Received April 25, 2020, accepted May 9, 2020, date of publication May 19, 2020, date of current version June 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2995568

Multi-View Low-Rank Coding-Based Network Data De-Anonymization

XINGPING XIAN[1], TAO WU[2], (Member, IEEE), SHAOJIE QIAO[3], (Member, IEEE), WEI WANG[4], YANBING LIU[5], AND NAN HAN[6]

[1]Department of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2]School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[3]School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China
[4]Institute of Cybersecurity, Sichuan University, Chengdu 610065, China
[5]Chongqing Engineering Laboratory of Internet and Information Security, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[6]School of Management, Chengdu University of Information Technology, Chengdu 610103, China

Corresponding authors: Tao Wu (wutaoadeny@gmail.com) and Shaojie Qiao (qiaoshaojie@gmail.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0904900 and Grant 2018YFB0904905, in part by the National Natural Science Foundation of China under Grant 61802039, Grant 61772091, Grant 61772098, and Grant 61802035, in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJQN201800630, and in part by the Innovative Talents Program under Grant BYJS201811.

ABSTRACT Social networks are extensively exploited by third-party consumers such as researchers and advertisers to understand user characteristics and behaviors. In general, before network data is published, sensitive relationships should be anonymized to prevent the compromise of individual privacy. To quantify the guarantee level of privacy-preserving mechanisms and mitigate users' privacy concerns, numerous studies concerning network data de-anonymization have been carried out. However, most existing studies focus on single-view data, and privacy protection for multi-view data that is ubiquitous in the era of big data has not been yet extensively explored. In this study, we are interested in answering the following question: Are the traditional privacy protection methods still valid for the anonymization of multi-view data? In this study, we propose a Multi-View Low-Rank Coding (MVLRC) based network data de-anonymization framework to assess the vulnerability of privacy protection techniques by accurately reconstructing a large portion of the original data. Specifically, the framework assumes that in principle, the target and auxiliary networks have common structural patterns, and they can be modeled together to infer the hidden structure of the target network. The essential components of our work include the following: (1) a robust network representation model for structural pattern learning; (2) the network representation based multi-view modeling of target network and auxiliary network; (3) the inference of the anonymized links via target network reconstruction. Experimental results on synthetic networks and three real-world networks demonstrate that auxiliary networks can be utilized by malicious adversaries for privacy inference attacks. Thus, the privacy protection of multi-view network data needs more sophisticated anonymization techniques.

INDEX TERMS Privacy preserving, de-anonymization, network data, multi-view learning, low-rank coding.

I. INTRODUCTION

The development of social networks, including online social networks, mobile social networks, vehicular social networks, etc., has led to a tremendous explosion of network data [1]. The unprecedented increase of network data provides a wonderful opportunity for both academia and industry to conduct appealing network theories and applications. For example, researchers design structure mining and information cascade models with the assistance of the data extracted from social network sites [2]–[4]; platforms deliver traffic updates or trip recommendation to users based on location-based services (LBS) [5]; employers find targets of recruitment advertising and position candidates from the professional experiences in LinkedIn [6]. Recently, privacy-preserving has gained popularity in many areas [7]–[10]. In practice, and more specifically in social networks, many kinds of individual information such as sexual contacts, purchase records,

The associate editor coordinating the review of this manuscript and approving it for publication was Jun Hu.

financial relationships, etc., are often considered to be privacy data. Accordingly, releasing the data collected from social networks would directly compromise users' privacy. Therefore, privacy preserving has become an indispensable task for network data publication [11].

A. LINK PRIVACY-PRESERVING

Because privacy data coexists with public data in social networks, there are three important privacy risks with social network data being published: content disclosure risk, identity disclosure risk, and link disclosure risk [12]–[14]. Link disclosure is the focus of our study, which refers to inference about the existence of a sensitive link or relationships between two individuals. For example, in financial transaction networks, email networks, and professional social networks, the existence of money transactions, private emails, and friendship between two individuals that are considered as sensitive links may be inferred based on the published data [15].

Recently, many anonymization methods have been proposed to limit link disclosure. The first group of methods achieve link anonymization by transforming networks to have some subgraph similarity, such as the k-isomorphism based method [16] and equivalence class partition based method [17]. The second group of methods partition the network into clusters, and each cluster is collapsed into one super-node thereby, hiding sensitive links [18], [19]. The third group of methods are based on random link perturbation, including adding non-existing links, deleting existing links, and switching links [14], [15], [20], [21].

B. LINK PRIVACY ATTACK

As an addition to data anonymization research, de-anonymization techniques have been proposed in the literature [22]–[24], which can be used to assess the privacy risk of anonymization techniques and probe their potential drawbacks. To infer sensitive links from the anonymized networks, subgraph attack methods [25] and semantics based methods [26] have been proposed. Moreover, in the link prediction problem [27], the existence of missing links can be estimated based on the intrinsic topological features of observed networks. Accordingly, various methods proposed for link prediction could be exploited by adversaries for network de-anonymization. For example, similarity based methods were used by attackers to infer the sensitive relationships of users, and reconstruction based de-anonymization methods have been proposed to recover the original networks according to the structural patterns of anonymized data [28]–[31].

TABLE 1. Different cases for multi-view data anonymization.

	Case 1	Case 2	Case 3	Case 4
View I Anonymization	×	×	√	√
View II Anonymization	×	√	×	√

C. MOTIVATION

Typically, in the era of big data, there are rich diversities, and the same object can be observed from different viewpoints or captured by distinct apparatus. Different views that are complementary to each other form the basis of multi-view learning. Specifically, each view of the data contains some specific information that others do not have. Thus, multiple view models can be applied to represent the data comprehensively [32]. For example, for authors in the scientific community, their relationships can be characterized based on co-authorship or citations. Another example is the real social network of individuals, where various social media applications (e.g, Sina Weibo, Wechat) capture the interactions between users from different viewpoints.

In this study, we explore the problem of network data de-anonymization from the perspective of multi-view learning and determine the possibility of anonymized links inference. In multi-view networks, existing anonymization techniques assume that it is enough to anonymize each of their views independently. Let us consider the simplest case in which published data includes only two views. We will then have options to anonymize the data as shown in Table 1: no anonymization for either view, anonymization for one view, and anonymization for both. With two views as shown in Table 1, case 4 is the backbone of the anonymization techniques for data publishing, which is clearly the strongest protection of privacy. Here, we are interested in answering the following research question: Is case 4, the strongest among four cases, sufficient for anonymizing multi-view network?

D. CONTRIBUTIONS

In this study, we seek to answer the question by taking an adversary approach to assay the privacy level of anonymized multi-view networks. The idea behind this is that the anonymized links of existing methods are generated without considering the structural patterns of the target network, and the resulting local subgraphs have inconsistent structural patterns compared with the normal ones [33]. In addition, we assume that the auxiliary networks have consistent structural information with the target network, and they can be utilized for structural patterns learning. Thus, we propose a novel network data de-anonymization framework, called Multi-view Low-rank Coding (MVLRC), to model the target network and auxiliary network together. Based on low-rank theory [34], we define a low-rank constrained network representation model and uncover the anonymized links by exploring the representation relationship among elemental subgraphs. The key contributions of this study include:

- We answer the problem of whether the traditional privacy protection methods are still valid for the anonymization of multi-view data and formulate the privacy-preserving oriented multi-view network de-anonymization framework. To the best of our knowledge, it is the first time that the privacy protection of multi-view network data has received attention.

- We develop the multi-view low-rank coding method MVLRC for network de-anonymization, in which the auxiliary network can be incorporated naturally with the target network for anonymized links inference.
- The promising accuracy of MVLRC demonstrates that the representative network anonymization approaches cannot be directly applied on multi-view network data. This observation will help researchers in designing multi-view network data anonymization methods by taking network structural patterns into consideration.

E. ORGANIZATION

The remainder of this paper is organized as follows: In Section 2, we introduce the related work. Section 3 presents the preliminaries and problem formulation, followed by the proposed MVLRC algorithm and the derivation of the optimal solutions. Experiments are reported in Section 5. The last Section concludes the study.

II. RELATED WORK

A. NETWORK PRIVACY PRESERVATION

When releasing data for analysis, privacy preserving of individuals has recently raised great concern in the data mining field. The main concern is that sensitive information should not be disclosed. There are different types of privacy models proposed for preserving data privacy such as K-anonymity, L-diversity, T-closeness, and differential privacy. The methods should limit the disclosure risks while maintaining the utility of the data. More details about existing research outputs and achievements of the privacy field can be found in [35]. However, most existing studies on preserving privacy in data publishing have focused on tabular data. Owing to the dependency and complexity of network data, privacy preserving about social network data is much more challenging than the anonymization of the conventional tabular data, and the anonymization techniques for tabular data cannot adapt to network data [12].

Generally, for network data publication, there are three important privacy risks: content disclosure risk, identity disclosure risk, and link disclosure risk [13], [14]. A large portion of studies on social network privacy has concentrated on identity disclosure, which reveals users' identifiable personal information (such as names and social security number) based on the structure features or descriptive attributes [22], [36]–[38]. Although a privacy-protection mechanism for social network data publishing should consider content, identity, and link disclosure threats, link disclosure often leads to both content and identity disclosures. Therefore, limiting link disclosure is more fundamental than the others.

Many anonymization methods have currently been proposed and can be categorized into two groups, i.e., generalization based approaches and perturbation based approaches. Specifically, the basic idea of generalization based methods is to replace the sensitive information with a less specific, but semantically consistent value [39]. Perturbation based methods include link modification strategy and randomization strategy, in which the former proposes link addition and deletion mechanism to meet the desired constrains, such as k-degree anonymity [40], and k-automorphism anonymity [41]; the latter attempts to change network structure by randomly adding and removing links. In addition, differential privacy methods [42], [43] are also proposed for network data anonymization.

Compared with the large number of generalization based approaches [19], [44] and k-anonymity based methods [16], [41], [45] proposed for user identity anonymization, the research on link privacy protection is insufficient. Initially, Zheleva and Getoor [18] focused on the problem of preserving the privacy of sensitive relationships in network data and defined the sensitive relationships inference problem based on anonymized network. Thereafter, the most conservative approach for link privacy protection was proposed to remove the sensitive relationships altogether, thus preserving any privacy that these relationships may compromise [18]. Moreover, the works [20], [46] presented random perturbation and random switching strategies for link privacy protection. Along this line, considering the structural proximity of nodes, some structure-aware randomization perturbation methods have been proposed, including the local perturbation based methods [14], [15] and the random walk based method [47]. Furthermore, the Gaussian noise based method [48] and differential privacy methods [49]–[51] were developed for link privacy protection.

