Research Article

Inferring Ties in Social IoT Using Location-Based Networks and Identification of Hidden Suspicious Ties

Nauman Ali Khan [iD],[1] **Sihai Zhang** [iD],[1] **Wuyang Zhou** [iD],[1] **Ahmad Almogren** [iD],[2] **Ikram Ud Din** [iD],[3] **and Muhammad Asif** [iD][1]

[1]*Key Laboratory of Wireless-Optical Communication, University of Science and Technology China, Hefei 230027, China*
[2]*Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11633, Saudi Arabia*
[3]*Department of Information Technology, The University of Haripur, Haripur 22620, Pakistan*

Correspondence should be addressed to Wuyang Zhou; wyzhou@ustc.edu.cn

Stochastic Internet of Things (IoT)-based communication behavior of the progressing world is tremendously impacting social networks. The growth of social networks helps to quantify the effect on the Social Internet of Things (SIoT). Multiple existences of two persons at several geographical locations in different time frames hint to predict the social connection. We investigate the extent to which social ties between people can be inferred by critically reviewing the social networks. Our study used Chinese telecommunication-based anonymized caller data records (CDRs) and two openly available location-based social network data sets, Brightkite and Gowalla. Our research identified social ties based on mobile communication data and further exploits communication reasons based on geographical location. This paper presents an inference framework that predicts the missing ties as suspicious social connections using pipe and filter architecture-based inference framework. It highlights the secret relationship of users, which does not exist in real data. The proposed framework consists of two major parts. Firstly, users' cooccurrence based on the mutual location in a specific time frame is computed and inferred as social ties. Results are investigated based upon the cooccurrence count, the gap time threshold values, and mutual friend count values. Secondly, the detail about direct connections is collected and cross-related to the inferred results using Precision and Recall evaluation measures. In the later part of the research, we examine the false-positive results methodically by studying the human cooccurrence patterns to identify hidden relationships using a social activity. The outcomes indicate that the proposed approach achieves comprehensive results that further support the theory of suspicious ties.

1. Introduction

A social network is a web of social ties among individuals. Social ties are the kind of one-to-one communication links among humans or Social Internet of Things [1, 2]. Formation of social ties depends upon many attributes, such as the location of living, personality, age, gender, workplace, activities, and many more [3, 4]. These ties are built based on some needs or relationships. People use many mediums for communication such as calls, chatting on social networking websites, reading and writing comments to person, and

reviewing and suggesting some purchasing mobile applications [5]. Investigating human behavior and machine performance, how they react to and participate in social networks remained a center of attention for researchers [2, 6]. Social network analysis is the computer science field that quantifies, evaluates, and analyzes human behavior [7, 8].

A concept of social communication links was proposed by Granovetter [9]. According to their research finding, the communication link between two people is considered as the social tie. Also, each communication link's strength was

further classified as strong or weak depending upon the frequency of communication, number of times, emotional attachment, number of mutual ties, relationship actions, and a combination of these mentioned parameters [5]. Equation (1) quantifies the strength of social ties, which is denoted by weight, such that higher weight tells stronger ties and vice versa [10]. wAB represents the weight of social tie between Node A and Node B, while CA and CB represent the degree of Node A and Node B; cAB is the number of mutual nodes between A and B. Community structures were also found one of the main reasons for social tie strength [11]; it was found that people from the same communities have strong ties as compared to different communities [5]:

$$wAB = \frac{cAB}{CA - 1 + CB - 1 - cAB}. \tag{1}$$

Various models and techniques are developed to infer the social network based on inadequate aspects [7, 12]. One of the specific categories belonging to such inference determines the cooccurrence based on time and location. Despite encountering many measures, there remains a deficiency in acquiring precise and accurate inferences. In our research, we consider several threshold parameters to quantify more precise inferences. We also develop a framework that infers existing social ties and the hidden relationships in a social network.

Initially, in our research, we present the inference of social ties among people by correlating to their physical presence at several sites and their direct connections. We define a social connection if two individuals X and Y cooccur in a s cell within t hr time frame, such that X calls to person R while connected to a $b1$ base station and in the same time frame Y calls a person S from the same $b1$ base station. Furthermore, we counted the number of cooccurrence of X and Y. Firstly, we find social ties depending upon the number of direct calls between two people. To ensure the correct social connection, we state a threshold, such that the count of direct calls is more than the threshold. Secondly, we evaluate the social relationship between two people by counting the number of calls by X and Y in a specific time frame. Figure 1 states an example that explains the procedure of quantification. Each hexagon represents an area of a single base station. X and Y are together 6 times in various base stations, and there is a variation in the gap of calls. In the first part of research, we use the CDR data set provided by the telecommunication company and two openly available location-based social network data sets, i.e., Brightkite and Gowalla [13]. All data sets resemble to the stated example in Figure 1. We counted the number of concurrences based on multiple gap time frame thresholds and mutual friends. Furthermore, we correlate results with the direct calls based on social connection using Precision and Recall evaluation measures.

In the second phase of research, we explore the false-positive results formed by the CDR-based social tie inference model. We state a missing tie as a suspicious tie between two people if they do not have any direct calls but are found together numerous times. Also, they have a certain number of mutual friends. In the literature, missing ties are defined as either nonresponse or absent ties [12]. In an activity, an actor does not give any information about a tie considered as nonresponse [14], while an absent tie means when an actor does not give any indication about the tie detail. A survey was conducted to monitor the social behavior of the boys' and girls' liking pattern. That was limited to binary data, such that one represents a tie while zero represents no tie. Figure 2 shows visual representation (block modeling) of adjacency matrix made according to survey data [14]. In Figure 2, green-filled slots represent the existence of tie, regardless of its strength, while white slots indicate either absent ties or nonresponse ties. In our research, we explore and classify a subset of missing ties as suspect ties. We conduct a social activity and simulation that generates a data set the same as the CDR data supporting this concept. Furthermore, we correlate the CDR-based social tie inference model's false-positive results with the activity and simulation results.

The contributions of this paper are as follows:

(1) We developed an inference model and a classifier that identifies location-based social ties. The inference model is tested on the CDR-based social network, Brightkite, and Gowalla, using Precision and Recall measures.

(2) We identify a class of suspect ties by examining the social tie inference model's false-positive results.

(3) We conducted an activity-based survey and a simulation that demonstrates and evaluates the suspects' ties.

The rest of this article is organized as follows: Section 2 describes the literature review. Section 3 presents the descriptions of cooccurrence count normalization, inference algorithm, and social tie inference. Brief concepts about the hidden relationship and suspicious links are described in Section 4. The proposed framework and an algorithm to infer suspicious relationships are given in Section 5. Section 6 describes the data set description, results, and analysis. Finally, the conclusion of this article is presented in Section 7.

2. Related Work

The physical world social network is represented as a graph, where nodes are treated as people, and edges are represented as the social tie between two people [15]. In the literature, edge weight is represented as the strength of that particular social tie [10, 16]. A social network such as Twitter forms a bidirectional graph, e.g., a fan follows a celebrity but the celebrity hardly ever follows back. Usage of bidirectional graphs investigates influential networks and most inflectional people [17–19]. Recommendation and targeted marketing are some of the essential objectives of exploring social ties [20, 21]. Theme-based model adopts dynamic programming to explore critical factors, for example, playing and dating are the kind of themes [22]. Social ties coupling and predicting the mobility of users were researched by
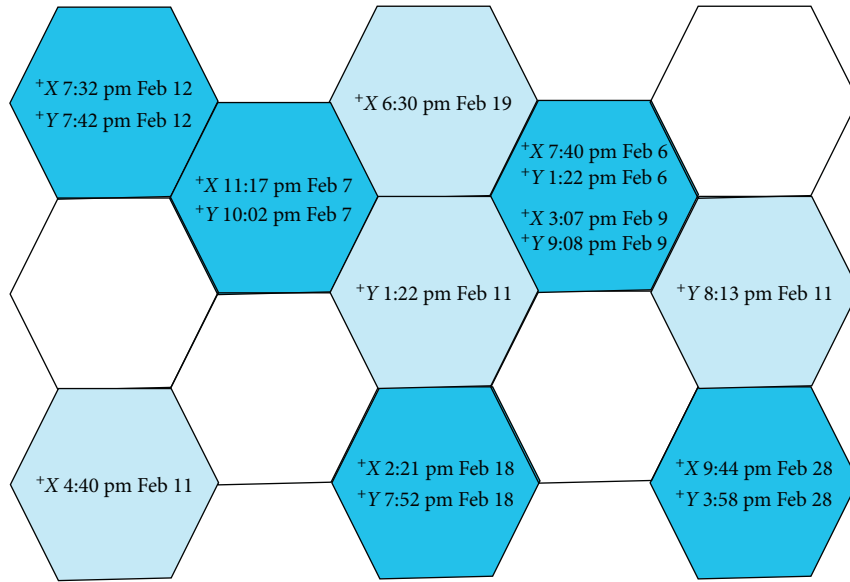
FIGURE 1: Illustration of social tie example for two individuals that have cooccurrence while connected to various base stations of different times.
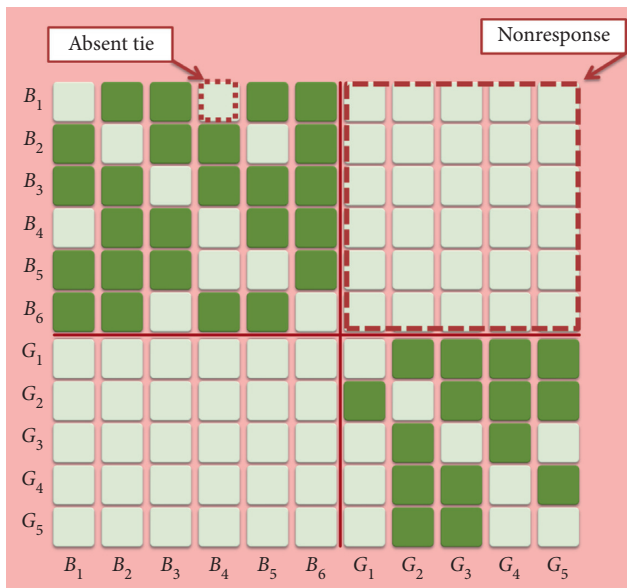

