Research Article

Building Security Mechanisms for Cross-Border Business Customer Data Analysis Based on Smart Computing

Yicai Li and Lin Shi

Guangdong Food and Drug Vocational College, Guangzhou 510520, Guangdong, China

Correspondence should be addressed to Yicai Li; liyicai@gdyzy.edu.cn

Received 13 January 2022; Revised 20 February 2022; Accepted 24 February 2022; Published 22 March 2022

Academic Editor: Naeem Jan

Copyright © 2022 Yicai Li and Lin Shi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dynamicsocialnetworksalsosufferfromprivacyviolationduetothecontinuousreleaseofdata.Inordertodefendagainstenemyattacks,anewdynamicprivacy-preservingmethodisinnovated,calledDynamic$k^w$Re-Structured Diversity Anonymity ($k^w$-SDA). This method limits the probability of disclosure of node/community identities to $1/k$ when data are continuously published by protecting individuals in groups. The graph based on the previous $w – 1$ is released, missing some graph alterations. The advantage of the proposed method is to save many features of the network while protecting privacy efficiently and effectively.

1. Introduction

Social networking or social network service (SNS) is a technical application architecture under the Web 2.0 system, specifically referring to Internet application services designed to help people build social networks [1]. In recent years, with the success of sites such as Facebook, SchoolNet, and Kaixin, social networking has become increasingly familiar and widespread around the world [2]. Official WeChat data show that at the end of 2013, there were 355 million active WeChat users and more than 6 billion social network accounts worldwide. In August 2014 alone, the number of social media users worldwide exceeded 2 billion [3]. By the end of 2014, there were 1.28 billion mobile phone users in China, with 77% of them accessing social networks. Social networks have become the most widespread and influential communication platform in terms of both domestic and international development. The social network atmosphere is ubiquitous, with mobile socialisation, socialisation of resources, socialisation of entertainment, socialisation of shopping, and socialisation of the Internet, and also, socialisation has become a fundamental application in the Internet [4].

There are currently more than 6 billion social networking accounts worldwide, sending or receiving structured, semistructured, and unstructured data, such as posts, documents, software, images, audio, and video through various social networking applications on a daily basis. The huge social network user base will generate huge amounts of social network information [5]. Sina Weibo, for example, had 143.8 million monthly active users and 66.6 million daily active users in April 2014, with 55 million users posting monthly messages, 57% of which were retrieved and commented on, and an average of more than 100 million tweets per day. This shows that social networks are an important source of big data.

In a narrow sense, social network information is a collection of online information resources, with the characteristics of general online information. In a broad sense, social network information includes not only the network information usually referred to but also the relationship between information production subjects [6].

Social network information is big data, which have four distinctive characteristics of big data:
(1) Large volume of data: The huge community of social network users generates huge amounts of social network information.

(2) There are many types of data: The information shared and exchanged in social networks is rich in semistructured and unstructured information, such as a large number of charts, dynamic sound and images, charming videos, and intricate social relationships.

(3) Low value density and high commercial value: In the huge amount of social network information, a lot of information is after-dinner gossip and infinite repetition, with very little truly valuable information [7].

(4) Fast processing speed (1-second law): That is, getting the information needed in real time from the massive amount of social network information requires highly optimised algorithms to achieve extremely fast data processing speed, which is the most significant feature of Big Data. These privacy-preserving techniques for social network data distribution are mainly aimed at static social networks. There is an urgent need to address the privacy concerns of continuous posting of dynamic social network data.

Figure 1 gives the continuous posting data of a hospital’s dynamic social network. The name of each individual is removed. The disease label indicates the community in which a vertex participates. The two social networks at times $t_1$ and $t_2$ satisfy the dual structural diversity [8], which ensures that for each vertex, at least 1 node in the different communities has the same degree as it. Therefore, using only the data published at $t_1$ or $t_2$, the node degree attack cannot uniquely identify nodes and communities again [9].

However, when publishing data consecutively, there are 2 drawbacks that can be used by the enemy:

(1) Bob is known to have visited the hospital for the first time in the recent past.

(2) Alice is known to have been taken to the hospital several times recently and the number of patients familiar to Alice has increased from 1 to 2. Since A is the only node whose node degree has increased from 1 to 2, Alice can be reidentified based on the 2 data releases, and information about her disease severe acute respiratory syndrome SARS is leaked.