B. NETWORK DATA DE-ANONYMIZATION

Network de-anonymization techniques are actively studied to explore the vulnerabilities of current network data privacy protection mechanisms. Most of the existing de-anonymization attacks focus on user identity de-anonymization. Typically, these approaches can be classified into two categories, attributes based identity de-anonymization as well as structure based identity de-anonymization. Recently, there has been a surge of interest in the topic of identity de-anonymization by involving the attributes information of users. Most of the methods extract features from public profile fields, such as user ID, location, etc., and content information, e.g., timestamps, geo-tags, etc. Thereafter, these methods adopt classifiers to infer whether the node pairs correspond to a similar identity [52]. Important studies on this topic are reviewed by Shu *et al.* [53].

The structure based identity de-anonymization, also called vertex re-identification, assumes that the accounts belonging to the same user across social networks have similar local structures. Thus, the subgraphs associated with target nodes can be used as background knowledge for user identification. Specifically, Nilizadeh *et al.* [37] matched the anonymized network with auxiliary network and identified user identity by considering the community structure. Lee *et al.* [54] incorporated multi-hop neighbors' information in network structures as novel features and optimized the matching for users between the anonymized network and the auxiliary network by leveraging a machine learning technique. Narayanan and Shmatikov [55] proposed a network topology based de-anonymization method that first identifies some seed nodes and then propagates the mapping to new nodes based on structure similarity. To match the nodes accurately, Ji *et al.* [56] defined a unified similarity measurement and proposed a de-anonymization framework based on it. Ji *et al.* [57] implemented comprehensive quantification of de-anonymizability of networks with seed information and provided theoretical foundation for structure-based de-anonymization attacks. Zhou *et al.* [58] proposed a cross-platform unsupervised user identification algorithm based on friend relationships. The above works demonstrate that privacy-preserving on network structure is necessary for the anonymization of user identity.

Besides identity de-anonymization, the link de-anonymization, i.e., link disclosure or link re-identification, which aims to identify sensitive relationships among users from anonymized networks, is also an important issue in the field of network privacy protection. Specifically, Ying and Wu [59] investigated the sensitive relationships protection problem and verified the value of similarity measures for link privacy breaching. To mitigate the vulnerability of network anonymization mechanisms, Zhang *et al.* [28] developed an enhanced network anonymization method by generating fake edges as plausible as possible. Fire *et al.* [29] presented a classifier based link reconstruction attack method to identify sensitive relationships. Wu *et al.* [30] defined a low-rank approximation based de-anonymization algorithm to reconstruct a network from link randomized observation. Vuokko and Terzi [31] proposed a maximum-likelihood-estimation-based method to reconstruct the original networks. These methods are mostly based only on the structural features of the anonymized network. In our work, we investigate how to utilize multi-view features from the target network and auxiliary network for network link de-anonymization.

C. MULTI-VIEW LEARNING AND PUBLISHING

Owing to the diverse domains and various feature extractors, multiple groups of features are currently available for specific learning problems, and each of them can be regarded as a particular view. Accordingly, multi-view learning paradigm is developed to exploit the useful information from different views. The existing multi-view learning algorithms can be classified into three groups [60]: 1) co-training, 2) multiple kernel learning, and 3) subspace learning. Importantly, co-training algorithms enhance the learning performance in different views by using the information from one another; multiple kernel learning algorithms define a kernel function for each view and thereafter combine the kernels together to improve learning performance; subspace learning algorithms aim to find a meaningful low dimensional embedding or latent subspace shared by all feature sets. Generally, existing multi-view learning methods mainly aim to maximize the agreement on multiple distinct views or exploit their complementary information and ensure their success. Recently, many multi-view algorithms have been proposed by taking into consideration the complementary information from different views, such as clustering [61] and subspace learning [62].

For privacy-preserving data publishing, Dou and Coulondre [63] presented a formal analysis of privacy violation in the context of multi-view tabular data. Yao *et al.* [64] defined k-anonymity based on relational view and concentrated on how to detect whether or not a given set of releasing views violates k-anonymity. In our study, we also analyze the privacy risk of multi-view data. However, our work aims to explore the network structure de-anonymization problem, which is different from the existing ones that mainly focus on tabular data.

III. PRELIMINARIES AND PROBLEM

Typically, to anonymize networks for publication, the sensitive relationships contained in original graphs are removed firstly, and then the anonymization strategies, such as sparsification, perturbation and switching methods, are applied to add or remove network links. To quantify the guarantee level and assess the privacy risk of state-of-the-art anonymization strategies for multi-view networks, the core task of this study reduces to recover the original social graph and identify anonymized links as accurate as possible based on target graph. Based on the recovered graph, the sensitive relationships can be accurately inferred with subgraph attacks, similarity measures, etc. For simplicity, this paper assumes that only one auxiliary graph is available for structure de-anonymization.

A. DEFINITIONS

Definition 1 (Original Graph): A social network can be modeled as a graph $SG = \{U, R\}$, in which U denotes the set of users and $R \subseteq U \times U$ indicates the set of relationships between the users. If the associated parties prefer to keep link $R_{i,j}$ in graph SG hidden, then this link $R_{i,j}$ is the sensitive relationship, and the graph SG can be regard as the original graph that needs privacy-protection.

Definition 2 (Target Graph): For a social graph $SG = (U, R)$ containing sensitive relationships, its structure is always modified based on a certain anonymization strategy $I(\cdot)$ to preserve privacy before publishing. We refer to the published social graph as the target graph $SG^T = \{U^T, R^T\}$, where U^T is the user set, $U^T = U$, and R^T is the relationship set, $R^T \neq R$.

Definition 3 (Anonymized Link Set): For a social graph SG, the difference between the link sets of SG and its target graph SG^T is defined as anonymized link set $\Re$, $\Re = R \backslash R^T$.

Definition 4 (Auxiliary Graph): For a target graph $SG^T = \{U^T, R^T\}$, if there is a published social graph $SG^H = \{U^H, R^H\}$ that describes a set of relationships on the same set of individuals with SG^T from a different viewpoint, i.e., $U^H = U^T$, and $R^H = \{R^H_{i,j} | i \in U^H, j \in U^H\}$ where $R^H \neq R^T$, the graph SG^H is defined as the auxiliary graph of SG^T.

TABLE 2. Notations and meanings.

Notations	Descriptions
SG	Original social graph
SG^T	Target graph, i.e., anonymized original graph
SG^H	Auxiliary graph corresponding to target graph
$\Re$	Anonymized link set
$I(\cdot)$	Anonymization strategy
$\Gamma(\cdot)$	De-anonymization algorithm
k	Anonymization coefficient
$A^{(i)}$	The adjacency matrix of network
$E^{(i)}$	The matrix denoting the fake links of $A^{(i)}$
λ, α	The trade-off parameter
$\hat{X}$	Representation matrix of social graphs
$X^{(1)}$	Representation matrix of anonymized SG^T
$X^{(2)}$	Representation matrix of anonymized SG^H

Problem Statement: For an original graph SG, given the target graph SG^T and its related auxiliary graph SG^H, the goal of the network structure de-anonymization is to develop an algorithm $\Gamma(\cdot)$ to generate a de-anonymized graph $SG^D = \Gamma(SG^T, SG^H)$, thereby approximating the original graph SG as much as possible.

The description of the notations used in this study is presented in Table 2.

B. NETWORK ANONYMIZATION MODEL

To evaluate the performance of the de-anonymization attack, we consider popular anonymization techniques that have been most widely used in structure anonymization works [12], [36] [20] [54] [65] . Specifically, the selected anonymization strategies are introduced as follows:

(1) Densification. This method ensures the anonymity of graph $SG = \{U, R\}$ by only adding $k|R|$ links randomly where k is the anonymization coefficient.

(2) Sparsification. This method obtains the anonymization of social graph $SG = \{U, R\}$ by randomly eliminating $k|R|$ links.

(3) Perturbation. This method first removes $k|R|$ links from a social graph $SG = \{U, R\}$ in the same way as the sparsification method does. Thereafter, it adds random false links until the number of links in the anonymized graph is the same as the original one.

(4) Switching. This method selects two random edges (i_1, j_1) and (i_2, j_2) from social graph $SG = \{U, R\}$ such that $\{(i_1, j_2) \notin R \wedge (i_2, j_1) \notin R\}$. Thereafter, it switches pairs of links, i.e. removes links (i_1, j_1) and (i_2, j_2) and adds new links (i_1, j_2) and (i_2, j_1) instead. This step is repeated $\frac{k|R|}{2}$ times, which results in $k|R|$ link removals/additions.

In this study, the original graph with sensitive links can be anonymized by the various anonymization strategies and generate different anonymized graphs on a same node set. Among the resulted networks, except the target graph, all the others can be selected as auxiliary graph. Thus, for original social graph de-anonymization, the target graph and the auxiliary graph can be assumed to be the views corresponding to a same group of users [66].

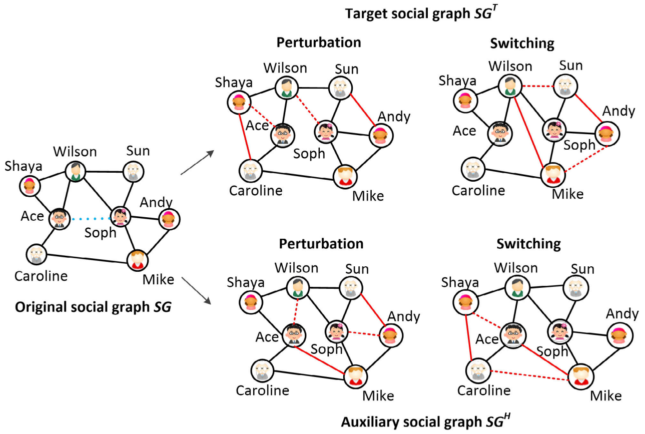

Target social graph SG^T
Perturbation
Switching
Wilson
Sun
Shaya
Ace
Andy
Soph
Caroline
Mike
Original social graph SG
Perturbation
Switching
Auxiliary social graph SG^H

FIGURE 1. Illustration of social network anonymization. In the original social graph, the sensitive link is indicated by the blue dotted line. In the target and auxiliary social graph, the red dotted lines represent the removed links, and the red solid lines mean the newly added links.

Example 1: Given an original social graph with sensitive links, data publishers always remove the sensitive links firstly and then apply anonymization strategies for privacy preservation. Data publishers independently perform privacy protection operations, resulting in multiple views. Specifically, as shown in the middle of Fig. 1, two publishers anonymize the original social graph SG with perturbation strategy, in which two links indicated by the red dotted line are deleted and the two links indicated by the red solid line are added. Similarly, with the switching anonymization strategy, the publishers anonymize SG by switching two links selected randomly, thereby generating graphs shown in the right of Fig. 1. Thus, for original social graph de-anonymization, the above graphs generated by the publishers can be viewed as target social graph SG^T and auxiliary social graph SG^H.