FIGURE 2: Boys and girls liking social network (absent tie and nonresponse).

seeing the physical and network properties (geosocial properties) [23, 24]. An effective prediction technique was proposed to find the typical patterns of two users by comparing the check-in details [24, 25]. Area significance is measured using a weight-assigning method by incorporating two users' cooccurrence for a specific area. A coffee shop is a more significant area as compared to an ordinary place [26]. The scoring mechanism helps in categorizing and labelling of social ties [10]. Inference about the any social network is incomplete if associated features are neglected. The baseline of any social network is the single social connection between two people. In the state of the art, social ties are generally categorized as (1) strong ties, (2) weak ties, and (3) absent

ties [9, 17], whereas the strength of tie depends upon (a) amount of time, (b) emotional intensity, (c) intimacy variables, and (d) social distance [3, 10]. The repeated presence of two individuals in a specific geographical location within a limited time also infers a social connection [7]. The strength of the social tie is directly proportional to the happening of such high cooccurred events.

IoT has emerged as one of the most powerful and impressive technological research domains [2]. IoT presents a novel connectivity concept, where machines can equally collaborate with humans based on actuators and sensors [27, 28]. One research forecasts that smart devices such as electronic medical kits and smart watches will reach up to the worth of USD 160 billion by the end of 2026 [28]. The communication network between smart devices and human forms Social Internet of Things (SIoT) and further opens up new research challenges for researchers. Managing problems such as data scalability, velocity, and variety are few of the emerging issues in SIoT [29]. Understanding social ties among human-to-human, human-to-machines, and machines-to-machines helps to quantify the network performance issue [30].

Social ties are the backbone of any social network [31]. Formation and deformation of social ties affect communities in a network [32]. Besides social tie strength: factors such as location, emotion, situation, age, gender, religion, personality, and many more have a substantial impact on the social connection [10, 33]. Granovetter highlighted the strong connection between weak social connections and finding jobs [34]. In the literature, sources of data commonly used for social analysis are call logs [35, 36], emails, and social-networking websites [5]. In the literature, extensive challenges associated with the integration of visible and invisible networks are highlighted [37]. Investigating criminal social networks using limited clues is one of the emerging research areas of social network analysis (SNA) [38, 39].

Statistically, there are always some hidden or visible associated parameters among cooccurred events. Social network analysis is performed to explore such intriguing knowledge. In the physical world, social network analysis is utilized in job searching [4], studying urban life psychology, investigation of guilt association [12], finding communities [40], spreading of news [41], and influential networks [18, 42]. In the recent era of information and technologies, massive logs are generating for each person, e.g., call records, bank transactions, online purchase records, daily emails, CCTV cameras, and much more mediums [7, 43, 44]. In contrast to the physical world, such mediums further concise the accuracy of results by highlighting such associated features. Despite numerous data sources, there is no optimal procedure to quantify stochastic human nature and social network evolution [45].

The grouping method identifies hidden social groups, which further explores the friend circles and focuses under high privacy settings [46]. Another research explores the hidden social ties using respondent sampling [47]. In the literature, hidden social ties refer to that population, which is hard to access. The population that tries to hide from the social network is hidden in a network [47]. In our research, suspect ties mean actors in a social network that are present and accessible, but they try to hide their social connections. Our second part of the research explores the suspicious ties within the existing network instead of a hidden population in a social network.

3. Data Set Characteristics and Evaluation Measure

3.1. Data Set Descriptions. In our research, we incorporated three large location-based data sets, i.e., CDR, Brightkite, and Gowalla [13]. The CDR large data set used in this study was provided by one of the Chinese mobile telecommunication operator companies. The data set contains 202,000 subscribes along with user demographic information. Calling detailed records contain six months (June 2014—December 2014), and calling detailed records contain these 202, 000 subscribes, which have 221, 451, 169 records. Each record of the data set is represented in the following format.

Caller ID	Call Type	Callee ID	Time
	Duration	LAC ID	CELL ID

Brightkite and Gowalla are openly available location-based social network data sets [13, 48]. Both data sets are gathered using the online social-networking websites. Websites maintain user check-in data by fetching mobile GPS location data. These services use to help people in finding the nearby users and to build social connection. Brightkite contains 58,228 nodes and 214,078 edges, and Gowalla contains 196,591 nodes and 950,327 edges. Other than social network data, both data sets also contain direct social tie data.
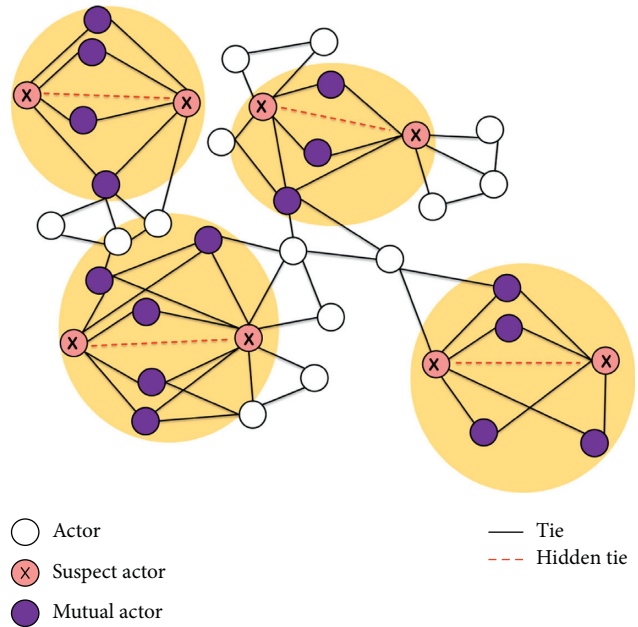

FIGURE 3: Hidden social ties referring suspects.

3.2. Abbreviations and Evaluation Measures. Figure 3 states the example of the social network, having a case of suspect actors and their hidden ties. Actors with several mutual friends but do not have direct connection may have a secret connection. This information helps in identifying them as a suspect tie. The social network evolves, and new connections expand the scope of the social network. One social network is a combination of multiple social networks involving different individuals [31]. A social network can be sliced based on starting and ending time. Social networks can also be divided into subsocial networks monthwise if it has been developed over one year [49, 50]. Social network slicing helps our research further to explore the missing ties between friends of friend relationships.

The following list of abbreviations is used for the quantification processes of Precision and Recall, which will also be used in several parts of the paper:

SK = calling record

The calling records represent the actual number of direct calls that occur between two users. The value of SK is counted to identify the social tie between two users.

CK = times of cooccurrence

Time cooccurrence represents the presence of two users in the range of a common base station. We counted CK when two users were connected to a common base station, and they called any other user.

G = time − frame gap value

The time-frame gap value represents the time interval between two users' calls while connected to a specific base station. For example, user X calls someone at 2 pm and user Y calls someone else at 4 pm; in this case, the gap between the calls is 2 hr. To quantify the results and

evaluate Precision and Recall values, we experimented on the following set of gap values: $G = 30$ mins, 1 hr, 2 hr, 6 hr, 12 hr, and 24 hr.

SNK = threshold for direct calls

The threshold for direct calls represents the set of threshold values for assessing direct calls between two users. We evaluated Precision and Recall curves on the bases of the following set of threshold values SNK = 2, 5, 10, and 15.

CTK = threshold for cooccurrence

The threshold for cooccurrence represents the set of threshold values for the two users' presence in a specific base station. We tested the performance on the CTK threshold values ranges from 1 to 40.

MuF = threshold for mutual friend

The threshold for mutual friend represents the set of threshold values for the two users' mutual friend. We tested the performance on the MuF threshold values ranges from 1 to 100:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3}$$

$$\text{where as} \begin{cases} TP = (SNK \geq SK \text{ and } CTK \geq CK), \\ FP = (SNK < SK \text{ and } CTK \geq CK), \\ FN = (SNK < SK \text{ and } CTK < CK), \\ TN = (SNK \geq SK \text{ and } CTK < CK), \end{cases} \tag{4}$$

$$F1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{5}$$

3.3. Cooccurrence Count Normalization Measure.

Cooccurrence count value CV tells the presence of two users in the region of one base station. An issue related to CV counting is explained and resolved using an example for the two users X and Y, shown in Figure 4. We counted CV when two users were connected to a common base station, and they called any other user in a specific time frame. The example is shown in Figure 4 states the call log details of users X and Y gathered in a time frame T.