To this end, this article investigates the node identity and community identity disclosure problem when a dynamic social network continuously publishes data under a node-degree attack. For the above time-evolving problem, it is assumed that an adversary can learn the node-degree information of the target victim by monitoring the victim over a continuous period $w$. To protect against enemy attacks, a new dynamic privacy protection method is introduced, called $k^w – S \ DA$. This method limits the probability of node/community identity disclosure to $1/k$ when continuously publishing data by protecting individuals in groups. Then, a scalable heuristic algorithm that can implement the dynamic $(k^w – S \ DA)$ algorithm is proposed. Finally, the effectiveness of this algorithm is verified through simulation experiments.

The arrangement of the paper is as follows: Section 1 consists of introduction. Section 2 explains the tasks related to our topic. Section 3 talks about multiview fusion framework for profiling. Section 4 elaborates the privacy model and problem definition. Section 5 describes the anonymisation algorithm. Section 6 discusses the simulation experiments. Section 7 concludes the article.

2. Related Works

Some of the literatures have investigated the privacy issues when continuous data are published in social networks. O’Leary [10] proposed protecting the label correlation between directly connected nodes. Ooi et al. [11] first proposed a private static model that can protect the identity of nodes in static social networks and then generalized their model to dynamic scenarios by generalizing node IDs. Um [12–15] proposed a privacy-preserving approach based on the GSNN algorithm. It achieves privacy protection by clustering the nodes in the social network and then generalizing the generated clusters within and between clusters to anonymous social network. Hwangbo et al. [16] proposed a trajectory privacy protection method PrivateCheckIn to address the trajectory privacy leakage problem of pseudonymous users in check-in services. Vazquez et al. [17] proposed a privacy protection algorithm with greedy perturbation to ensure that the initial shortest path of the social network remains unchanged and its length is close to that after perturbation. However, most of the above studies do not consider the use of structural diversity to protect privacy. Inadequate consideration of continuous posting of data leaves the opportunity for an enemy to attack and gain privacy by continuously collecting information about the victim and comparing multiple postings of data.

User portraits (persona), i.e., user information labeled data, label user data models based on user behavioural characteristics, social attributes, habits, consumption behaviour, job nature, and other information, are virtual representations of real users in the digital world. It helps enterprises to finally locate the user population and specify precise management strategies, thus improving the daily management of enterprises and enhancing the user business experience in all aspects.

At present, user profiling technology has been widely used and developed in many fields, such as online marketing, social media, mobile phone bills, and e-commerce. Sheikh et al. [18] proposed a method to build a context-aware portrait of mobile users by analysing user profile log data. Yang [19] found similar users of smart terminals by analysing the situational behaviour patterns of smart terminal users and construct a sparse situational portrait of users with historical logs. Zafar et al. [20] constructed evolutionary models based on users’ social network relationships and interests to depict user portraits under dynamic evolutionary conditions of social elements. Tang and Zhang [21] proposed a multilabel classification method to predict the gender and age of users. Qiu constructed a twitter user profile based on
2.1. Key Models for the Commercial Exploitation. In recent years, the business value of social network big data has become increasingly important to entrepreneurs or for-profit organisations due to its low acquisition cost, high marginal revenue, high timeliness, circle effect, and viral spread effect. Through the commercial exploitation of social network big data, such as collection, analysis, and processing, it is possible to tap and grasp users’ structure, interests, value orientation, demand trends, behavioural habits, personality traits, interpersonal relationships, action trajectories, and so on, in order to carry out social network-based online marketing and personalised information recommendation services. There are five main modes of commercial exploitation of social network big data, and the process of implementation is shown in Figure 2.

The operator model refers to operators providing social network application services (e.g., Sina and Tencent) using their own position as a bridge between social network information providers and recipients to mine, analyse, and represent the relationships between knowledge resources and their owners using visual graphics on the collected network operation data, build a knowledge graph, and depict the correlation and interaction between social network users and the entire social network. The knowledge graph is used to depict the correlation and interaction between social network users and the whole social network characteristics and to analyse and discover the attributes, interests, and hobbies of user groups in the social network. In the era of big data processing, knowledge graph, with its quantitative and visual analysis means, provides a breakthrough for network service providers to tap into social network resources and is a powerful tool to open up this treasure.