IV. NETWORK STRUCTURE DE-ANONYMIZATION

In this section, we will introduce a principal and explainable network representation model. In order to utilize the complementary information from auxiliary network for network structure de-anonymization, the proposed model is extended to a multi-view scenario.

A. NETWORK REPRESENTATION MODELING

Based on empirical analysis, real-world networks have been proven to have some common topological characteristics, such as small-world, scale-free, and core-periphery features [67]. Hence, networks are always assumed to have specific structural patterns for structure modeling [68]. Moreover, Koutra *et al.* [69] found that network structures can be summarized and compressed by using an enriched set of representative subgraphs as building blocks, such as cliques, stars, chains, and bipartite cores. Inspired by these works, in this study, networks are viewed as the linear summation of a set of elemental subgraphs with a specific interaction pattern i.e., networks can be represented by using the elemental subgraphs as structural bases. Specifically, let $A \in R^{n\times m}$ denote the adjacency matrix of an anonymized graph that consists of m neighborhood structures, i.e., $[A_{:,1}, A_{:,2}, \ldots, A_{:,m}]$. Given a complete basis matrix $D = [D_{:,1}, D_{:,2}, \ldots, D_{:,m}] \in R^{n\times m}$, each neighborhood structure $A_{:,i}$ can be represented as a linear combination of the bases, which is defined as follows:

$$A_{:,i} = [D_{1,:}X_{:,i}, D_{2,:}X_{:,i}, \ldots, D_{n,:}X_{:,i}]^T = \sum_{k=1}^{m} D_{:,k}X_{k,i}, \quad (1)$$

where $X_{k,i}$ denotes the weight corresponding to the structural basis $D_{:,k}$. Consequently, the adjacency matrix A can be represented by $A = DX$, in which the representation matrix $X \in R^{m\times n}$ captures the structural patterns of networks, and the matrix D indicates the set of representative subgraphs. To recognize the structural patterns of the network, the best candidate for the basis matrix D is the adjacency matrix A, and the network can be represented by the following equation:

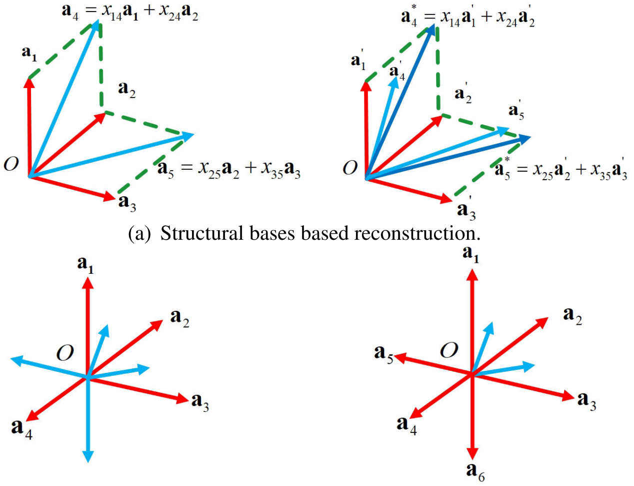
$$A = AX \quad (2)$$

Since social networks often contain frequent subgraphs, the columns of the representation matrix X corresponding to the subgraphs should be correlated. Thus, X is expected to be low-rank. Moreover, individuals may have different interaction patterns in reality. Thus, the modeling of real-world networks should be node-oriented. Because each column of the adjacency matrix A represents the interactions between a node and the rest of the nodes, to characterize the node-specific corruptions in networks and learn the representation matrix, $\ell_{2,1}$ norm, i.e., $||\cdot||_{2,1}$, is adopted in our model to capture the difference between the adjacency matrix A and the graph representation AZ in terms of the graph node. Based on the above observations, social networks can be modeled via the following structured low-rank representation:

$$\min_{X,E} rank(X) + \lambda||E||_{2,1} \quad s.t. A = AX + E. \quad (3)$$

where E is the noise term, $\lambda \geq 0$ is a trade-off parameter used to balance the low-rank and noise terms.

In this study, we assume that the anonymization process does not alter the network structure significantly. Thus, the original network SG can be inferred based on the learned structural patterns from anonymized network. Fig. 2 provides an intuitive illustration of our low-rank representation-based network structure de-anonymization method. To a specified network, by solving the structured low-rank representation model, three structural bases a_1, a_2, and a_3 are identified, and the network can be represented based on them, as shown in the left of Fig. 2 (a). Thus, to an anonymized social graph where some neighborhood structures are perturbed for privacy-preserving, such as a'_4 and a'_5, the original network structure a^*_4 and a^*_5 can be recovered based on the identified structural bases and the learned representation relationships. To the networks with the same dimension, we argue that the lesser the number of its structural bases, the higher the proportion of the redundant structure in the networks and the more possible it is for the anonymized structure to be recovered, as shown in Fig. 2 (b).



(a) Structural bases based reconstruction.

(b) Structural bases of networks with same dimension.

FIGURE 2. The profound meaning of structured low-rank representation based network structure de-anonymization. In (a), the anonymized information can be recovered based on the principle that redundant structures can be represented by structural bases. In (b), the proportion of redundant structures in networks means the upper limit of the possibility of being de-anonymized.

B. REGULARIZED MULTI-VIEW LOW RANK REPRESENTATION

Most of the existing network structure anonymization strategies for privacy-preserving do not take into account the underlying structural characteristics of networks. Consequently, the difference between the original network SG and the target network SG^T, i.e., the anonymized link set, follows different structural patterns with the original network SG. Thus, we argue that the anonymized link set can be identified via the structural patterns centered network representation model.

In the previous section, the network representation model is proposed based on the assumption that only the single view data is available, i.e., the target network, for structural patterns learning. Nevertheless, in reality, multiple related social networks from different viewpoints on the same set of users contain valuable information and can be adopted as auxiliary networks for structural patterns characterization, i.e., optimizing the accuracy of the identified structural bases and the learned representation relationships, thereby improving the performance of network structure de-anonymization.

Let $A^{(i)}, i = 1, 2$ indicate the relationship set of target network SG^T and auxiliary network SG^H respectively, and they can be represented as follows:

$$A^{(1)} = A^{(1)}X^{(1)} + E^{(1)} \quad (4)$$

$$A^{(2)} = A^{(2)}X^{(2)} + E^{(2)} \quad (5)$$

where $X^{(1)}$ and $X^{(2)}$ are representation matrices and $E^{(1)}$ and $E^{(2)}$ are noise terms. Because the target network SG^T and the auxiliary network SG^H capture the interactions between the

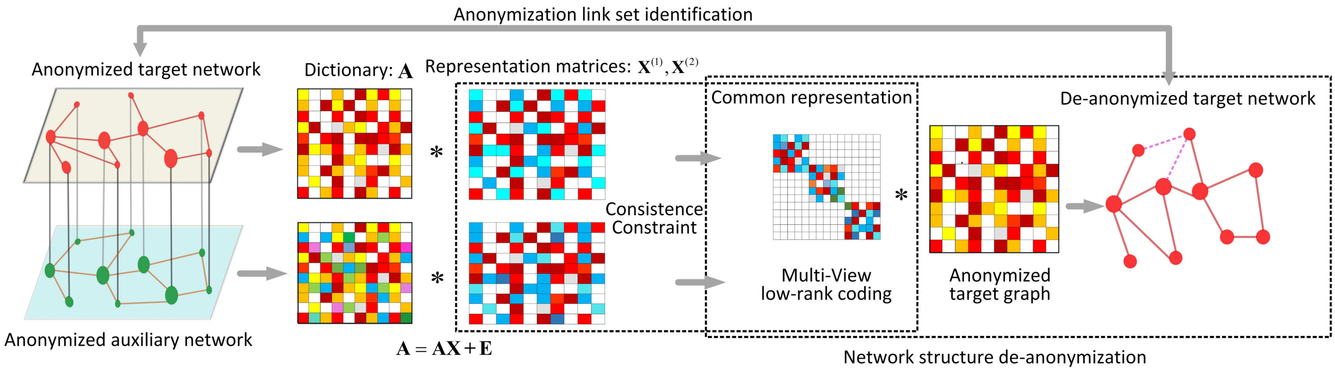

Anonymization link set identification
Anonymized target network
Dictionary: A
Representation matrices: $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$
Common representation
De-anonymized target network
Consistence Constraint
Multi-View low-rank coding
Anonymized target graph
Anonymized auxiliary network
A = AX + E
Network structure de-anonymization

FIGURE 3. Architecture of the proposed multi-view learning model for network structure de-anonymization.

same group of users from different viewpoints, we expect the multiple view-specific networks to embody consistent structural patterns and be complementary to each other, which can be used for the de-anonymization of any one of them. Consequently, we define a regularizer by pushing the representation matrices closer to ensure the consistence, i.e., minimizing the following problem:

$$\Omega(X) = \|X^{(1)} - X^{(2)}\|_{2,1} \tag{6}$$

Based on the regularizer term $\Omega(X)$, the view divergence between the published anonymized networks could be well mitigated. Therefore, based on structured low-rank representation in Equation (3), the regularized multi-view low-rank representation problem can be formulated as follows:

$$\begin{aligned} &\min_{X^{(i)},E^{(i)}} \sum_{i=1}^{2} (rank(X) + \lambda\|E^{(i)}\|_{2,1}) + \alpha\Omega(X) \\ &s.t.\ A^{(i)} = A^{(i)}X^{(i)} + E^{(i)}, \quad i = 1, 2 \end{aligned} \tag{7}$$

where $\|E^{(i)}\|_{2,1}$ represents the $\ell_{2,1}$ of i^{th} view, λ and α represent trade-off parameters. Because the rank($\cdot$) minimization problem in objective function (7) is difficult to solve, nuclear norm $||X||_*$, i.e., the sum of its singular values, was proposed as a good surrogate for the rank minimization problem [70], and we subsequently come up with the following problem formulation:

$$\begin{aligned} &\min_{X^{(i)},E^{(i)}} \sum_{i=1}^{2} (\|X^{(i)}\|_* + \lambda\|E^{(i)}\|_{2,1}) + \alpha\Omega(X) \\ &s.t.\ A^{(i)} = A^{(i)}X^{(i)} + E^{(i)}, \quad i = 1, 2 \end{aligned} \tag{8}$$

To solve the optimization problem as shown in the objective function (8), we adopt the recently proposed inexact augmented Lagrange multiplier (inexact ALM) algorithm in [71]. To facilitate the optimization, the auxiliary variables Q and V are introduced to make the objective function separable:

$$\begin{aligned} &\mathcal{L}(X^{(i)}, Q^{(i)}, E^{(i)}, V) \\ &\quad = \sum_{i=1}^{2} (\|Q^{(i)}\|_* + \lambda\|E^{(i)}\|_{2,1}) \\ &\qquad + \alpha\|V\|_{2,1} + \sum_{i=1}^{2} (\langle Y^{(i)}, A^{(i)} - A^{(i)}X^{(i)} - E^{(i)}\rangle \\ &\qquad + \frac{\mu}{2}\|A^{(i)} - A^{(i)}X^{(i)} - E^{(i)}\|_{\mathcal{F}}^2 + \langle K^{(i)}, X^{(i)} - Q^{(i)}\rangle \\ &\qquad + \frac{\mu}{2}\|X^{(i)} - Q^{(i)}\|_{\mathcal{F}}^2) + \langle W, V - (X^{(1)} - X^{(2)})\rangle \\ &\qquad + \frac{\mu}{2}\|V - (X^{(1)} - X^{(2)})\|_{\mathcal{F}}^2 \\ &s.t.\ A^{(i)} = A^{(i)}X^{(i)} + E^{(i)}, \\ &\quad X^{(i)} = Q^{(i)}, \quad i = 1, 2, \\ &\quad V = X^{(1)} - X^{(2)} \end{aligned} \tag{9}$$

where $Y^{(i)}$, $K^{(i)}$ and W are the Lagrange multipliers, and $\mu > 0$ is a penalty parameter. The initializations for each variable and the complete optimization algorithm for solving the problem (9) are shown in Algorithm 1. Consequently, the optimal value of $X^{(1)}$ can be combined with the target graph SG^T for network structure de-anonymization.