Let $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ and $Y = \{y_1, y_2, y_3\}$, then

$$CV = \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_i\}, \tag{6}$$

where $\alpha_i = \text{Time Gap}(x_i, Y)$, $i = 1$ to 6, $j = 1$ to 3,

$$\text{Time Gap}(x_i, Y) = \min\{(x_i, y_1), (x_i, y_2), \ldots, (x_i, y_j)\}. \tag{7}$$

In Figure 4, x_1 and x_2 call times have the closest call time to y_1 call. In this case, a count value of CV can be calculated as 2. However, such counting may lead to a wrong inference. It is the same as if one person calls once, and another person calls n-times within a specific time frame, equals n as the count value. To resolve this issue, we propose a normalization equation that decreases the count value periodically. We introduce Beta (β) value as a periodic normalizing factor.

Let X denotes a set of calls by user X and Y denotes a set of calls by user Y; if $(X \geq Y)$, then $X \times Y$ else $Y \times X$.

According to the example stated in Figure 4, we assumed β for set $Y = (y_1\beta, y_2\beta, y_3\beta)$.

For first match value of $\beta = 1$.

For the second match values of $\beta = \beta/2$.

Likewise, for the nth match value of $\beta = \beta/n$,

$$= \left[((x_1, y_1) \times (y_1\beta)) + \left((x_2, y_1) \times \left(\frac{y_1\beta}{2} \right) \right) \right]$$

$$+ \left[((x_3, y_2) \times (y_2\beta)) + \left((x_4, y_2) \times \left(\frac{y_2\beta}{2} \right) \right) + \left((x_5, y_2) \times \left(\frac{y_2\beta}{3} \right) \right) \right] \tag{8}$$

$$+ [(x_6, y_3) \times (y_3\beta)],$$

$$= \sum_{m=1}^{mk} \sum_{n=1}^{nk} \left((x_n, y_m) \times \left(\frac{y_m\beta}{n} \right) \right). \tag{9}$$

In equation (9), mk refers to the total number of calls made by user X to Y, while nk refers to the total number of calls made by user Y to X. equation (9) finds intermediate value for CV, i.e., 4.33, instead of maximum 6 or minimum 3 values.

4. Social Tie Inference

We initially investigated direct social ties formed by CDR data sets and compared them to the indirect social ties formed based on common location using Algorithm 1. By
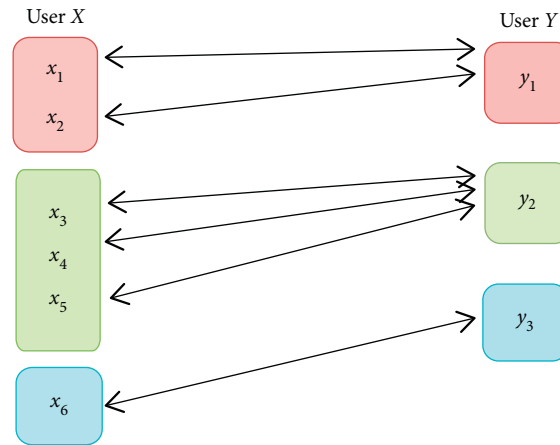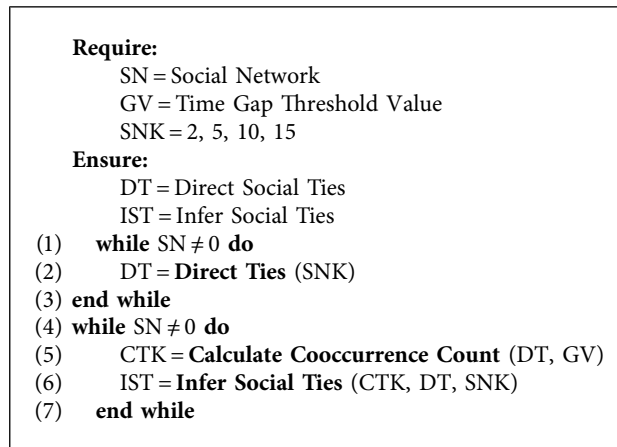
FIGURE 4: User X and user Y cooccurrence count comparison case.

Require:
      SN = Social Network
      GV = Time Gap Threshold Value
      SNK = 2, 5, 10, 15
Ensure:
      DT = Direct Social Ties
      IST = Infer Social Ties
(1)   **while** SN $\neq$ 0 **do**
(2)   DT = **Direct Ties** (SNK)
(3) **end while**
(4) **while** SN $\neq$ 0 **do**
(5)   CTK = **Calculate Cooccurrence Count** (DT, GV)
(6)   IST = **Infer Social Ties** (CTK, DT, SNK)
(7)   **end while**

ALGORITHM 1: Social tie inference.

direct ties, we mean calling or direct connection. For example, person A calls person B refers to a direct tie between A and B. Algorithm 1 takes GV, SNK, and CDR data sets (social network) as inputs. Furthermore, the algorithm has two parts; initially, it finds the direct ties between two individuals depending upon the SNK threshold value. Secondly, it counts the presence of two individuals based on several parameters. The **Calculate Cooccurrence Count()** function finds the number of cooccurrences using equation (9), explained in the previous section. **Infer Social Ties()** function finds the social connections depending upon CTK, DT, and SNK and inferred them as the social ties.

4.1. CDR-Based Social Tie Inference. A social tie is inferred between two persons if they are found together at several sites numerous times. The inference algorithm identifies two sets of results, i.e., direct social ties and inferred social ties. For the cross-validation of results, we correlate the direct tie results with the inferred ones. Precision and Recall evaluation measures are used to examine the results. We tested all records based on threshold values, K is the direct calls, M is the times of cooccurrence, and G is the time frame gap value. While N as direct call count shows the degree of friendship,

more value of N indicates the friendship strength. Figure 5 shows the Precision graph, which contains four sets, Figures 5(a)–5(d). The whole data set is examined based on K and the value of M and v.

In Figure 5(a), the value of K is 15 which represents the users with direct calls between each other equals to or greater than 15. The value of M is the number of cooccurrence for two different users. The Precision values are comparatively significantly less for M in the range of 0 to 10. In contrast, the value of Precision increases exponentially for the value of M in the range of 10 to 30. The higher value of M indicates higher cooccurrence of users. A positive correlation can be observed between the values of M and Precision. It infers that cooccurrence is a significant attribute that affects positively in identifying social ties. All graphs in Figure 5 have six different lines; each line represents the different time gap ranges. It can also be seen that the values of gap value 30 minutes are having more Precision while the rest lines of 1 hour, 2 hours, 6 hours, 12 hours, and 24 hours are having less Precision. This also clues that the strength of ties has a specific effect on Precision. Users having strong social connections, most of the time, are found together in certain areas. This pattern is explicitly observed by looking cooccurrence value $M = (20 \text{ to } 40)$ and gap time frame $G = 30$
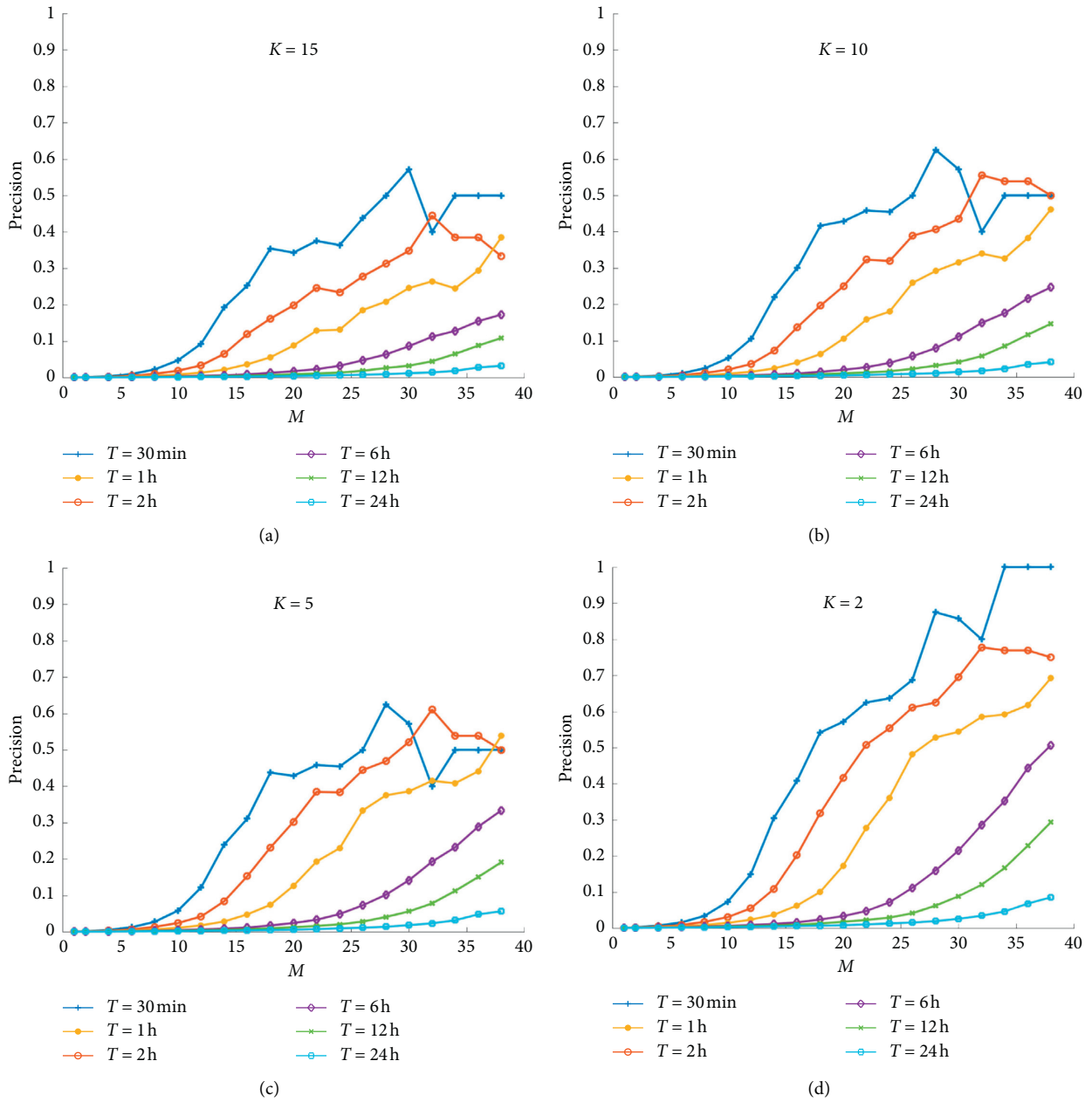
FIGURE 5: The Precision formed between two callers due to social contact depending on the number of times they have connected through the same base station within a specific time frame. Figure 5(a) shows the Precision for the six different time frame lines, whereas M is the number of cooccurrences and K is the number of direct calls. (a) $K = 15$, (b) $K = 10$, (c) $K = 5$, and (d) $K = 2$.