The profit-driven model refers to businesses, organisations, or individuals who, in pursuit of the huge commercial value contained in social network big data, adopt all technologies and means to recklessly collect, disseminate, sell, analyse, and process big data in social networks under the drive of profit. In this exploitation model, the temptation of profit, coupled with the weak awareness and skills of social network users in security protection, weak government supervision, insufficient industry self-regulation, lack of information ethics and morality of developers, and deficiencies in relevant laws and regulations, has led to this model being the hardest hit area for infringement of personal privacy in the commercial exploitation of big data in social networks at present. In recent years, some illegal business organisations have been tempted by the huge potential business opportunities of social networks to illegally collect, mine, and exploit social network big data, effectively obtaining an image metadata tagging and other means. Theft of personal information (e.g., mobile phone numbers, passwords, e-mail accounts, address books, and real personal information) from social networks to spread online rumours and bad online culture commit online fraud and theft.

3. Multiview Fusion Framework for User Profiling

3.1. Dual Channel Modelling. The information of the social business customers of StateNet is the basic data; users change their travel mode for each trip in response to the change of the target location and choose the recommended hotels according to the target location in accordance with the target distance, star rating, rating, comfort, etc. Through statistics, it is found that there are more users who check the standard few times throughout the year, and the corresponding sensitivity accounts for a smaller proportion, so they are defined as low active users; while there are fewer users who check the standard many times, the sensitivity accounts for a larger proportion, so they are defined as high active users. The distribution of low- and high-activity users in customer information is shown in Figure 3.

There are obvious differences between the characteristics of low active users and highly active users, mainly in the following two aspects:

(1) Different perspectives on user profiling: For low-activity users, the focus is on the standard key content, while for high-activity users, the system focuses on relevant links such as search frequency and browsing time.

(2) Different ways of constructing user characteristics: Low-activity users have only a few search records with little corresponding data items, while high-activity users have multiple corresponding data items.
records with multiple values for the same item. Based on the different portrayal perspectives and the different ways of constructing user characteristics, the low and high active users are modelled independently and predicted separately.

3.2. Building a Multisource User Profile System. Feature clusters are built around three different dimensions of social business customer information: business approval information, business control information, financial settlement information, and around user category features, numerical features, clustering features, text features, temporal features, and statistical features.

User consultation business types reflect the different user needs of social business customer information, such as business control, ticket booking, expense reimbursement, tariff consultation, and business approval. Statistical analysis of the urban and rural categories shows that users in the city centre and mountainous areas behave differently in terms of sensitivity characteristics.
For highly active users, the Bag-of-words model, a common method of document representation in information retrieval, is used to characterise them. In information retrieval, a document is regarded as a collection of words. Each word existing independently of the other, and a set of unordered words is used to represent the document, ignoring the syntax and order of the text. It is necessary to consider not only whether the sample has a category X (Has_X) but also the proportion (Ratio_X) and number (Count_X) of the corresponding category X.

Cluster analysis refers to the process of analysing multiple classes consisting of objects from a collection of physical or abstract objects, measuring the similarity between data sources and grouping similar data sources into the same clusters. For example, after statistics found that there are fixed coding rules for unit code fields, clustering feature analysis is performed on unit codes according to the coding rules to summarise the class characteristics.

A total of 174 codes were resolved by intercepting the user unit code fields according to their lengths, including 1 quaternary code of length 11; 96 tertiary codes of length 9; 75 secondary codes of length 7; and 12 primary codes of length 5. 3 digits formed the user category, and the first digit was used for clustering features. In clustering analysis, the degree of similarity between two instance points in the feature space is reflected by the distance between the two instance points, and the distance between points in the high-dimensional vector space can be generalized to the Lp distance formula for solving, which is divided into different forms at different orders, number into different forms:

\[
L_1(x_i, y_j) = \sum_{i=1}^{n} |x_i^{(1)} - x_j^{(1)}| \\
L_2(x_i, y_j) = \left( \sum_{i=1}^{n} |x_i^{(1)} - x_j^{(1)}| \right)^{1/2}
\]

For highly active users, the Bag-of-words model, a common method of document representation in information retrieval, is used to characterise them. In information retrieval, a document is regarded as a collection of words. Each word existing independently of the other, and a set of unordered words is used to represent the document, ignoring the syntax and order of the text. It is necessary to consider not only whether the sample has a category X (Has_X) but also the proportion (Ratio_X) and number (Count_X) of the corresponding category X.

Cluster analysis refers to the process of analysing multiple classes consisting of objects from a collection of physical or abstract objects, measuring the similarity between data sources and grouping similar data sources into the same clusters. For example, after statistics found that there are fixed coding rules for unit code fields, clustering feature analysis is performed on unit codes according to the coding rules to summarise the class characteristics.

