunori Kondo, Ken Osaka, and Jun Aida. Association between

social isolation and smoking in japan and england. *Journal of Epidemiology*, page JE20200138, 2020.

[34] Angela K Perone, Berit Ingersoll-Dayton, and Keisha Watkins-Dukhie. Social isolation loneliness among lgbt older adults: Lessons learned from a pilot friendly caller program. *Clinical Social Work Journal*, 48(1):126–139, 2020.

[35] Stephen J Macdonald, Lesley Deacon, Jackie Nixon, Abisope Akintola, Anna Gillingham, Jacqueline Kent, Gillian Ellis, Debbie Mathews, Abolaji Ismail, Sylvia Sullivan, et al. 'the invisible enemy': disability, loneliness and isolation. *Disability & Society*, 33(7):1138–1159, 2018.

[36] Danielle Kelly, Artur Steiner, Micaela Mazzei, and Rachel Baker. Filling a void? the role of social enterprise in addressing social isolation and loneliness in rural communities. *Journal of rural studies*, 70:225–236, 2019.

[37] Nicola Petersen, Hans-Helmut König, and André Hajek. The link between falls, social isolation and loneliness: a systematic review. *Archives of gerontology and geriatrics*, 88:104020, 2020.

[38] Julianne Holt-Lunstad, Timothy B Smith, and J Bradley Layton. Social relationships and mortality risk: a meta-analytic review. *PLoS medicine*, 7(7):e1000316, 2010.

[39] Raffaella Calati, Chiara Ferrari, Marie Brittner, Osmano Oasi, Emilie Olié, André F Carvalho, and Philippe Courtet. Suicidal thoughts and behaviors and social isolation: A narrative review of the literature. *Journal of affective disorders*, 245:653–667, 2019.

[40] Karen Windle, Jennifer Francis, Caroline Coomber, et al. *Preventing loneliness and social isolation: interventions and outcomes*. Social Care Institute for Excellence London, 2011.

[41] Joyce Siette, Megan Cassidy, and Stefan Priebe. Effectiveness of befriending interventions: a systematic review and meta-analysis. *BMJ open*, 7(4), 2017.

[42] Andy P Dickens, Suzanne H Richards, Colin J Greaves, and John L Campbell. Interventions targeting social isolation in older people: a systematic review. *BMC public health*, 11(1):647, 2011.

[43] A Wigfield, S Alden, and E Kispeter. Age uk's fit for the future social prescribing extension project evaluation report http://www. ageuk. org. uk/documents. *EN-GB/fit% 20as% 20a% 20fiddle/fftf% 20extension% 20project-Final% 20Evaluation% 20Report*, 20, 2016.

[44] Sara J Czaja. The role of technology in supporting social engagement among older adults. *Public Policy & Aging Report*, 27(4):145–148, 2017.

[45] Tine Buffel and Sophie Handler. *Age-friendly cities and communities: A global perspective*. Policy Press, 2018.

[46] Seyed Ebrahim Hosseini and Khaled Goher. Personal care robots for older adults: An overview. 2017.

[47] Jonathan G Shaw, Monica Farid, Claire Noel-Miller, Neesha Joseph, Ari Houser, Steven M Asch, Jay Bhattacharya, and Lynda Flowers. Social isolation and medicare spending: Among older adults, objective isolation increases expenditures while loneliness does not. *Journal of Aging and Health*, 29(7):1119–1143, 2017.

[48] Lourdes Margain, Alberto Ochoa, Teresa Padilla, Saúl González, Jorge Rodas, Odalid Tokudded, and Julio Arreola. Understanding the consequences of social isolation using fireworks algorithm. In *Innovations in Bio-Inspired Computing and Applications*, pages 285–295. Springer, 2016.

[49] Vivienne J Zhu, Leslie A Lenert, Brian E Bunnell, Jihad S Obeid, Melanie Jefferson, and Chanita Hughes Halbert. Automatically identifying social isolation from clinical narratives for patients with prostate cancer. *BMC medical informatics and decision making*, 19(1):1–9, 2019.

[50] Sushant Ranjan and Rama Shankar Yadav. Social isolation: development and validation of measures. *Benchmarking: An International Journal*, 2019.

[51] Alicia Martinez Rebollar, Miguel Gonzalez Mendoza, Hugo Estrada Esquivel, Wilfrido Campos Francisco, and Virginia Campos Ortiz. Automatic detection of social isolation based on human behavior analysis. *Computación y Sistemas*, 24 (4), 2020.

[52] Honglei Zhang, Serkan Kiranyaz, and Moncef Gabbouj. Outlier edge detection using random graph generation models and applications. *Journal of Big Data*, 4 (1):11, 2017.

[53] Xiaowei Xu, Nurcan Yuruk, Zhidan Feng, and Thomas AJ Schweiger. Scan: a structural clustering algorithm for networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 824–833, 2007.

[54] Hiroaki Shiokawa, Yasuhiro Fujiwara, and Makoto Onizuka. Scan++ efficient algorithm for finding clusters, hubs and outliers on large-scale graphs. *Proceedings of the VLDB Endowment*, 8(11):1178–1189, 2015.

[55] Felix R Day, Ken K Ong, and John RB Perry. Elucidating the genetic basis of social interaction and isolation. *Nature communications*, 9(1):1–6, 2018.

[56] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117, 2009.

[57] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4):046110, 2008.

[58] Similarity in graphs: Jaccard versus the overlap coefficient — by brad rees — rapids ai — medium. `https://medium.com/rapids-ai/similarity-in-graphs-jaccard-versus-the-overlap-coefficient-610e083b877d`. (Accessed on 01/26/2021).

Vita Auctoris

NAME:	Saghi Khani
PLACE OF BIRTH:	Esfahan, Iran
YEAR OF BIRTH:	1993
EDUCATION:	Islamic Azad University Najafabad Branch, B.Sc in Information Technology, Najafabad, Esfahan, Iran, 2016
	University of Windsor, M.Sc in Computer Science, Windsor, Ontario, Canada 2021