minutes. Another positive correlation is found between the degree of friendship and physical presence at a specific place.

To see the effect of friendship strength, we evaluated results for the four different K values, i.e., 2, 5, 10, and 15. A typical pattern is found in all the graphs shown in Figure 5. It shows that Precision is less for people whose mutual presence is less at different sites. Also, people with strong social ties spent less than 1 hr time together at a specific location. To understand the graph's actual meaning, we quantify and reconcile with the actual direct social ties. It is observed that a positive correlation in results infers that

people with strong social connections often visit places together.

Figure 6(a) represents the Recall results. We tested and evaluated Recall based on the same measures as Precision, i.e., direct calls, cooccurrence, and gap time frame. Figure 6(a) states that the value of Recall is at a maximum when we consider a less number of commonplaces. An inverse trend is observed between the values of Recall and M, specifically for the M value ranging between 0 and 3. The same as Precision, Recall is also evaluated based on six different time gap values.
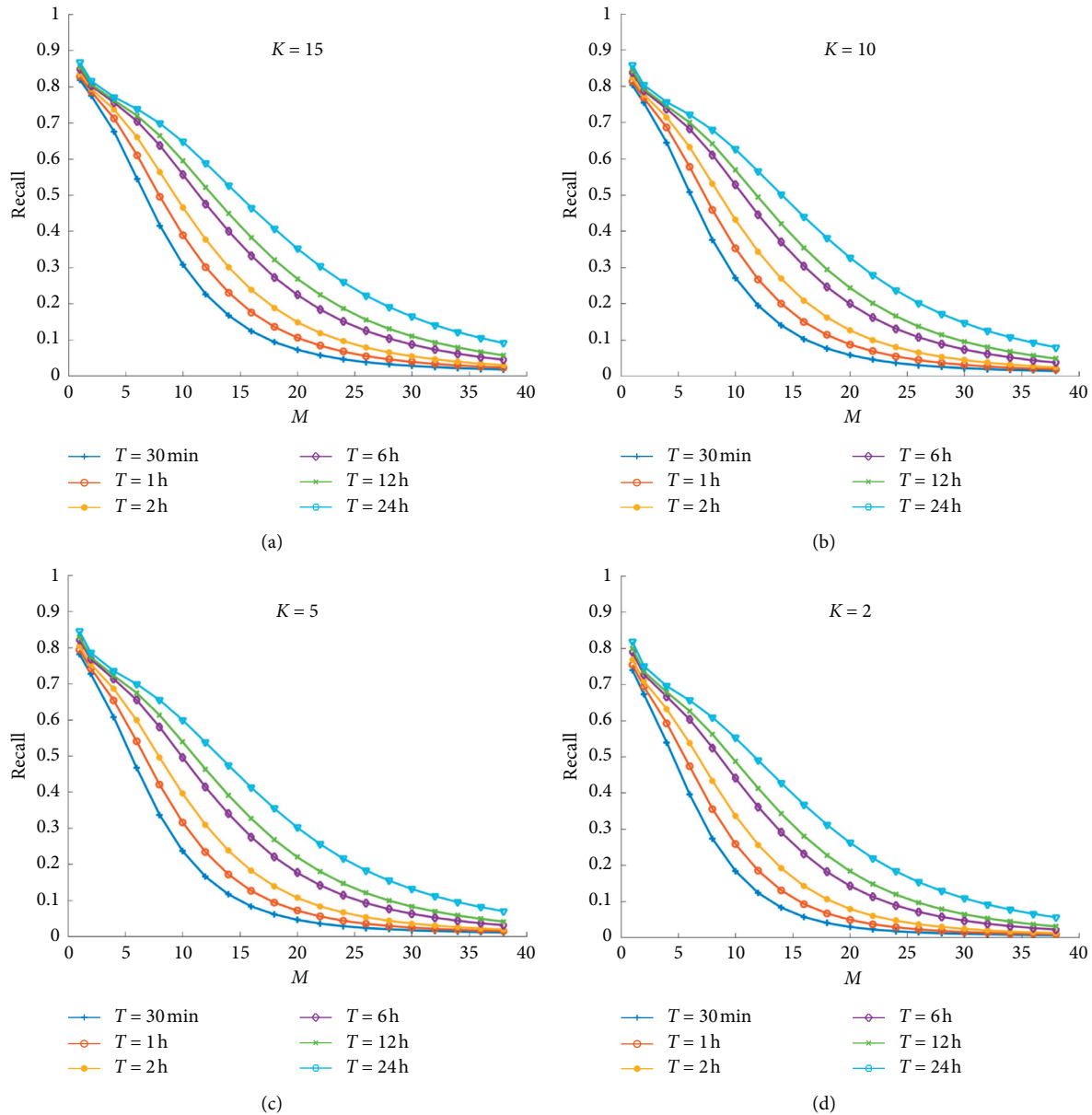
FIGURE 6: The Recall formed between two callers due to social contact depending upon the number of times they have connected through the same base station within a specific time frame. Figure 6(a) shows the Recall for the six different time frame lines, whereas M is the number of cooccurrences and K is the number of direct calls. (a) $K = 15$, (b) $K = 10$, (c) $K = 5$, and (d) $K = 2$.

This part of the research finds people's cooccurrence based on the same base station connectivity in a specific time frame and infers them as social ties. Furthermore, it cross relates the inferred results with direct call results.

4.2. Brightkite- and Gowalla-Based Social Tie Inference. Brightkite and Gowalla data sets contain direct social ties as well as the check-in information of each user. In our study, we investigated both data sets based on several dimensions and found some of the very interesting facts. During the analysis, we observed positive correlations between mutual friend count values MuF and user check-in details. In Brightkite, MuF ranging 6 to 40 and in Gowalla, MuF

ranging 25 to 90 shows the positive corelation with Precision. We also measure the effect of gap time value on the Precision and Recall and evaluated results based on several gap time values.

Figures 7(a) and 7(c) of line graph show the relation between mutual friend count values MuF and Precision, while Figures 7(b) and 7(d) show the relation between mutual friend count values MuF and Recall. In Figure 7, results shows that people having a certain number of mutual friends use to visit place together or with little gap of interval.