C. MULTI-VIEW LOW-RANK CODING METHOD

The proposed regularizer in Equation (6) models the correlation between the representation matrices $X^{(1)}$ and $X^{(2)}$ via $\ell_{2,1}$ norm to utilize the complementary information. However, the values of $X^{(1)}$ and $X^{(2)}$ learned from Algorithm 1 are the approximate representations of the structural patterns of the anonymized social networks and are not accurate enough. To model the shared structural patterns of multi-view networks directly, we define a common representation matrix $\hat{X}$ and propose a novel method called MVLRC. The architecture of the proposed multi-view learning model for network structure de-anonymization is given in Fig. 3. Here, we present the details of MVLRC.

1) ALGORITHM EXPLANATION

The optimal value of $X^{(1)}$ in Algorithm 1, i.e., the estimation of the structural patterns of the target network, plays an essential role in network structure de-anonymization. To characterize the structural patterns effectively, the regularization term $\|X^{(1)} - X^{(2)}\|_{2,1}$ is introduced to encourage the consistent structural information and restrain the discrepancy between the anonymized and the auxiliary networks. Consequently, the optimal values of $X^{(1)}$ and $X^{(2)}$ are robust to the corruptions coming from anonymization manipulations and prone to the common knowledge of the multi-view networks.

Algorithm 1 Solving Problem (9) by Inexact ALM

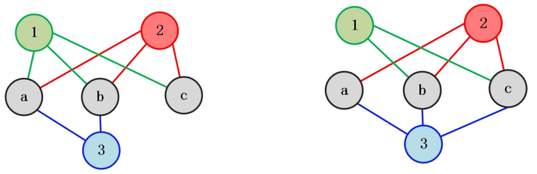
```
Input: The adjacency matrices $A^{(1)}$,$A^{(2)}$ of the target and
    auxiliary networks, trade-off parameter $\lambda$ and $\alpha$.
Output: The representation matrix $X^{(i)}$, error matrix $E^{(i)}$, $i$
    = 1,2.
Initial $X^{(i)} = Q^{(i)} = Y^{(i)} = E^{(i)} = K^{(i)} = 0$, $W = 0$,
$\mu = 10^{-6}$, $\rho = 1.1$, $\varepsilon = 10^{-8}$, $\max_{\mu} = 10^{10}$;
while not converged do
    Fix the other variables and update $Q^{(i)}$ by
    $Q^{(i)} = \arg\min_{Q^{(i)}} \sum_{i=1}^{2} (\frac{1}{\mu}\|Q^{(i)}\|_* + \frac{1}{2}\|Q^{(i)} - (X^{(i)} + \frac{K^{(i)}}{\mu})\|_{\mathcal{F}}^2)$;
    Fix the other variables and update $X^{(1)}$ by
    $X^{(1)} = (2I + A^{(1)^T}A^{(1)})^{-1}(A^{(1)^T}(A^{(1)} - E^{(1)}) + Q^{(1)} + V + X^{(2)} + \frac{A^{(1)^T}Y^{(1)} - K^{(1)} + W}{\mu})$;
    Fix the other variables and update $X^{(2)}$ by
    $X^{(2)} = (2I + A^{(2)^T}A^{(2)})^{-1}(A^{(2)^T}(A^{(2)} - E^{(2)}) + Q^{(2)} - V + X^{(1)} + \frac{A^{(2)^T}Y^{(2)} - K^{(2)} - W}{\mu})$;
    Fix the other variables and update $V$ by
    $V = \arg\min_{V} \frac{\alpha}{\mu}\|V\|_{2,1} + \frac{1}{2}\|V - (X^{(1)} - X^{(2)} + \frac{W}{\mu})\|_{\mathcal{F}}^2$;
    Fix the other variables and update $E^{(i)}$ by
    $E^{(i)} = \arg\min_{E^{(i)}} \sum_{i=1}^{2} (\frac{\lambda}{\mu}\|E^{(i)}\|_{2,1} + \frac{1}{2}\|E^{(i)} - (A^{(i)} - A^{(i)}X^{(i)} + \frac{Y^{(i)}}{\mu})\|_{\mathcal{F}}^2)$
    Update the multipliers
    $Y^{(i)} = Y^{(i)} + \mu(A^{(i)} - A^{(i)}X^{(i)} - E^{(i)})$;
    $K^{(i)} = K^{(i)} + \mu(X^{(i)} - Q^{(i)})$;
    $W = W + \mu(V - (X^{(1)} - X^{(2)}))$;
    Update the parameter $\mu$ by $\mu = \min(\rho\mu, \max_{\mu})$;
    Check the convergence conditions
    $\|A^{(i)} - A^{(i)}X^{(i)} - E^{(i)}\|_\infty < \varepsilon$ and
    $\|X^{(i)} - Q^{(i)}\|_\infty < \varepsilon$ and
    $\|V - (X^{(1)} - X^{(2)})\|_\infty < \varepsilon$
end while
output $X^{(i)}$, $E^{(i)}$, $i = 1, 2$.
```

However, they are still not an accurate reflection of the networks' structural patterns. To solve the problem, we define the representation matrix $\hat{X}$ to characterize the common structural patterns and modify the regularized multi-view low-rank representation as follows.

$$\min_{\hat{X}, E^{(i)}} \sum_{i=1}^{2} (\|\hat{X}\|_* + \lambda\|E^{(i)}\|_{2,1})$$
$$s.t.\ A^{(i)} = A^{(i)}\hat{X} + E^{(i)}, \quad i = 1, 2 \tag{10}$$

Lemma 1: Solving the network representation model defined in the objective function (10) can accurately characterize the common structural patterns of multi-view social networks with the matrix $\hat{X}$.

Proof: For the network representation model $A^{(1)} = A^{(1)}\hat{X} + E^{(1)}$ and $A^{(2)} = A^{(2)}\hat{X} + E^{(2)}$, the low-rank pursuit of $\|\hat{X}\|_*$ and noise minimization $\|E^{(i)}\|_{2,1}$, $i = 1, 2$, collectively require the models to reconstruct the networks with neighborhood structures i.e., structural bases, and noises that are as few as possible. Because of the consistency between $A^{(1)}$ and $A^{(2)}$, the networks have similar structural patterns. Consequently, the structural bases and their contribution to network reconstruction are basically the same. Thus, the networks $A^{(1)}$ and $A^{(2)}$ could be inferred approximately based on a common representation matrix $\hat{X}$, i.e., $A^{(1)}\hat{X}$ and $A^{(2)}\hat{X}$, with the sparse differences being modeled by $E^{(i)}$, $i = 1, 2$. Finally, the common structural patterns of multi-view networks can be captured by the optimal value of the matrix $\hat{X}$ accurately.



$$\begin{bmatrix}1&1&1\\1&1&1\\1&1&0\end{bmatrix}*\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}+\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix};$$
$$\begin{bmatrix}1&1&1\\1&1&1\\1&1&0\end{bmatrix}*\begin{bmatrix}1&1&0\\0&0&0\\0&0&1\end{bmatrix}+\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix};$$
$$\begin{bmatrix}1&1&1\\1&1&1\\1&1&0\end{bmatrix}*\begin{bmatrix}0&0&0\\1&1&1\\0&0&0\end{bmatrix}+\begin{bmatrix}0&0&0\\0&0&0\\0&0&-1\end{bmatrix}.$$
$$\mathbf{A}^{(1)} = \mathbf{A}^{(1)}\mathbf{X}^{(1)} + \mathbf{E}^{(1)}$$

(a)

$$\begin{bmatrix}0&1&1\\1&1&1\\1&1&1\end{bmatrix}*\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}+\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix};$$
$$\begin{bmatrix}0&1&1\\1&1&1\\1&1&1\end{bmatrix}*\begin{bmatrix}1&0&0\\0&1&1\\0&0&0\end{bmatrix}+\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix};$$
$$\begin{bmatrix}0&1&1\\1&1&1\\1&1&1\end{bmatrix}*\begin{bmatrix}0&0&0\\1&1&1\\0&0&0\end{bmatrix}+\begin{bmatrix}-1&0&0\\0&0&0\\0&0&0\end{bmatrix}.$$
$$\mathbf{A}^{(2)} = \mathbf{A}^{(2)}\mathbf{X}^{(2)} + \mathbf{E}^{(2)}$$

(b)

FIGURE 4. Example of the multi-view low-rank coding model.

Example 2: Here, we consider the multi-view networks contained in correlated subgraphs, as shown in Fig. 4 (a) and Fig. 4 (b). To the nonzero regions of their adjacency matrices, we present three different network representations with various constraints. Specifically, the first line in Fig. 4 (a) and Fig. 4 (b) is the network representation in which the representation matrices are full rank and the error matrices are empty. The second line is the network representation in which the rank of representation matrices are reduced to 2, and the error matrices still remain empty. According to the third line, the network representations of the two subgraphs have the same representation matrix with the lowest rank value, i.e., 1, and sparse error matrices. By comparing the three cases, we can conclude that the low-rank and sparse constrains collectively propel the network representation model to represent the multi-view networks with a common representation matrix.

To solve the MVLRC model, we first introduce the auxiliary variable $\hat{Q}$ to make the objective function (10) separable.