A total of 174 codes were resolved by intercepting the user unit code fields according to their lengths, including 1 quaternary code of length 11; 96 tertiary codes of length 9; 75 secondary codes of length 7; and 12 primary codes of length 5. 3 digits formed the user category, and the first digit was used for clustering features. In clustering analysis, the degree of similarity between two instance points in the feature space is reflected by the distance between the two instance points, and the distance between points in the high-dimensional vector space can be generalized to the Lp distance formula for solving, which is divided into different forms at different orders, number into different forms:

\[
L_1(x_i, y_j) = \sum_{i=1}^{n} |x_i^{(1)} - x_j^{(1)}| \\
L_2(x_i, y_j) = \left( \sum_{i=1}^{n} |x_i^{(1)} - x_j^{(1)}| \right)^{1/2}
\]

Demand varies seasonally and over time, with large differences in the number of business-sensitive users at different times of the year. The statistical results of the number of high- and low-sensitivity users are shown in Figure 4.

In Figure 4, the horizontal axis represents the number of days in a year, the vertical axis represents the number of people, and the sensitive users are shown as curves with diamond boxes, while the nonsensitive users are shown as curves with squares. By looking at the annual distribution of the number of high and low sensitive users, we find that there are few sensitive users from April to July, while the number of sensitive users increases sharply in January, and the number of low sensitive users and high sensitive users changes regularly in cycles of about 20 days. Considering the importance of time-factor features in the user feature system, for low active users, two granularity categories of year and month are constructed for feature analysis, whereas for high active users, three granularity Bag-of-words features of month-day and hour are constructed as well as numerical features.

3.3. Multiview Fusion Model. Constructing data features with multiple perspectives to build a multisource feature system for the State Grid social commerce user information table. A multiview fusion model based on a two-layer XGBoost is proposed to effectively use multisource features and solve the problem of high-dimensional features. XGBoost is an optimised boosted tree model with the advantages of high task generality and fast running speed and is widely used in data mining and machine learning. The model structure is shown in Figure 5. In this model, start from user of two types, i.e., low and highly active users, and then enter into the multifeature system. After this, XG boot system start working and working on different trees results in form of bagging and analyse the low and highly active users as sensitive users.

4. Privacy Model and Problem Definition

In this article, dynamic social networks are modelled as a time-stamped graph. Specifically, \( G_t (V^t, E^t, C^t) \) denotes the dynamic social network at time t, \( V^t \) denotes the set of nodes corresponding to individuals, \( E^t \) denotes the set of edges between individuals, and \( C^t \) denotes the set of communities.
4.1. Multicommunity Identity Privacy Protection. For the basic case when \( w = 1 \), the background knowledge for enemy processing includes the data release \( G' \) and the node degree \( d_v \). In order to protect the multi-community identities of nodes from node degree attacks, it is necessary not only for many nodes to share the same node degree but also for these nodes to protect their respective community identities from each other. Define a k-heavy protection group in which nodes have the same node degree and can protect their respective multicommunity identities from each other.

An example of the above definition is given in Figure 6(a). The number of each node indicates the multicommunity identity of the node, i.e., \( C^{t}_{A} = \{1, 3\} \), \( \theta^{t}_{v} = \{B, C, E, F, G, H\} \) indicates a 4-guard group. Since in a subset \( \theta = \{B, E, F, G\} \) of size 4, the multicommunity identities of any two nodes are disjoint sets.

At least \( k \) nodes in a k-heavy guard group \( \theta^{t}_{v} \) have the same node degree. Therefore, the probability of identifying a node from \( \theta^{t}_{v} \) with a known node degree \( d \) is limited to \( 1/k \). Furthermore, there are at least \( k \) nodes in k-repopulation \( \theta^{t}_{v} \) in which the multicommunity identities of any pair of nodes do not intersect. Therefore, the probability identity of a single node is compromised and is also limited to \( 1/k \).

4.2. Dynamic Privacy Model. The concept of k-reprotection groups can now be extended to dynamic scenarios when \( w > 1 \). In order to avoid privacy breaches when successive data releases occur, anonymity graphs \( G'_{t} \) should be generated based on previous data releases \( G'_{t-1}, G'_{t-2}, \ldots, G'_{t-w+1} \) so that the enemy no longer has the knowledge advantage resulting from previous data releases. In other words, to protect the node degree at a given moment, it is necessary to protect the sequence of node degrees at time \( w \) against the node degree sequence information held by the enemy. Similar to the concept in the base case, in order to prevent the use of \( \Delta^{w} \) to obtain node identities, there should be multiple nodes with the same node degree during \( w \). To protect against multiple community identities, these nodes should participate in unrelated communities at each time. Using these factors as a basis, we first define k-heavy protection consistency groups and then propose a new privacy model, namely, a dynamic \( k^{w} \) structural diversity model.