The social tie inference framework infers some of the absent ties as social ties in the form of false-positive results. To further extract the actual meaning of such incorrect
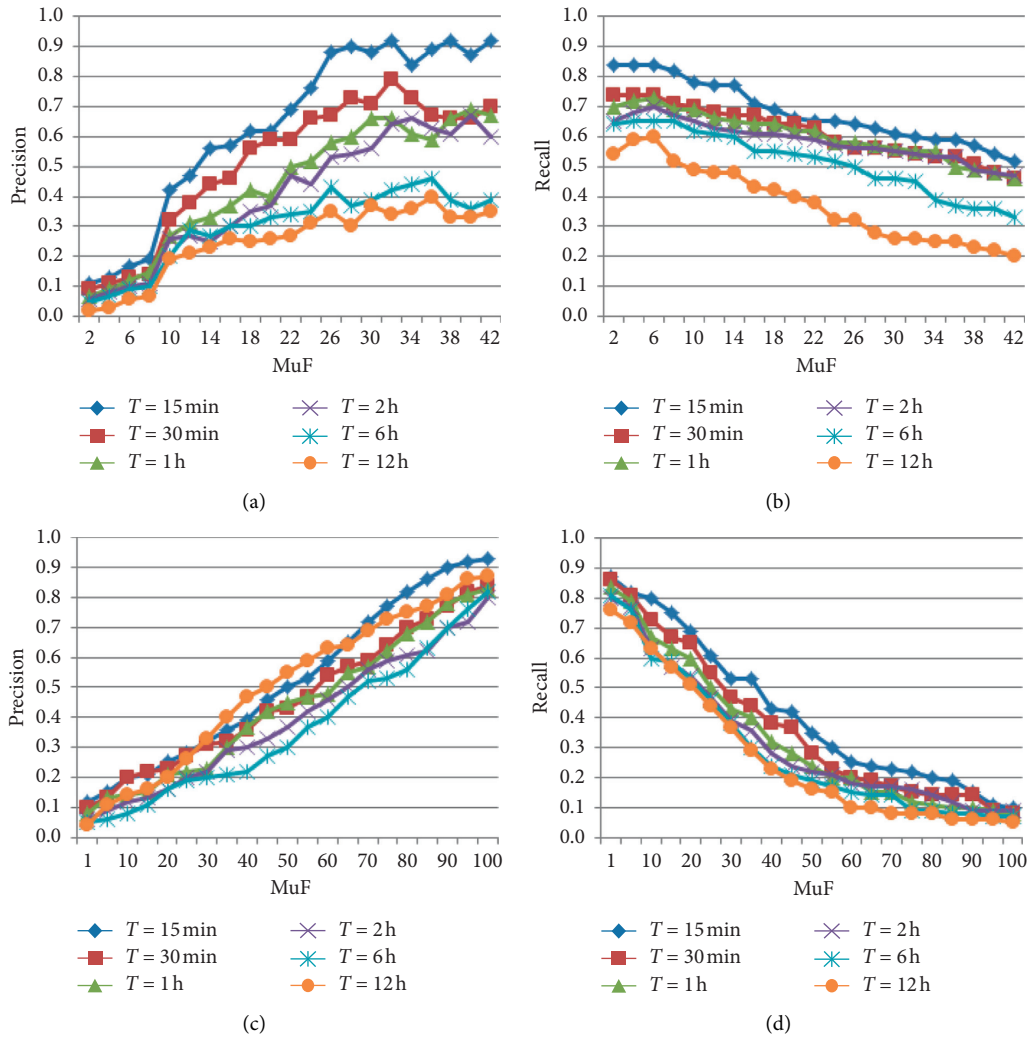
FIGURE 7: The Precision and Recall formed between two users due to social contact depending on the number of times they have checked in to certain location within a specific time frame. Figure 7(a) shows the Precision for the six different time frame lines, whereas MuF is the number of mutual friend count. (a) Brightkite Precision, (b) Brightkite Recall, (c) Gowalla Precision, and (d) Gowalla Recall.

inferences, we conducted a social activity and simulation. The false-positive results of the first part of the research serve as the foundation for the second part. Activity under the first part data set is conducted, and the false-positive results are examined by studying the human cooccurrence patterns, described in the next section Suspicious Ties. This stage of research gave us a clue to further exploit the category of missing ties.

5. Suspicious Ties

An absent tie can be inferred as a suspect tie, if it satisfies the following properties:

    (1) M, number of mutual friends

    (2) C, number of cooccurrence on different sites

In the CDR data set, each cell is treated as a single cell of the base station. Our model adopts the following four features for evaluation:

(1) Base station

(2) User ID

(3) Gap time threshold

(4) Call time stamp

    The following mathematical model explains the problem and its formulation:

    Let BS denotes a set $\{bs_i\}$ of bs_n points, and bs_i is called as the distinct base station cell

    Let DU denotes a set $\{du_i\}$ of du_n points, and du_i is called as the distinct user

Let TH denotes a set $\{th_i\}$ of th_n points, and th_i is called as the threshold value for timeframe

Let C denotes a set $\{c_i\}$ of c_n points, and c_i is called as the call information,

$$c_i = (\text{id}, \text{ct}), \quad \text{where} \begin{cases} \text{id as call id,} \\ \text{ct as outgoing call time.} \end{cases} \quad (10)$$

Let R denotes a set $\{r_i\}$ of r_n points, then

$$r_i = [(du_i, c_i(\text{id})) \times (du_i, c_i(\text{id}))] \wedge du_i \neq du_i. \quad (11)$$

Let RT denotes a set $\{rt_i\}$ of rt_n points, then

$$rt_i = [(du_i, c_i(\text{ct})) - (du_i, \text{ct}_i(\text{ct}))] \wedge du_i \neq du_i. \quad (12)$$

Let S denotes a set $\{s_i\}$ of s_n points, then

$$s_i = [(r_i \in bs_i) \wedge (rt_i \leq th_i)], \quad (13)$$

where s_i is a set of elements that identifies distinct callers based on the same base station connectivity and a definite number of calls in a specific time frame.

6. Suspect Inference Framework

We studied the pattern of exceptional cases belong to the false-positive set and described a subset of the false-positive set as suspect ties. Physical activity was designed and conducted to investigate the formation of suspect social ties. Activity consisting of 50 people, and a data set was generated within almost 4-5 hours. A basketball court was utilized for the activity. Nine circles were drawn physically on the basketball court, assumed as the base station cell. Out of 50 people, nine were directed to act as a base station. The boundary of each circle was considered as a range of the base station cell. Rests of the 41 persons were directed to perform the following two steps.

6.1. Selection Step. Initially, each person from 41 people was asked to choose two sets of friends. One set as obvious friends and the second set of hidden friends such that the size of the hidden friend set should be at most 1/5 of the obvious friend set size, e.g., if one person has five people in apparent friend set, he can have no more than one hidden friend. After the selection of both sets by each individual, information was shared with one of our representatives.

6.2. Operation Step. In this phase, each person was directed to follow the following rules:

(1) You should not call your hidden friend

(2) You should call all of your obvious friends at least once

(3) You should conduct a maximum number of calls to your closest obvious friend and second maximum to a second level obvious friend and likewise to the least friend

(4) You must try to meet your hidden friend as much as possible physically

The method of calling is like, if person A wants to call person B from base station $B1$, the person has to go to the base station $B1$ and register a call with a person acting and standing in base station $B1$. Respective base station person will write and make an entry with five parameters, i.e., *Caller Name*, *Callee Name*, *Time*, *From Base Station Name*, and *To Base Station Name*. The data set was gathered in the following format. An example is given below.

Caller Name	Callee Name	Time Stamp	From Base Station	To Base Station
User 1	User 12	02/03/2019 10:15:08	BS 7	BS 2
..	..	..	..	..

For the understanding of the variations and patterns, the same activity was also designed using simulation. The whole simulation followed the same conditions, and another data set was generated using a random function. Based on the activity, a framework is designed to separate a class of suspect ties. Proposed inference framework work is designed and implemented using pipe and filter architecture, shown in Figure 8. Algorithm 2 shows the implementation of suspicious tie inference framework, explained in Figure 8. The framework takes the social network matrix, count threshold value, gap time value, and mutual friend count value as inputs and filters the result. Initially, **Calculate Levels()** function finds the five level depth information for each distinct user. Let us say if A calls B, B calls C, C calls D, and D calls E, it implies that $A = 0$, $B = 1$, $C = 2$, $D = 3$, and $E = 4$ represent five levels. This step ensures that all the levels have distinct users. Secondly, **Find Suspects()** function selects only those sets of users from level 1 and level 3 that do not have any direct calls and M number of mutual friends. Furthermore, **Calculate Subsocial Network()** function generates subgraphs using level 1 and level 3 details depending upon the gap time value. Results are filtered on these bases of time gap value, e.g., two users called some other user while connected to the same base station within the given time frame, explained in the previous section. After that, **Infer Hidden Ties()** function uses the proposed normalization method to find the number of the count, defined in equation (9), and then all results are filtered according to the cooccurrence count CV and the mutual friend threshold value M. Based on mentioned parameters and thresholds, Algorithm 2 significantly identifies the subclass of missing ties as suspicious ties.

The results of the activity and simulation are computed and evaluated using Precision and Recall measures.
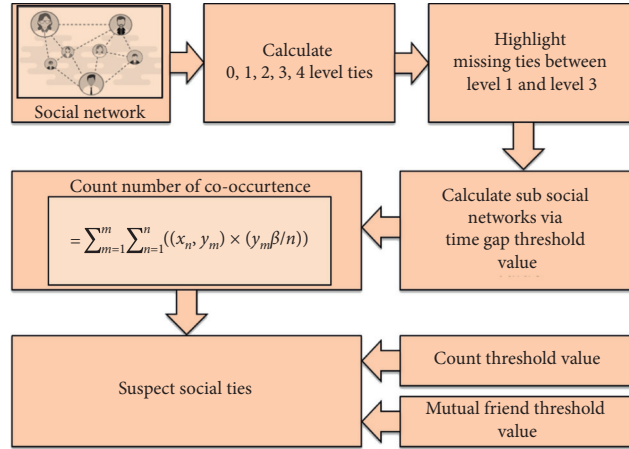
FIGURE 8: Suspect social tie inference framework.