The problem can subsequently be transformed as follows:

$$\begin{aligned}
&\min_{\hat{X},\hat{Q},E^{(i)}} \sum_{i=1}^{2} (\|\hat{Q}\|_{*} + \lambda \|E^{(i)}\|_{2,1}) \\
&s.t. \;\; A^{(i)} = A^{(i)}\hat{X} + E^{(i)}, \quad i = 1, 2 \\
&\qquad \hat{X} = \hat{Q}
\end{aligned} \tag{11}$$

Thus, the augmented Lagrangian function of the objective function (11) is defined below.

$$\begin{aligned}
&\mathcal{L}(\hat{X}, \hat{Q}, E^{(i)}) \\
&= \sum_{i=1}^{2} (\|\hat{Q}\|_{*} + \lambda \|E^{(i)}\|_{2,1}) + \langle K, \hat{X} - \hat{Q} \rangle \\
&\quad + \frac{\mu}{2}\|\hat{X} - \hat{Q}\|_{\mathcal{F}}^{2} + \sum_{i=1}^{2} (\langle Y^{(i)}, A^{(i)} - A^{(i)}\hat{X} - E^{(i)} \rangle \\
&\quad + \frac{\mu}{2}\|A^{(i)} - A^{(i)}\hat{X} - E^{(i)}\|_{\mathcal{F}}^{2})
\end{aligned} \tag{12}$$

where $Y^{(i)}$ and K are Lagrangian multipliers and $\mu > 0$ is a penalty parameter.

2) OPTIMIZATION

Here, we adopt the inexact ALM algorithm to solve the optimization problem in (12). We alternatively update the variables $\hat{X}$,$\hat{Q}$ and $E^{(i)}$ while fixing the other variables.

Update $\hat{Q}$: By ignoring the irrelevant variables w.r.t. $\hat{Q}$ in (12), the subproblem is given as follows:

$$\begin{aligned}
\hat{Q} &= \arg\min_{\hat{Q}} \|\hat{Q}\|_{*} + \langle K, \hat{X} - \hat{Q} \rangle + \frac{\mu}{2}\|\hat{X} - \hat{Q}\|_{\mathcal{F}}^{2} \\
&= \arg\min_{\hat{Q}} \frac{1}{\mu}\|\hat{Q}\|_{*} + \frac{1}{2}\|\hat{Q} - (\hat{X} + \frac{K}{\mu})\|_{\mathcal{F}}^{2}
\end{aligned} \tag{13}$$

This problem can be effectively solved by using the singular value thresholding (SVT) operator [34]. Let $\Delta Q = \hat{X} + \frac{K}{\mu}$, the optimal solution of (13) is $\hat{Q} = U_{\hat{Q}} \Omega_{1/\mu}(\sum_{\hat{Q}}) V_{\hat{Q}}$, where $\Omega_{\delta}(H) = \max(H - \delta, 0) + min(W + \delta, 0)$ is the soft-thresholding operator [71].

Update $E^{(i)}$: Here we show how to update $E^{(i)}$ (i = 1,2) with fixed $\hat{Q}$ and $\hat{X}$ variables. After dropping the irrelevant terms w.r.t. $E^{(i)}$ (i = 1,2), the function (12) can be transformed as follows:

$$\begin{aligned}
E^{(i)} &= \arg\min_{E^{(i)}} \lambda \|E^{(i)}\|_{2,1} + \langle Y^{(i)}, A^{(i)} - A^{(i)}\hat{X} - E^{(i)} \rangle \\
&\quad + \frac{\mu}{2}\|A^{(i)} - A^{(i)}\hat{X} - E^{(i)}\|_{\mathcal{F}}^{2} \\
&= \arg\min_{E^{(i)}} \frac{\lambda}{\mu}\|E^{(i)}\|_{2,1} + \frac{1}{2}\|E^{(i)} - (A^{(i)} - A^{(i)}\hat{X} + \frac{Y^{(i)}}{\mu})\|_{\mathcal{F}}^{2}
\end{aligned} \tag{14}$$

The solution to the problem is presented in [34]. Specifically, let $\Psi = A^{(i)} - A^{(i)}\hat{X} + \frac{Y^{(i)}}{\mu}$, the k-th column of $E^{(i)}$ is given as follows:

$$E^{(i)}(:, k) = \begin{cases} \dfrac{\|\Psi_k\| - \frac{\lambda}{\mu}}{\|\Psi_k\|} \Psi_k, & if \; \dfrac{\lambda}{\mu} < \|\Psi_k\|, \\ 0, & otherwise. \end{cases} \tag{15}$$

Update $\hat{X}$: After dropping the terms independent of $\hat{X}$, Equation 12 can be transformed as follows:

$$\begin{aligned}
\hat{X} &= \arg\min_{\hat{X}} \sum_{i=1}^{2} ((\langle Y^{(i)}, A^{(i)} - A^{(i)}\hat{X} - E^{(i)} \rangle) + \langle K, \hat{X} \\
&\quad - \hat{Q} \rangle + \frac{\mu}{2}\|\hat{X} - \hat{Q}\|_{\mathcal{F}}^{2} + \frac{\mu}{2}\|A^{(i)} - A^{(i)}\hat{Q} - E^{(i)}\|_{\mathcal{F}}^{2}) \\
&= (A^{(1)^T} A^{(1)} + A^{(2)^T} A^{(2)} + I)^{-1} [A^{(1)^T} A^{(1)} \\
&\quad + A^{(2)^T} A^{(2)} + \hat{Q} - A^{(1)^T} E^{(1)} - A^{(2)^T} E^{(2)} \\
&\quad + \frac{1}{\mu}(A^{(1)^T} Y^{(1)} + A^{(2)^T} Y^{(2)} - K)]
\end{aligned} \tag{16}$$

The above process is repeated until convergence. The detail of the MVLRC algorithm for finding the common structural patterns is presented in Algorithm 2.

Algorithm 2 Solving Problem (12) by Inexact ALM

```
Input: The adjacency matrices $A^{(1)}$,$A^{(2)}$ of the target and auxiliary networks, trade-off parameter $\lambda$.
Output: The representation matrix $\hat{X}$, error matrix $E^{(i)}$, $i$ = 1,2.
Initial $\hat{X} = \hat{Q} = K = 0$, $Y^{(i)} = E^{(i)} = 0$, $\mu = 10^{-6}$,$\rho = 1.1$, $\varepsilon = 10^{-8}$, $\max_{\mu} = 10^{10}$;
while not converged do
    Fix the other variables and update $\hat{Q}$ via (13);
    Fix the other variables and update $\hat{X}$ via (16);
    Fix the other variables and update $E^{(i)}$ via (14);
    Update the multipliers $K$ and $Y^{(i)}$
    $Y^{(i)} = Y^{(i)} + \mu(A^{(i)} - A^{(i)}X^{(i)} - E^{(i)})$;
    $K = K + \mu(\hat{X} - \hat{Q})$;
    Update the parameter $\mu$ by $\mu = \min(\rho\mu, \max_{\mu})$;
    Check the convergence conditions
    $\|A^{(i)} - A^{(i)}X^{(i)} - E^{(i)}\|_{\infty} < \varepsilon$ and
    $\|\hat{X} - \hat{Q}\|_{\infty} < \varepsilon$
end while
output $\hat{X}$, $E^{(i)}$
```

D. DE-ANONYMIZATION ALGORITHM

Given an target graph SG^T, the goal of our study is to generate a de-anonymized graph SG^D based on the topology of SG^T and the learned optimal structural patterns, thereby inferring the anonymized links that are perturbed for privacy-preserving.

According to the network representation model, the target network can be inferred based on the linear combination of the basis matrix with the representation weight. To maintain the data utility for subsequent data analysis, the anonymized operations for privacy-preserving are always limited, and the anonymized network largely retains the intrinsic structural features of the original network. Thus, the adjacency matrix $A^{(1)}$ of the target network is the strongest candidate to be the basis matrix. In addition, the representation matrix $X^{(*)}$, learned from the anonymized networks by Algorithm 1 or Algorithm 2, captures the structural patterns and can be used for target network reconstruction. Accordingly, we can infer the target network structure by $A^{(1)}X^{(*)}$.

The whole procedures for network structure de-anonymization are shown in Algorithm 3. Specifically, in the first step, the structural patterns, captured by the representation matrix of the target network can be efficiently learned from the multi-view social networks by Algorithm 1 (i.e., $X^{(1)}$) or Algorithm 2 (i.e., $\hat{X}$). Then, the adjacency matrix O of the target network can be calculated in Step 2.

```
Algorithm 3 Network Structure De-Anonymization
Input: The adjacency matrices $A^{(1)}$, $A^{(2)}$ of target and auxiliary networks.
Output: De-anonymized target network.
1: Learn the structural pattern $X'$ of the available anonymized networks by Algorithm 1 or Algorithm 2;
2: Reconstruct the target network by $O = A^{(1)}X'$;
3: Return De-anonymized target network $O$.
```

E. COMPUTATIONAL COMPLEXITY

In this section, we provide a detailed complexity analysis of the proposed algorithms wherein, the main computation complexity mainly focuses on nuclear norm computation, matrix inversion, and multiplication computation. Specifically, exact SVD of an $n \times m$ matrix has time complexity $\mathcal{O}(min\{nm^2, n^2m\})$. In case of a matrix with size $m \times m$, time complexity of SVD is $\mathcal{O}(m^3)$. It will be time consuming if m is large, i.e., the number of data samples is large. Fortunately, the SVD of an $m \times m$ matrix can be accelerated to $\mathcal{O}(r^2m)$ according to [72], where r is the rank of the low-rank matrix. In addition, the computation complexity of matrix inversion and multiplication computation all cost $\mathcal{O}(m^3)$.

Suppose the $A^{(1)} \in R^{n\times m}$ and $A^{(2)} \in R^{n\times m}$ matrices, the main time-consuming components of Algorithm 1 concentrates on the solving of $Q^{(1)}$ and $Q^{(2)}$ in Step 3 and the updating of $X^{(1)}$ and $X^{(2)}$. Since $Q^{(1)}$, $Q^{(2)} \in R^{n\times m}$, the time complexity of SVD on $Q^{(1)}$ and $Q^{(2)}$ is $\mathcal{O}(l_1m^3)$ and $\mathcal{O}(l_2m^3)$ respectively, where l_1 and l_2 are the total number of SVD. Meanwhile, in Algorithm 1, the complexity of matrix inverse and multiplication in $X^{(1)}$ and $X^{(2)}$ costs $\mathcal{O}(l_3m^3)$ and $\mathcal{O}(l_4m^3)$ respectively, where l_3 and l_4 are the total number of matrix inverse and multiplication operations. Therefore, the total computation complexity of Algorithm 1 is approximately $\mathcal{O}(tl_1m^3 + tl_2m^3 + tl_3m^3 + tl_4m^3)$, assuming that there are t iterations. Moreover, the main time-consuming processes of Algorithm 2 are SVD computation used in solving $\hat{Q}$, and matrix inverse and multiplication in solving $\hat{X}$; the total complexity of Algorithm 2 is $\mathcal{O}(tf_1m^3 + tf_2m^3)$, where f_1 represents the total number of SVD computations, f_2 is the total number of matrix inverse and multiplication operations, and t is the iteration number. In Algorithm 3, the main complexity comes from the de-anonymization of the target network in Step 2, and the matrix multiplication costs $\mathcal{O}(nm^2)$ for $A^{(1)} \in R^{n\times m}$ and $X' \in R^{n\times m}$. Thus, the total complexity of de-anonymization method combining Algorithm 1 and Algorithm 2 is $\mathcal{O}((l_1 + l_2 + l_3 + l_4 + f_1 + f_2)tm^3)$. Because $l_1, l_2, l_3, l_4, f_1, f_2, t$ are all small constants, the complexity is $\mathcal{O}(m^3)$. Similarly, the total complexity of de-anonymization method combining Algorithm 1 and Algorithm 3 is $\mathcal{O}(m^3)$ $+\mathcal{O}(nm^2)$.