5. Anonymization Algorithms

The approach in this article consists of two parts:

1. Construction and maintenance of the CS table
2. Anonymization process

First, in order to avoid searching for similar nodes for \( w \) distribution for the sake of canonization and reducing information distortion, the CS table is constructed, and the nodes are sorted according to their node degree sequence. Thus, after the node degree of the current table is attached, the CS table can be maintained efficiently by reordering some of the smaller sets of nodes. Thus, after the CS table has been updated at each moment according to the current graph, the nodes can be anonymous one by one according to their ordering in the CS table. During the accommodation phase, the CS table is used to quickly find nodes with a similar sequence of node degrees. In particular, for node \( v \), two randomisation methods are considered. One is to make \( v \) belong to one of the current k-heavy guard consistency groups. In this case, because the CS table places the nodes with the most similar node degree sequence nearby, node \( v \) is assigned to the closest k-heavy guard consistency group in the CS table. Another approach is to create a new k-heavy guard consistency group using V and the other \( k-1 \) nodes.

Specifically, it is known that \( G^{1}_{t} \), the CS table, includes \( G^{1} \) nodes in descending order of node degree. This requires that all nodes be sorted. When anonymizing \( G^{1}_{t} \), we change the
CS table at the same time. After anonymisation, each node belongs to a k-reguard consistency group. Once $G^2$ is known, we append the node information for $G^1$ to the corresponding record. From now on, instead of sorting all nodes, only nodes within the same group are sorted, since the nodes are already sorted in descending order according to their node degree. Thus, the sorting time is reduced $G^2$. After anonymisation, each node is in the k-reguard consistency group $\Theta^k_G$. For subsequent input snapshots, similar steps can be performed until $G^w$ is anonymized.

$$\text{mergecost}_E(v, D^w) = \begin{cases} 
0, & \text{if } D^{[1, \ldots, 1]} = D^{[1, \ldots, 1]} - D^{[1]} \leq R_v; \\
D^{[1, \ldots, 1]} - D^{[1]} - R_v, & \text{if } D^{[1, \ldots, 1]} = D^{[1, \ldots, 1]} - D^{[1]} \leq R_v + N_v; \\
N_v, & \text{if } D^{[1, \ldots, 1]} = D^{[1, \ldots, 1]} - D^{[1]} - D^{[1]} \leq R_v + N_v; \\
\infty, & \text{otherwise,}
\end{cases}$$

where $\Delta [a, b]$ denotes the sequence of nodes at time $[a, b]$. MergeCost$_E(v, D^w)$ denotes the number of additional edges required to protect the nodes in the current k-heavy guard consistency group $\Theta^k_G$. For the second anonymisation method, let $U$ denote the set of nodes of size $k$ and the multicomunity identities of any pair of nodes in the set do not intersect at time $w$. The cost of anonymisation is defined as

$$\text{CreateCost}_E(v) = \min_{u \in U} \sum_{w \in U} \text{MergeCost}_E(u, \Delta^w_v).$$

CreateCost$_E(v)$ for a new k-reguard consistency group to be created when the sequence of nodes is $\Delta^w_v$. When only adjusting and adding edges does not protect $v$ effectively, the AddingVertex operation should also be performed.

In summary, the general steps of the algorithm in this article are as follows: First, a CS table is created to summarise the node information from the previous data release. Then, for the current graph, the nodes are anonymous according to their sequence in the CS table. Specifically, the minimum MergeCost$_v$ and CreateCost$_v$ are calculated for the highest ordered node $v$ in the CS table that has not yet been anonymously to determine what method can be used to minimise changes to the set of nodes and to anonymous $v$. If the two costs are equal, the minimum MergeCost$_v$ and CreateCost$_v$ are calculated. If the two costs are equal, calculate the minimum MergeCost$_E$ and CreateCost$_E$ to determine the method that minimizes information distortion and anonymity $v$. If the two costs are equal, calculate the minimum MergeCost$_E$ and CreateCost$_E$ to determine the method that minimizes information distortion and anonymity $v$. Note that the CS table arranges the nodes in descending order of node degree sequence.