Require:
    SN = Social Network
    GV = Time GapThreshold Value
    CV = Cooccurrence Count Threshold Value
    M = Mutual Friend Count
Ensure:
    ST = Inferred Social Suspect Ties
(1)   **while** SN ≠ 0 **do**
(2)   Levels [5, n] = **Calculate Levels()**
(3)   Suspects = **Find Suspects (**Level 1, Level 3**)**
(4)   SG = **Calculate Subsocial Network (**Suspects, GV**)**
(5)   ST = **Infer Suspect Ties (**SG, CV, M**)** using equation (9)
(6)   **end while**

ALGORITHM 2: Suspicious hidden social tie inference.

TABLE 1: Evaluation of simulation and social activity.

| | Gap time | $M \geq 2$ | | | | | | $M \geq 5$ | | | | | |
| | | Simulation | | | Social activity | | | Simulation | | | Social activity | | |
		P	R	$F1$	P	R	$F1$	P	R	$F1$	P	R	$F1$
CV ≥ 5	GV < 10	0.12	0.68	0.20	0.12	1.00	0.21	0.13	0.62	0.21	0.12	0.79	0.21
	GV < 20	0.10	0.42	0.16	0.15	0.79	0.25	0.15	0.31	0.20	0.13	0.42	0.20
	GV < 30	0.21	0.22	0.21	0.71	0.59	0.64	0.25	0.18	0.21	0.23	0.28	0.25
	GV ≥ 30	0.34	0.23	0.27	1.00	0.12	0.21	0.30	0.10	0.15	0.32	0.19	0.24
CV ≥ 10	GV < 10	0.10	0.33	0.15	0.12	0.48	0.19	0.09	0.30	0.14	0.11	0.61	0.19
	GV < 20	0.11	0.28	0.16	0.27	0.50	0.21	0.09	0.18	0.12	0.10	0.39	0.16
	GV < 30	0.13	0.18	0.15	0.29	0.30	0.29	0.09	0.10	0.09	0.09	0.20	0.12
	GV ≥ 30	0.13	0.08	0.10	0.26	0.05	0.08	0.10	0.04	0.06	0.13	0.07	0.09
CV ≥ 20	GV < 10	0.04	0.18	0.07	0.08	0.23	0.12	0.05	0.29	0.09	0.11	0.39	0.17
	GV < 20	0.07	0.08	0.07	0.05	0.24	0.08	0.04	0.11	0.06	0.10	0.28	0.15
	GV < 30	0.04	0.06	0.05	0.09	0.09	0.09	0.03	0.05	0.04	0.09	0.19	0.12
	GV ≥ 30	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.05	0.02

$^*P$ = Precision, R = Recall, and $F1$ = $F1$ Score.
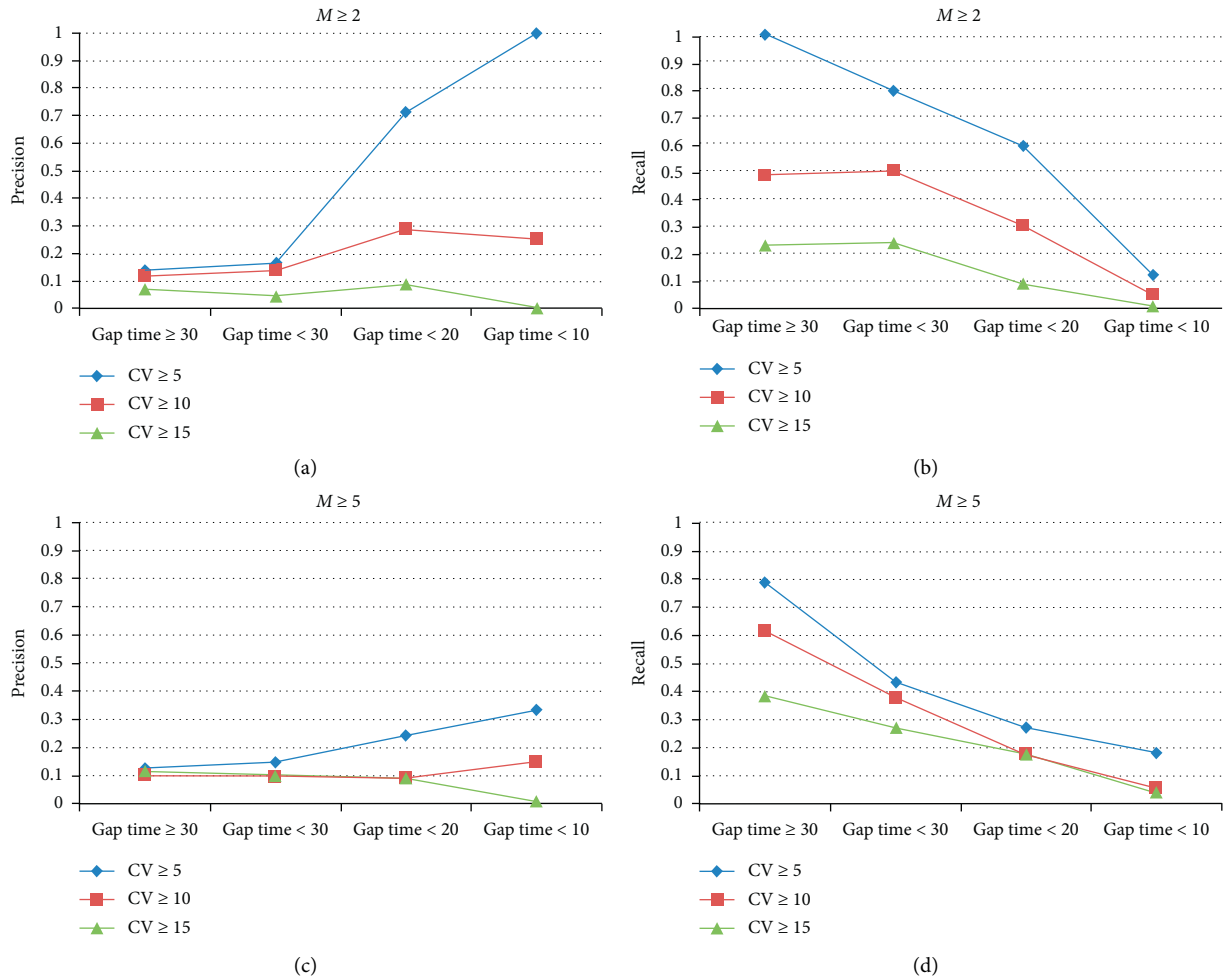
(a)

(b)

(c)

(d)

Figure 9: Evaluation of social activity using Precision and Recall.

Evaluation results of simulation and social activity conducted are shown in Table 1. Precision, Recall, and $F1$ Score measures are used to evaluate the framework that is further calculated based on the cooccurrence count CV, mutual friend count M, and gap time GV parameter setting values. $F1$ Score is calculated using equation (5). Definitions of the related parameters are given as follows.

TP	Which were hidden friends and system infer as hidden friends
TN	Which were not hidden friends and system infer as not hidden friends
FP	Which were not hidden friends and system infer as hidden friends
FN	Which were hidden friends and system infer as not hidden friends

6.3. Results and Discussion. We tested and evaluated all records based on the cooccurrence count value CV, mutual friend M, and gap time value GV as the gap time. Figures 9(a)–9(d) show the evaluation of results for the activity, based on three different values of the cooccurrence count value CV, i.e., CV = 5, 10, and 15, and two different values of the mutual friend M, i.e., $M \geq 2$ and ≥ 5. Likewise, Figures 10(a)–10(d) show the evaluation of results for the simulation data set along with mentioned CV and M values. Results were generated on four values of gap time, i.e., ≥ 30, < 30, < 20, and < 10.

In Figures 9(a) and 9(b), Recall is maximum and Precision is less where the value of GV ≥ 30, CV ≥ 5, and $M \geq 2$. The system obtains a maximum number of relevant hidden ties along with false-positive results. By GV ≥ 30, CV ≥ 5, and $M \geq 2$, it means that the gap between the two calls is 30 mins or more while the count of cooccurred events is kept minimum five and mutual friend count as two or less. The system's performance drops when gap time is reduced to < 30, < 20, and < 10. Even though there is a drop in true-positive values, a significant drop in the false-positive values can be seen. Results become concise, with the variation in both values of GV and M. The limited data set collected using activity highlights the occurrences of hidden ties. Whole activity and simulation were designed to get similar fields of data as CDR so that the proposed framework is compatible with the CDR data set.