It is worthwhile to note that the complexity of the proposed method is a one-time cost and may be performed off-line. Therefore, it is feasible for graphs with a couple of thousands of nodes. For very large graphs with millions of nodes, randomized algorithms may be used to figure out the SVD. Furthermore, when compared with the existing graph data de-anonymization algorithms, the complexity of the proposed methods are competitive. For example, the method proposed by Narayanan *et al.* [73] is $\mathcal{O}(n^4)$, the method in [74] costs $\mathcal{O}(n^3)$, and the time complexity of the methods in [75] is $\mathcal{O}(n^3)$.

V. EXPERIMENTS

In this section, the performance of the proposed MVLRC algorithm is evaluated on two synthetic networks, and three real-world datasets which are obtained from Stanford Network Analysis.[1] We adopt Reliability and AUC as performance metrics, and verify the superiority of MVLRC algorithm over the comparison methods under various anonymization techniques. In addition, we evaluate the robustness of MVLRC algorithm with diverse parameters setting, and visually compare the true and identified anonymized links to demonstrate the effectiveness of the MVLRC algorithm.

A. EXPERIMENTAL SETTING

Our approach is compared to three state-of-the-art single-view and multi-view structure de-anonymization methods. The details of the methods are given as follows:

- **RPCA based Recovery Approach**. RPCA [76], [77] is applied for subspace segmentation and link prediction. Thus, we identify the true network structure and infer the anonymized link set by conducting RPCA on the anonymized network (referred to as "RPCA").
- **LRR based Recovery Approach**. LRR [34] is a representative method to recover the original row space from a set of corrupted observations. Here, we utilize LRR to model the anonymized network where the anonymized links are viewed as noise, outliers, and sample-specific corruptions (referred as "LRR").
- **MVLRR**. The method recovers the target network by combining Algorithm 1 and Algorithm 3, where the anonymized auxiliary network is incorporated by regularization.
- **MVLRC**. The method recovers the target network by combining Algorithm 2 and Algorithm 3, where the common structural patterns are characterized by a specific representation matrix.

[1] https://snap.stanford.edu/data/index.html

TABLE 3. Reliability of different algorithms on small-world network under various anonymization coefficient *k*.

Privacy	Densification				Sparsification				Perturbation				Switching			
	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC
k=0.10	0.172	0.240	0.363	**0.716**	0.092	0.123	0.153	**0.293**	0.160	0.218	0.264	**0.575**	0.200	0.306	0.437	**0.678**
k=0.15	0.193	0.285	0.439	**0.755**	0.079	0.137	0.173	**0.319**	0.170	0.240	0.279	**0.597**	0.235	0.378	0.469	**0.685**
k=0.20	0.226	0.306	0.479	**0.773**	0.071	0.144	0.181	**0.349**	0.200	0.243	0.356	**0.624**	0.283	0.442	0.528	**0.702**
k=0.25	0.255	0.311	0.512	**0.776**	0.052	0.150	0.204	**0.344**	0.210	0.252	0.398	**0.641**	0.310	0.471	0.540	**0.697**
k=0.30	0.272	0.379	0.556	**0.782**	0.050	0.166	0.212	**0.321**	0.230	0.284	0.405	**0.646**	0.349	0.498	0.555	**0.691**

TABLE 4. AUC of network de-anonymization methods on small-world network under various anonymization coefficient *k*.

Algorithm	k=0.1		k=0.15		k=0.20		k=0.25		k=0.30	
	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification
RPCA	0.503	0.500	0.524	0.568	0.536	0.537	0.558	0.562	0.605	0.543
LRR	0.675	0.785	0.710	0.840	0.780	0.835	0.785	0.775	0.700	0.825
MVLRR	0.855	0.965	0.895	0.925	0.910	0.930	0.891	0.920	0.820	0.880
MVLRC	**0.968**	**0.990**	**0.963**	**0.978**	**0.955**	**0.950**	**0.943**	**0.955**	**0.953**	**0.948**
	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching
RPCA	0.528	0.565	0.518	0.504	0.538	0.509	0.528	0.510	0.548	0.505
LRR	0.753	0.790	0.725	0.753	0.748	0.698	0.730	0.710	0.688	0.663
MVLRR	0.870	0.873	0.840	0.818	0.853	0.778	0.828	0.745	0.788	0.695
MVLRC	**0.953**	**0.904**	**0.919**	**0.869**	**0.903**	**0.833**	**0.865**	**0.793**	**0.826**	**0.736**

TABLE 5. Reliability of different algorithms on LFR network under various anonymization coefficient *k*.

Privacy	Densification				Sparsification				Perturbation				Switching			
	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC
k=0.10	0.219	0.344	0.391	**0.509**	0.072	0.222	0.297	**0.401**	0.179	0.354	0.389	**0.525**	0.193	0.365	0.406	**0.558**
k=0.15	0.247	0.472	0.508	**0.578**	0.045	0.227	0.297	**0.440**	0.234	0.450	0.498	**0.660**	0.228	0.411	0.463	**0.551**
k=0.20	0.275	0.538	0.577	**0.635**	0.041	0.248	0.303	**0.456**	0.217	0.491	0.536	**0.663**	0.245	0.446	0.501	**0.573**
k=0.25	0.268	0.590	0.623	**0.671**	0.036	0.231	0.287	**0.412**	0.246	0.530	0.586	**0.701**	0.294	0.466	0.515	**0.594**
k=0.30	0.406	0.616	0.656	**0.695**	0.041	0.274	0.328	**0.434**	0.299	0.559	0.605	**0.707**	0.314	0.482	0.527	**0.609**

TABLE 6. AUC of network de-anonymization methods on LFR network under various anonymization coefficient *k*.

Algorithm	k=0.1		k=0.15		k=0.20		k=0.25		k=0.30	
	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification
RPCA	0.725	0.605	0.680	0.535	0.700	0.500	0.640	0.540	0.715	0.545
LRR	0.875	0.805	0.890	0.795	0.865	0.830	0.855	0.755	0.850	0.748
MVLRR	**0.890**	0.965	0.890	0.965	0.885	0.945	0.868	0.895	0.885	0.890
MVLRC	**0.890**	**0.970**	**0.910**	**0.975**	**0.895**	**0.955**	**0.925**	**0.910**	**0.905**	**0.915**
	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching
RPCA	0.643	0.555	0.642	0.535	0.615	0.570	0.625	0.515	0.573	0.505
LRR	0.833	0.725	0.823	0.700	0.758	0.658	0.785	0.633	0.763	0.598
MVLRR	0.873	0.793	0.868	**0.780**	0.850	0.733	**0.868**	0.655	0.805	**0.678**
MVLRC	**0.898**	**0.800**	**0.908**	0.770	**0.900**	**0.743**	0.858	**0.703**	**0.835**	**0.678**

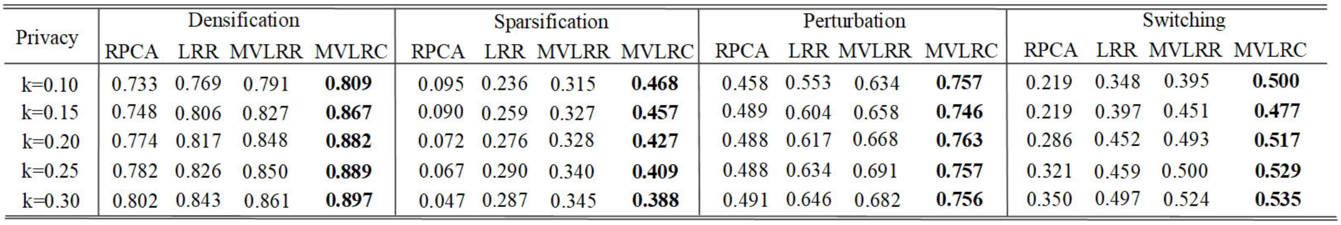
To measure the accuracy of the proposed MVLRC method for network structure de-anonymization and anonymized links identification, we adopt reliability as the evaluation metric, defined as follows:

$$Reliability = \frac{AN + DN}{TAN + TDN} \tag{17}$$

where AN is the number of added links being accurately found, DN is the number of deleted links being accurately found, and TAN and TDN are the total number of added links and deleted links for network structure anonymization, respectively. The more the anonymized links being identified by the de-anonymization algorithms, the higher the value of the reliability metric. Moreover, the metric AUC (Area Under the Receiver operating characteristic curve) is also adopted for the performance evaluation of network structure de-anonymization i.e., a higher value of AUC means a better network structure de-anonymization performance.