6. Simulation Experiments

A partial DBLP data set was used for performance evaluation. The data set consists of 94,414 nodes and 436,002 edges; nodes correspond to the authors who have published in 55 conferences such as AAAI, ACM Multimedia, and WWW, and edges indicate coauthorship. The 3 conferences are most commonly used
by authors to publish are seen as multicomunity attributes of authors because authors generally only publish articles related to their own research area. In order to simulate a dynamic scenario, data before 1991 is used as the initial version of the network, data before 1992 as version 2, and so on until 2011. Thus, the data set has a total of 21 consecutive data releases.

6.1. Performance Analysis of Anonymized DBLP with Different Values of $k$. The performance of this algorithm is described in terms of average shortest path length (ASPL), clustering coefficient (CC), and node degree centralization (DC) for different privacy levels $k$. The $x$-axis in all plots represents time, where the dynamic graph is anonymity according to the current graph and the previous 9 releases, i.e. $w = 10$. Because there are 21 consecutive data releases in the test data set, the $x$-axis in each plot has 12 time stamps.

Figure 7(a) shows the ASPL values for the initial DBLP graph and the anonymity DBLP graph. The trends of the curves are similar for the different $k$ values, with little difference between the curves. This is because the randomisation depends on both the current graph and the same number of previous data releases. Also, note that the information distortion is smaller for smaller $k$ values. However, ASPL values for larger $k$ values are comparable to the results for smaller $k$ values because the accommodation also depends on previous data releases.

Figure 7(b) shows the CC values of the initial DBLP graph and the anonymity DBLP graph. K values are different and all CC values deviate slightly from the original values, indicating that the CC values are well preserved by this method. Figure 7(c) shows the DC values of the DBLP data set. DC values of the anonymity curves are almost identical to the initial graph. The reason for this is that the algorithm minimizes the variation of nodal degree during successive
canonization processes. When the value is given, the smaller the $k$ value, the smaller the information distortion tends to be. All the above results show that the algorithm in this article is able to preserve the graphical features under various values of $k$.

6.2. Analysis of the Defense Effect of the Algorithm under Different Values of $k$. Figure 8 shows the comparison of the effectiveness of the proposed algorithm with the MR algorithm under different $k$ values. Here, the roentgenography rate is used as a performance indicator to measure the effectiveness of the algorithm. As can be seen from Figure 8, as the privacy level $k$ increases, the roentgenography rate of both algorithms becomes smaller and the degree of privacy leakage decreases step by step. The node degree of this algorithm changes less during the continuous accommodation process and contains more similar nodes, which make the attacker’s probability of finding the target node much lower. The MR algorithm is based on the value domain mapping, which has a limited degree of spatial division and contains a smaller number of similar nodes; therefore, the defense is less effective.

6.3. Comparative Experiments. The gradient boosting decision tree (GBDT), linear logistic regression (LR), decision tree (DT), and support vector machine (SVM) models were all evaluated with the same training features and other conditions. The experimental results of each classification model are shown in Table 1.

From the comparison experiments, it can be seen that the XGBoost model algorithm performs best in identifying high-activity users and low-activity users, and in order to verify the algorithmic advantages of the multiview fusion model, a comparison is made with other commonly used fusion models.

7. Conclusion

Based on the analysis and mining of web social commerce user data, we propose constructing a multiview fusion framework model for user profiling, which can quickly and accurately identify social commerce sensitive users. A new privacy model is proposed to protect the node and multi-community identities of each individual in the continuous posting of data in dynamic social networks. By analysing the characteristics of social commerce users, a dual-channel modelling algorithm is used to predict low-activity users and high-activity users separately, and multiple feature data extraction methods are proposed to construct a multisource feature system for social commerce users; a two-layer XGBoost-based multiperspective fusion model is proposed to solve the problem of high-dimensional features with multisource features, and the effectiveness of the method is demonstrated through experiments. The algorithm can effectively protect privacy while retaining most of the features of dynamic social networks, providing an important reference basis for the accurate identification of sensitive social business users.
Data Availability
The data sets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest
The author declares no conflicts of interest.

Acknowledgments
This research has been financed by Health Commission of Guangdong Province Research on Innovation Chronic Disease Management Model of Home Medical Devices Based on the Perspective of “Internet Plus” (B2020203) and School Level “Design and Research of Community Elderly Based on the Perspective of “Internet Plus” (B2020203) and Disease Management Model of Home Medical Devices Guangdong Province” Research on Innovation Chronic


References