Results shown in Figures 9 and 10 exhibit the existence of hidden relationships. The parameter, such as the gap time value GV, helps to identify the time frame selection such that
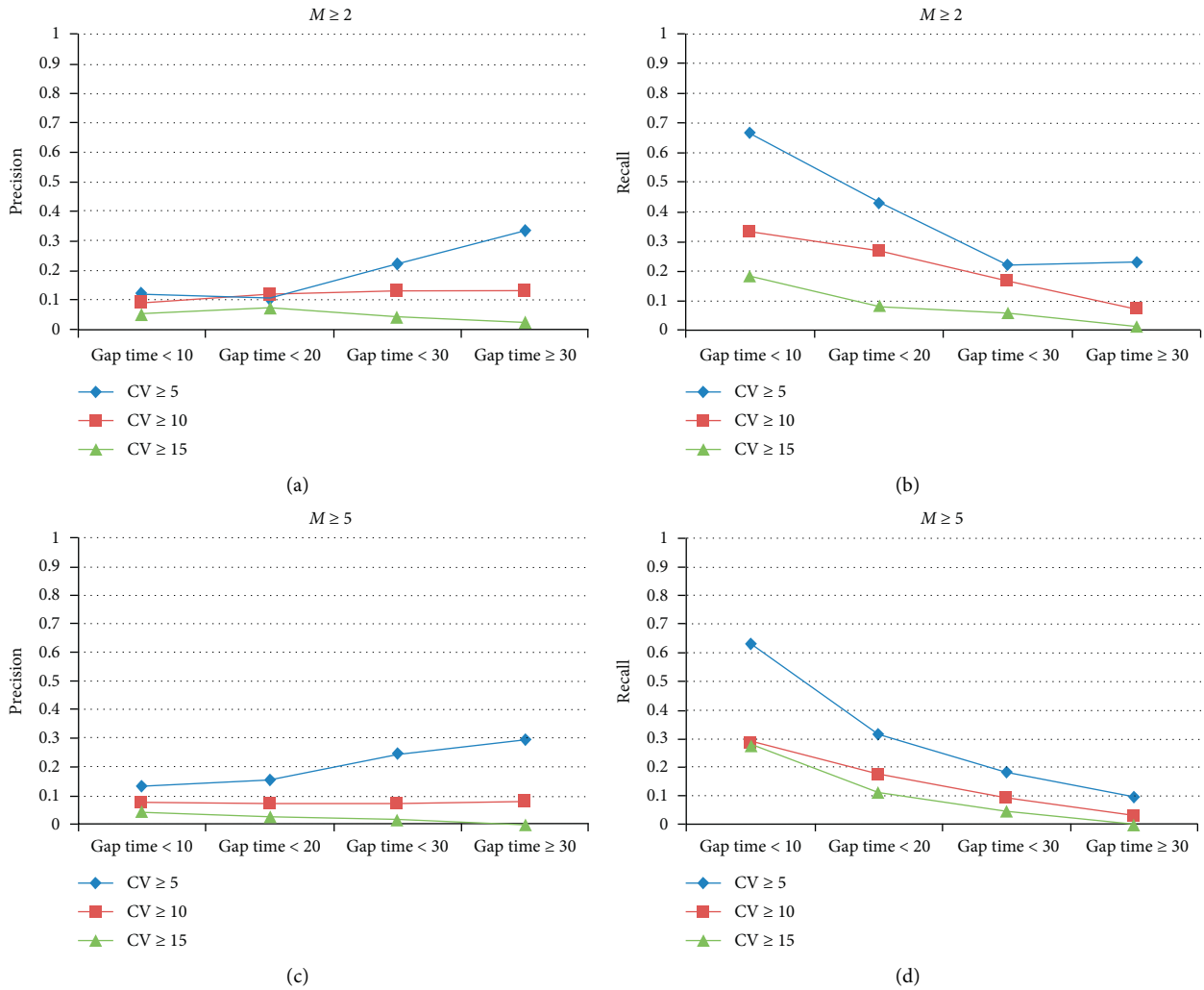
(a)

(b)

(c)

(d)

FIGURE 10: Evaluation of simulation using Precision and Recall.

what gap value is suitable to get optimal performance? Likewise, Figures 10(a) to 10(d) represent the Precision and Recall curves for the simulation data set. Results are gathered by setting the same values for the threshold parameters CV, GV, and M. Figures 10(a) and 10(d) give the least values for Precision and Recall, which tells that if threshold values are tightened, so does the count relevant retrieve results get low. Out of all the results, Precision and Recall values for both simulation and activity data set show maximum, if threshold values GV $\geq$ 30, CV $\geq$ 5, and $M \geq 2$ are set. We found some exciting dissimilarities in the results between simulation and activity data set results during the complete evaluation process. The simulation results do not give higher Recall and Precision value compared to activity data results. We concluded that the data set of simulation is generated using random function while the activity had the human hiding patterns. These results also help to infer human psychology in building hidden ties. The simulation data set generator works based on the same constraints mentioned as rules of the activity. However, the key difference between simulation and activity is the selection of friends, hidden friend, and the

pattern of calling. While conducting the simulation, trivial variation in results was observed as the random selection has random patterns. According to our findings, simulation and activity both exhibit patterns of hidden ties. However, activity results are more pronounced and significantly identifying the hidden relationships.

It is essential to highlight some critical questions and dependent variables that help to find hidden social ties between two people, for example, why two people hide their social ties? Is this deliberate or unintentional action? What if they are deliberately hiding their social tie for a purpose? In such a scenario, extracting a social tie for two people is a kind of intense problem. It is a kind of investigation process which explores a clue to draw some relationship between two people. Investigation designates if two people are posting a picture on social media to be more cautious about not identifying their social ties. While if they are doing some private activity, they will be less careful, for example, posting a picture on Facebook or Flickr compared to calling to a person from a specific location. Although extracting hidden social ties include various privacy issues. We designed

activity and simulation that generated the data set by the caller data record (CDR) data set. We thoroughly investigated the patterns of connectivity and established a framework to infer social ties. This research has opened up a new direction further to explore connectivity in the Social Internet of Things (SIoT), specifically machine-to-machine direct communication and machine-to-human or human-to-machine hidden relationships. In many cases, several machines work together but they rarely have a direct connection.

7. Conclusion and Future Work

In this research, we have examined the developments of social connection patterns based on physical gathering. In the first part of our research, we explored the correlation between direct communication and two individuals' physical presence. To check the system's performance and evaluation, we examined all results by utilizing Precision and Recall evaluation measures. We also present a periodic normalization equation for the cooccurrence count. In the second phase, we propose the suspect tie inference framework. False-positive results of the first part of the research are the ground to the second part of the study. The proposed framework adopts pipe and filter architecture, where the threshold values control each filter. The framework's fundamental objective is to take the data set such as CDR (caller data record) and infer suspect social ties, depending upon the specified threshold values. Analyzing the results critically, we propose a theory that identifies suspect social ties. Besides this, for comparison and evaluation, we conducted real-time human-based activity and simulation. Keeping in mind the structure of the actual CDR data set, the whole activity was designed and evaluated. In contrast to existing work, our research focus is on hidden ties instead of the hidden actors. In the future, we are aiming to explore the homophilic nature of suspect ties within the Social Internet of Things (SIoT).

Data Availability

The data used can be found at http://snap.stanford.edu/data/index.html#locnet.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this work.

Acknowledgments

References

[1] L. Weng, M. Karsai, N. Perra, F. Menczer, and A. Flammini, "Attention on weak ties in social and communication networks," in *Complex Spreading Phenomena in Social Systems*, pp. 213–228, Springer, Berlin, Germany, 2018.

[2] A. M. Ortiz, D. Hussein, S. Park, S. N. Han, and N. Crespi, "The cluster between internet of things and social networks: review and research challenges," *IEEE Internet of Things Journal*, vol. 1, no. 3, pp. 206–215, 2014.

[3] P. Luarn and Y.-P. Chiu, "Key variables to predict tie strength on social network sites," *Internet Research*, vol. 25, no. 2, pp. 218–238, 2015.

[4] D. Goel and K. Lang, "Social ties and the job search of recent immigrants," *ILR Review*, vol. 72, no. 2, pp. 355–381, 2019.

[5] H. Kizgin, A. Jamal, N. Rana, Y. Dwivedi, and V. Weerakkody, "The impact of social networking sites on socialization and political engagement: role of acculturation," *Technological Forecasting and Social Change*, vol. 145, pp. 503–512, 2019.

[6] P. Suebvises, "Social capital, citizen participation in public administration, and public sector performance in Thailand," *World Development*, vol. 109, pp. 236–248, 2018.

[7] Z. Liu, Y. Qiao, S. Tao, W. Lin, and J. Yang, "Analyzing human mobility and social relationships from cellular network data," in *Proceedings of the 2017 13th International Conference on Network and Service Management (CNSM)*, pp. 1–6, Tokyo, Japan, November 2017.

[8] L. Matosas-López and A. Romero-Ania, "The efficiency of social network services management in organizations. an in-depth analysis applying machine learning algorithms and multiple linear regressions," *Applied Sciences*, vol. 10, no. 15, p. 5167, 2020.

[9] M. S. Granovetter, "The strength of weak ties," in *Social Networks*, pp. 347–367, Elsevier, Amsterdam, Netherlands, 1977.