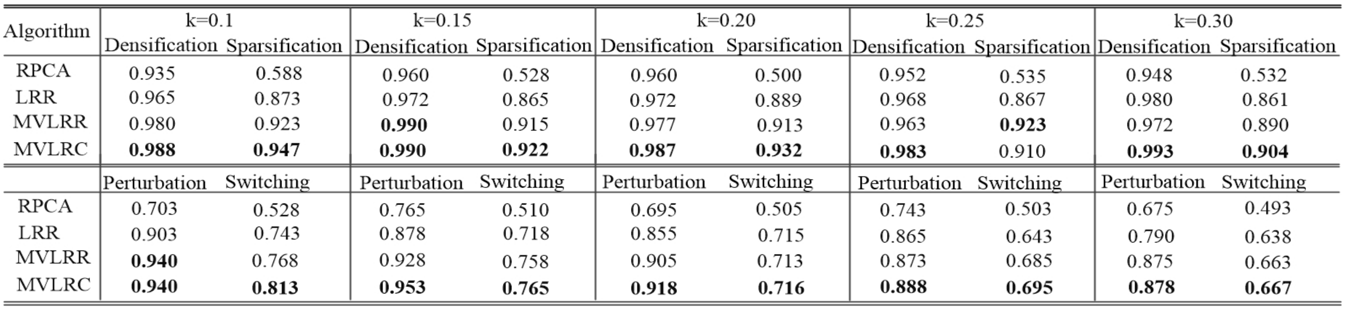
Algorithm 4 details the overall evaluation process of MVLRC. Firstly, based on the target network and auxiliary network, we infer target networks using the de-anonymization methods. Then, we compare the inferred

TABLE 7. **Reliability of different algorithms on Email-EuAll database under various anonymization coefficient *k*.**



Privacy	Densification				Sparsification				Perturbation				Switching			
	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC
k=0.10	0.733	0.769	0.791	**0.809**	0.095	0.236	0.315	**0.468**	0.458	0.553	0.634	**0.757**	0.219	0.348	0.395	**0.500**
k=0.15	0.748	0.806	0.827	**0.867**	0.090	0.259	0.327	**0.457**	0.489	0.604	0.658	**0.746**	0.219	0.397	0.451	**0.477**
k=0.20	0.774	0.817	0.848	**0.882**	0.072	0.276	0.328	**0.427**	0.488	0.617	0.668	**0.763**	0.286	0.452	0.493	**0.517**
k=0.25	0.782	0.826	0.850	**0.889**	0.067	0.290	0.340	**0.409**	0.488	0.634	0.691	**0.757**	0.321	0.459	0.500	**0.529**
k=0.30	0.802	0.843	0.861	**0.897**	0.047	0.287	0.345	**0.388**	0.491	0.646	0.682	**0.756**	0.350	0.497	0.524	**0.535**

TABLE 8. **AUC of network de-anonymization methods on Email-EuAll database under various anonymization coefficient *k*.**



Algorithm	k=0.1		k=0.15		k=0.20		k=0.25		k=0.30	
	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification
RPCA	0.935	0.588	0.960	0.528	0.960	0.500	0.952	0.535	0.948	0.532
LRR	0.965	0.873	0.972	0.865	0.972	0.889	0.968	0.867	0.980	0.861
MVLRR	0.980	0.923	**0.990**	0.915	0.977	0.913	0.963	**0.923**	0.972	0.890
MVLRC	**0.988**	**0.947**	**0.990**	**0.922**	**0.987**	**0.932**	**0.983**	0.910	**0.993**	**0.904**
	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching
RPCA	0.703	0.528	0.765	0.510	0.695	0.505	0.743	0.503	0.675	0.493
LRR	0.903	0.743	0.878	0.718	0.855	0.715	0.865	0.643	0.790	0.638
MVLRR	**0.940**	0.768	0.928	0.758	0.905	0.713	0.873	0.685	0.875	0.663
MVLRC	**0.940**	**0.813**	**0.953**	**0.765**	**0.918**	**0.716**	**0.888**	**0.695**	**0.878**	**0.667**

target network with the original target network and obtain the difference between them. Next, we sort the possible links and compare them with the real anonymized link set to calculate the evaluation metric.

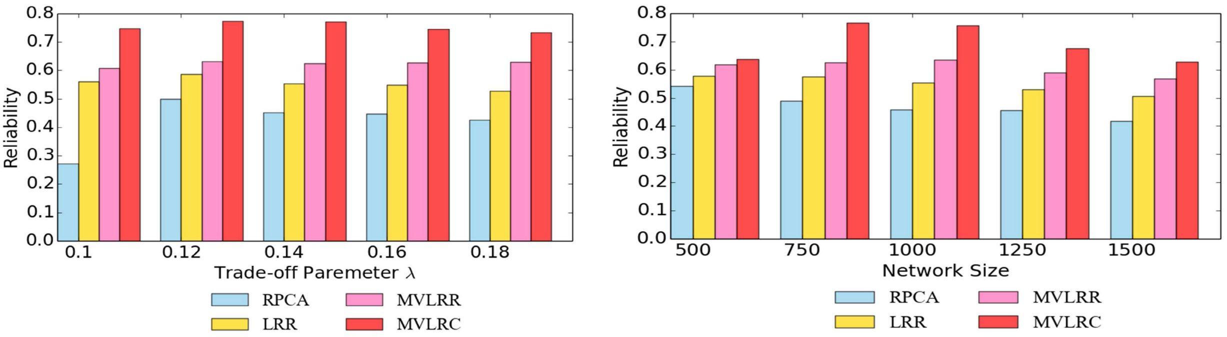
```
Algorithm 4 Experimental Evaluation Process
Input: The adjacency matrix $A$ of the target network, the adjacency matrices $A^{(1)}$, $A^{(2)}$ of the target network and auxiliary network, and the real anonymized link set $I$ generated by various anonymization techniques in the target network.
Output: The values of evaluation metric Reliability and AUC.
1: Estimate target network $O$ based on $A^{(1)}$ and $A^{(2)}$ using all candidate structure de-anonymization methods;
2: Calculate the discrepancy link set by $S = A - O$;
3: Rank the entries of $S$ based on their absolute values $|s_{i,j}|$, and select the top $|I|$ entries as the inferred anonymized link set $P$ in which the negative items correspond to the deleted links, and the positive items correspond to the added links in the anonymization process;
4: Calculate the evaluation metric Reliability and AUC by comparing $P$ and $I$;
5: Return the values of Reliability and AUC.
```

B. RESULTS ON SYNTHETIC NETWORKS

In this section, we conduct experiments on synthetic networks to confirm the expectation that MVLRC would perform well in this case. Based on Newman-Watts-Strogatz model [78], we generate a small-world network in which the node number is set to 1000, average node degree is set to 8 and the probability of random reconnection is set to 0.3. Meanwhile, we produce LFR community network with Lancichinetti-Fortunato-Radicchi (LFR) model [79] where the node number is set to 1000, average node degree is 5, and the mixing ratio is 0.2. Table 3 and Table 4 show the reliability and AUC results of RPCA, LRR, MVLRR, and MVLRC under different anonymization techniques on the small-world network. It can be clearly seen that MVLRC outperforms better than the other methods for anonymized links inference, and the reason is that the structural patterns of small-world network can be better captured by MVLRC. Moreover, a similar conclusion can be drawn from the experimental results on LFR network, as shown in Table 5 and Table 6. Therefore, the proposed MVLRC method is effective for the de-anonymization of synthetic networks.

C. RESULTS ON EMAIL-EuAll DATABASE

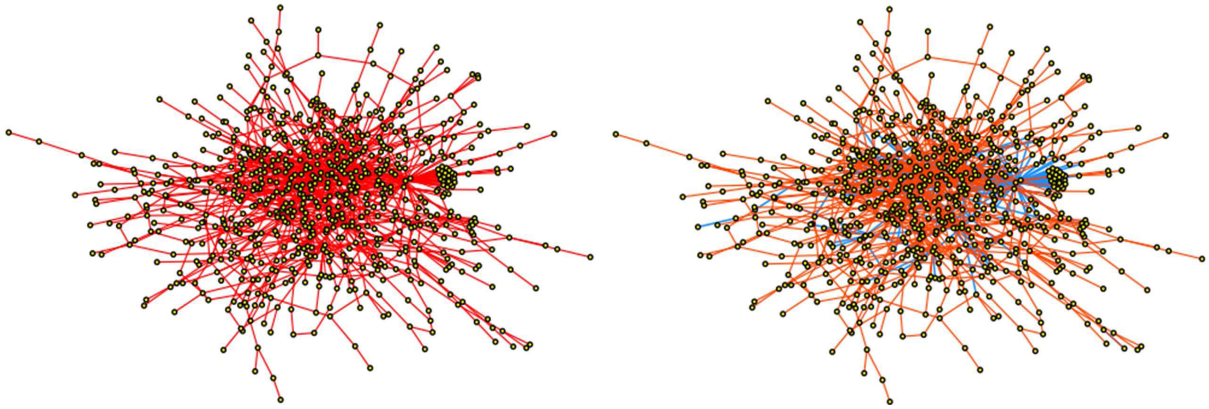
The Email-EuAll Database [80] is obtained from a large, undisclosed European research institution, and contains 3,038,531 emails between 287,755 different email addresses. Nodes represent individual persons who sent or received email messages, and links denote emails having been sent or received from one person to the others. We view the database as a simple, undirected graph. Because the large scale of real-world networks always make the experiments based directly on the them to be time-consuming and sometimes impractical, researchers related to social network analysis often sample the real-world networks firstly and then conduct experiments on the sampled networks. Similarly, in our study, all algorithms are tested on the networks with the size of 1000 randomly sampled from the database. The sampled networks are anonymized with the techniques



(a) The reliability under different λ values.

(b) The reliability under different network sizes.

FIGURE 5. Reliability of network de-anonymization methods on Email-EuAll database under various settings. The anonymization strategy of Email-EuAll adopts *Perturbation* mechanism, anonymization coefficient $k = 0.1$. In (a), the network size equals to 1000. In (b), the tradeoff parameter λ equals to 0.13.

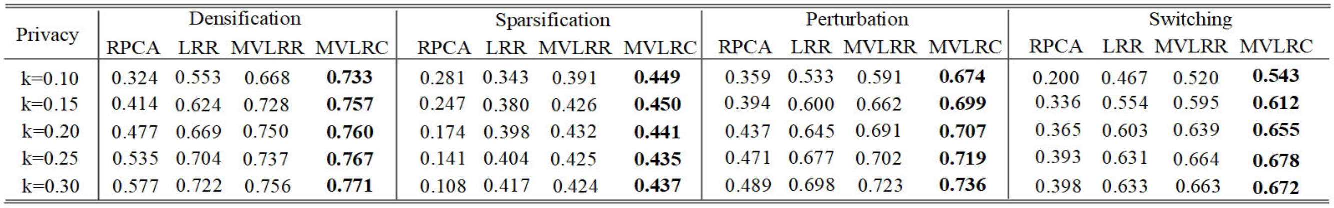


(a) The true anonymized link set.

(b) The detected anonymized link set.

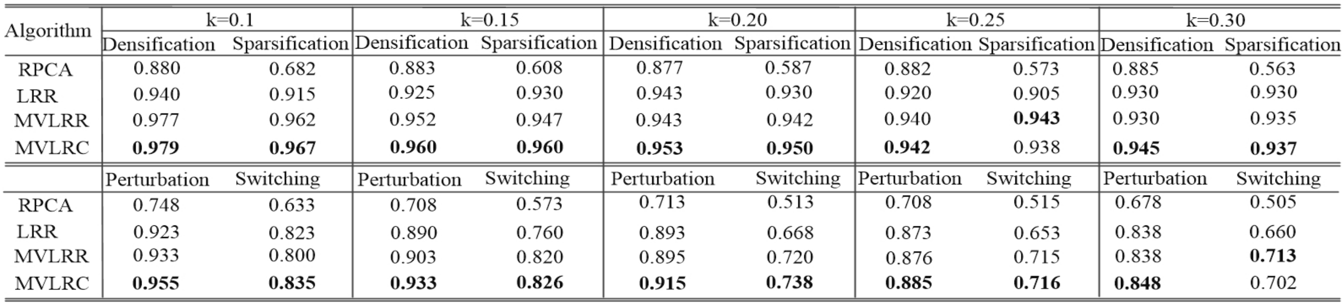
FIGURE 6. The visualization of true and detected anonymized link set in Email-EuAll database. The anonymization strategy of Email-EuAll adopts *Perturbation* mechanism, anonymization coefficient k equals to 0.2, trade off parameter λ equals to 0.13, and network size equals to 1,000. In (b), the orange lines represent the detected anonymized links, while the blue lines represent the undetected links.