[10] T. Nguyen, L. Zhang, and A. Culotta, "Estimating tie strength in follower networks to measure brand perceptions," in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 779–786, Vancouver, Canada, August 2019.

[11] H. Mattie, K. Engø-Monsen, R. Ling, and J.-P. Onnela, "Understanding tie strength in social networks using a local bow tie framework," *Scientific Reports*, vol. 8, no. 1, p. 9349, 2018.

[12] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg, "Inferring social ties from geographic coincidences," *Proceedings of the National Academy of Sciences*, vol. 107, no. 52, pp. 22436–24441, 2010.

[13] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1082–1090, San Diego, CA, USA, August 2011.

[14] A. Žnidaršič, P. Doreian, and A. Ferligoj, "Absent ties in social networks, their treatments, and blockmodeling outcomes," *Advances in Methodology & Statistics/Metodoloski Zvezki*, vol. 9, no. 2, 2012.

[15] S. A. Myers, A. Sharma, P. Gupta, and J. Lin, "Information network or social network?: the structure of the twitter follow graph," in *Proceedings of the 23rd International Conference on World Wide Web*, pp. 493–498, Seoul, Republic of Korea, April 2014.

[16] X. Zuo, J. Blackburn, N. Kourtellis, J. Skvoretz, and A. Iamnitchi, "The power of indirect social ties," 2014, http://arxiv.org/abs/1401.4234.

[17] M. A. Brandão and M. M. Moro, "Tie strength in co-authorship social networks: analyses, metrics and a new computational model," in *Proceedings of the Anais Estendidos do XXV Simpósio Brasileiro de Sistemas Multimídia e Web*, pp. 17–20, Rio de Janeiro, Brazil, 2019.

[18] H. A. Khanday, A. H. Ganai, and R. Hashmy, "A comparative analysis of identifying influential users in online social networks," in *Proceedings of the 2018 International Conference on Soft-Computing and Network Security (ICSNS)*, pp. 1–6, Coimbatore, India, February 2018.

[19] R. Garcia-Guzman, Y. A. Andrade-Ambriz, M.-A. Ibarra-Manzano, S. Ledesma, J. C. Gomez, and D.-L. Almanza-Ojeda, "Trend-based categories recommendations and age-gender prediction for pinterest and twitter users," *Applied Sciences*, vol. 10, no. 17, p. 5957, 2020.

[20] A. Talukder, M. G. R. Alam, N. H. Tran, D. Niyato, and C. S. Hong, "Knapsack-based reverse influence maximization for target marketing in social networks," *IEEE Access*, vol. 7, pp. 44182–44198, 2019.

[21] N. Ampazis, T. Emmanouilidis, and F. Sakketou, "A matrix factorization algorithm for efficient recommendations in social rating networks using constrained optimization," *Machine Learning and Knowledge Extraction*, vol. 1, no. 3, p. 928, 2019.

[22] N. Zhou, X. Zhang, and S. Wang, "Theme-aware social strength inference from spatiotemporal data," in *Proceedings of the International Conference on Web-Age Information Management*, pp. 498–509, Macau, China, June 2014.

[23] P. A. Grabowicz, J. J. Ramasco, B. Gonçalves, and V. M. Eguíluz, "Entangling mobility and interactions in social media," *PLoS One*, vol. 9, no. 3, Article ID e92196, 2014.

[24] D. Yang, B. Qu, J. Yang, and P. Cudre-Mauroux, "Revisiting user mobility and social relationships in LBSNs: a hypergraph embedding approach," in *Proceedings of the World Wide Web Conference*, pp. 2147–2157, San Francisco, NY, USA, May 2019.

[25] R.-H. Li, J. Liu, J. X. Yu, H. Chen, and H. Kitagawa, "Co-occurrence prediction in a large location-based social network," *Frontiers of Computer Science*, vol. 7, no. 2, pp. 185–194, 2013.

[26] V. Jayadevan, K. Bharadwaj, A. Kumar, and P. Khandelwal, "Discovering local social groups using mobility data," *International Journal of Computer Applications*, vol. 120, no. 21, 2015.

[27] A. J. Jara, Y. Bocchi, and D. Genoud, "Social internet of things: the potential of the internet of things for defining human behaviours," in *Proceedings of the 2014 International Conference on Intelligent Networking and Collaborative Systems*, pp. 581–585, Salerno, Italy, September 2014.

[28] W. A. D. M. Jayathilaka, K. Qi, Y. Qin et al., "Significance of nanomaterials in wearables: a review on wearable actuators and sensors," *Advanced Materials*, vol. 31, no. 7, Article ID 1805921, 2019.

[29] J. Ren, H. Guo, C. Xu, and Y. Zhang, "Serving at the edge: a scalable iot architecture based on transparent computing," *IEEE Network*, vol. 31, no. 5, pp. 96–105, 2017.

[30] I. Seeber, E. Bittner, R. O. Briggs et al., "Machines as teammates: a research agenda on ai in team collaboration," *Information & Management*, vol. 57, no. 2, Article ID 103174, 2020.

[31] T. Pi, L. Cao, P. Lv, Z. Ye, and H. Wang, "Inferring implicit social ties in mobile social networks,," in *Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, Barcelona, Spain, April 2018.

[32] N. Tahir, A. Hassan, M. Asif, and S. Ahmad, "MCD: mutually connected community detection using clustering coefficient approach in social networks," in *Proceedings of the 2019 2nd International Conference on Communication, Computing and Digital Systems (C-CODE)*, pp. 160–165, Islamabad, Pakistan, March 2019.

[33] V. A. Lewis, C. A. MacGregor, and R. D. Putnam, "Religion, networks, and neighborliness: the impact of religious social networks on civic engagement," *Social Science Research*, vol. 42, no. 2, pp. 331–346, 2013.

[34] M. S. Granovetter, "The strength of weak ties: a network theory revisited," *Sociological Theory*, vol. 1, pp. 201–233, 1983.

[35] J. Gui, Z. Zheng, X. Zhao, and Z. Qin, "Statistical properties and temporal properties of calling behavior," in *Proceedings of the 2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, pp. 1940–1944, Chengdu, China, December 2019.

[36] I. H. Sarker, "A machine learning based robust prediction model for real-life mobile phone data," *Internet of Things*, vol. 5, pp. 180–193, 2019.

[37] M. S. Bernstein, E. Bakshy, M. Burke, and B. Karrer, "Quantifying the invisible audience in social networks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 21–30, Paris France, April 2013.

[38] R. Montasari, R. Hill, V. Carpenter, and F. Montaseri, "Digital forensic investigation of social media, acquisition and analysis of digital evidence," *International Journal of Strategic Engineering*, vol. 2, no. 1, pp. 52–60, 2019.

[39] M. Yip, N. Shadbolt, and C. Webber, "Structural analysis of online criminal social networks," in *Proceedings of the 2012 IEEE International Conference on Intelligence and Security Informatics*, pp. 60–65, Arlington, VA, USA, June 2012.

[40] Z. Zhao, C. Li, X. Zhang, F. Chiclana, and E. H. Viedma, "An incremental method to detect communities in dynamic evolving social networks," *Knowledge-Based Systems*, vol. 163, pp. 404–415, 2019.

[41] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[42] Y. Fang, J. Gao, Z. Liu, and C. Huang, "Detecting cyber threat event from twitter using IDCNN and BILSTM," *Applied Sciences*, vol. 10, no. 17, p. 5922, 2020.

[43] P. Lu, L. Zhang, X. Liu, J. Yao, and Z. Zhu, "Highly efficient data migration and backup for big data applications in elastic optical inter-data-center networks," *IEEE Network*, vol. 29, no. 5, pp. 36–42, 2015.

[44] O. Sharif, M. Hoque, A. Kayes, R. Nowrozy, and I. Sarker, "Detecting suspicious texts using machine learning techniques," *Applied Sciences*, vol. 10, no. 18, p. 6527, 2020.

[45] K. C. Roy, M. Cebrian, and S. Hasan, "Quantifying human mobility resilience to extreme events using geo-located social media data," *EPJ Data Science*, vol. 8, no. 1, p. 18, 2019.

[46] A. Squicciarini, S. Karumanchi, D. Lin, and N. Desisto, "Identifying hidden social circles for advanced privacy configuration," *Computers & Security*, vol. 41, pp. 40–51, 2014.

[47] L. Chen, F. W. Crawford, and A. Karbasi, "Seeing the unseen network: Inferring hidden social ties from respondent-driven sampling," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 1174–1180, Phoenix, AZ, USA, February 2016.

[48] J. Leskovec and A. Krevl, "SNAP datasets: stanford large network dataset collection," 2014, http://snap.stanford.edu/data.

[49] X. Zheng, Y. Wang, and M. A. Orgun, "Contextual sub-network extraction in contextual social networks," in *Proceedings of the 2015 IEEE Trustcom/BigDataSE/ISPA*, pp. 119–126, Helsinki, Finland, August 2015.

[50] A. Madani and M. Marjan, "Mining social networks to discover ego sub-networks," in *Proceedings of the 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, pp. 1–5, Muscat, Oman, March 2016.