TABLE 9. Reliability of different algorithms on the Facebook database under various anonymization coefficients k.

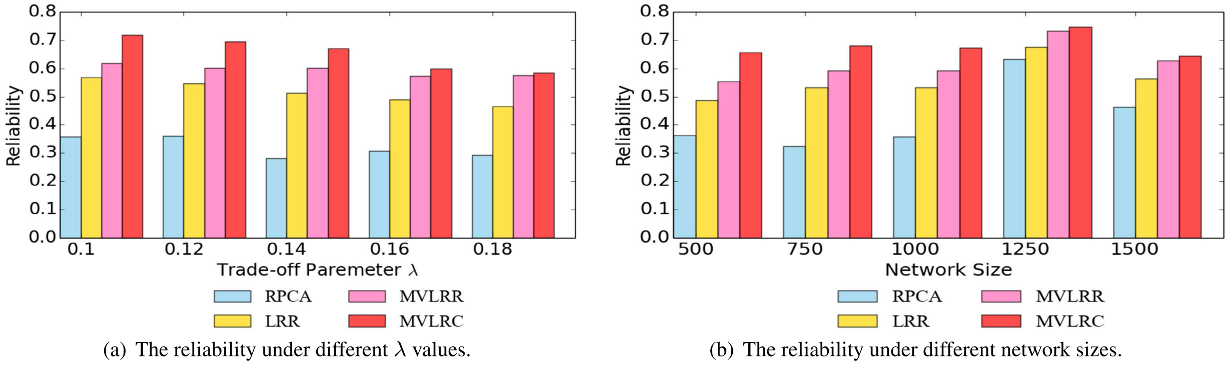


Privacy	Densification				Sparsification				Perturbation				Switching			
	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC
k=0.10	0.324	0.553	0.668	**0.733**	0.281	0.343	0.391	**0.449**	0.359	0.533	0.591	**0.674**	0.200	0.467	0.520	**0.543**
k=0.15	0.414	0.624	0.728	**0.757**	0.247	0.380	0.426	**0.450**	0.394	0.600	0.662	**0.699**	0.336	0.554	0.595	**0.612**
k=0.20	0.477	0.669	0.750	**0.760**	0.174	0.398	0.432	**0.441**	0.437	0.645	0.691	**0.707**	0.365	0.603	0.639	**0.655**
k=0.25	0.535	0.704	0.737	**0.767**	0.141	0.404	0.425	**0.435**	0.471	0.677	0.702	**0.719**	0.393	0.631	0.664	**0.678**
k=0.30	0.577	0.722	0.756	**0.771**	0.108	0.417	0.424	**0.437**	0.489	0.698	0.723	**0.736**	0.398	0.633	0.663	**0.672**

TABLE 10. AUC of network de-anonymization methods on the Facebook database under various anonymization coefficients k.



Algorithm	k=0.1		k=0.15		k=0.20		k=0.25		k=0.30	
	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification
RPCA	0.880	0.682	0.883	0.608	0.877	0.587	0.882	0.573	0.885	0.563
LRR	0.940	0.915	0.925	0.930	0.943	0.930	0.920	0.905	0.930	0.930
MVLRR	0.977	0.962	0.952	0.947	0.943	0.942	0.940	**0.943**	0.930	0.935
MVLRC	**0.979**	**0.967**	**0.960**	**0.960**	**0.953**	**0.950**	**0.942**	0.938	**0.945**	**0.937**
	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching
RPCA	0.748	0.633	0.708	0.573	0.713	0.513	0.708	0.515	0.678	0.505
LRR	0.923	0.823	0.890	0.760	0.893	0.668	0.873	0.653	0.838	0.660
MVLRR	0.933	0.800	0.903	0.820	0.895	0.720	0.876	0.715	0.838	**0.713**
MVLRC	**0.955**	**0.835**	**0.933**	**0.826**	**0.915**	**0.738**	**0.885**	**0.716**	**0.848**	0.702



(a) The reliability under different λ values.

(b) The reliability under different network sizes.

FIGURE 7. Reliability of network de-anonymization methods on Facebook database under various settings. The anonymized strategy of Facebook adopts *Perturbation* mechanism, anonymization coefficient *k* equals to 0.1. In (a), the network size equals to 1,000. In (b), the tradeoff parameter λ equals to 0.13.

presented in Section 2.2 to generate the target and auxiliary networks. For each sampled network, we repeat our experiments 10 times and report the average result.

Table 7 shows the results of RPCA, LRR, MVLRR, and MVLRC under different anonymization techniques. We can observe that multi-view approaches MVLRR and MVLRC outperform the single view methods RPCA and LRR in terms of reliability. The results demonstrate that the auxiliary network is valuable for de-anonymization optimization, and the proposed multi-view framework is effective for complementary information modeling. In addition, Table 7 shows that LRR outperforms RPCA when only the target network is available. This can be explained by the reason that LRR has a stronger expressive capability than that of RPCA. For the multi-view modeling based network structure de-anonymization, Table 7 illustrates that MVLRC outperforms MVLRR. Here, the advantages of MVLRC are mainly due to its methodology. Specifically, MVLRC directly targets on learning the common structural patterns by a specific representation matrix, which determines the de-anonymization results. In contrast, MVLRR is proposed for learning the target network's structural patterns and the auxiliary network's structural patterns respectively, constrained by a regularization term. Moreover, according to Table 7, the values of the reliability metric under densification anonymization strategy obviously increase with anonymization coefficient k increasing, which indicates that the densification strategy tends to be inefficient for network structure anonymization. Meanwhile, Table 8 shows the de-anonymization results in terms of AUC metric. The results show that MVLRC generally obtains the best performance among the de-anonymization methods under various anonymization techniques and coefficients, and the multi-view de-anonymization algorithms perform better than the single-view ones.

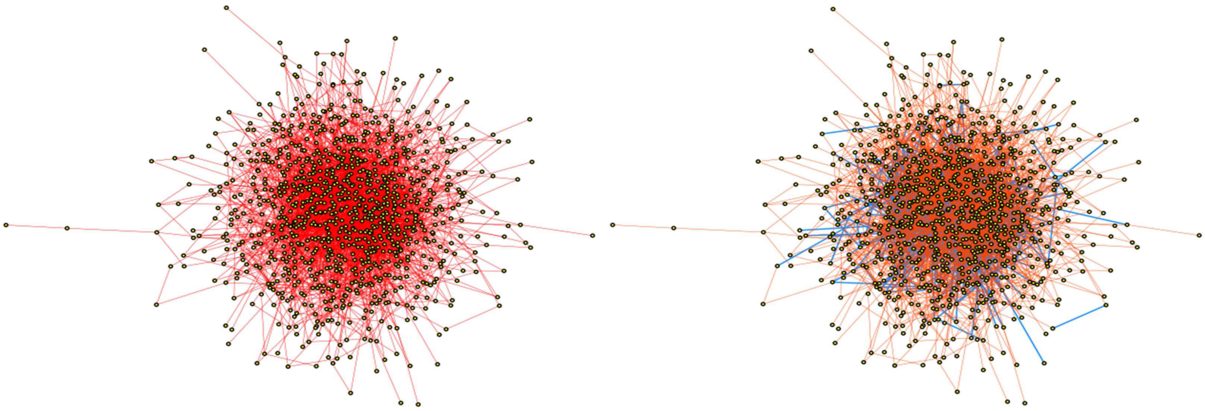
Besides the superiorities in terms of de-anonymization accuracy, another advantage of MVLRC is that it works well under a wide range of parameter specifications, as shown in Fig. 5. It can be seen that the MVLRC algorithm is better than the other methods as the parameter λ varies from 0.10 to 0.18, as shown in Fig. 5 (a). Moreover, notice that MVLRC is not sensitive to the parameter λ on this dataset. With the size of the sampled network growing from 500 to 1,500, MVLRC is better than the other methods in all cases, as shown in Fig. 5 (b). It is worthy to note that there have been similar results when the other anonymization techniques are adopted.

To test the effectiveness of MVLRC for network structure de-anonymization, we visually compare the true anonymized link set and the identified anonymized links, as shown in Fig. 6. In detail, the added and deleted links in the anonymized network are presented in Fig. 6 (a). By comparison, in Fig. 6 (b), the orange lines indicate the identified anonymized links while the blue lines represent the undetected ones. It is worth noting that most of the anonymized links are identified correctly.

D. RESULTS ON FACEBOOK DATABASE

The Facebook Database [81] contains information of nearly 10 million pairs of users on Facebook. The website aims to promote and facilitate the interactions across friends, colleagues, etc. For example, if user A and user B are friends, or they have same political tendency and hobbies, the network would create a link between them with a high probability. Here we sample the dataset randomly into a network with 1,000 nodes and conduct RPCA, LRR, MVLRR, and MVLRC on it. Table 9 shows the de-anonymization results under different anonymization strategies and various anonymization coefficients. It can be seen that our proposed MVLRC algorithm performs better than the others in terms of reliability metric. The experimental results agree with the discussions addressed in Section 4, which shows that the auxiliary network does contain valuable information for de-anonymization, and the MVLRC method can perform well in capturing common structural information. Table 10 shows the performance of de-anonymization methods in term of AUC. The results demonstrate that MVLRC outperforms the other methods.

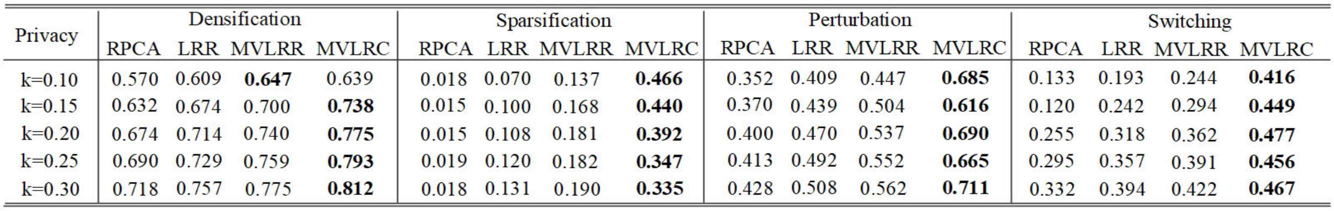
To examine the robustness of the proposed method, we perform experiments with various trade-off parameters and different network sizes. We present the reliability values corresponding to the de-anonymization methods in Fig. 7. The



(a) The true anonymized link set. (b) The detected anonymized link set.

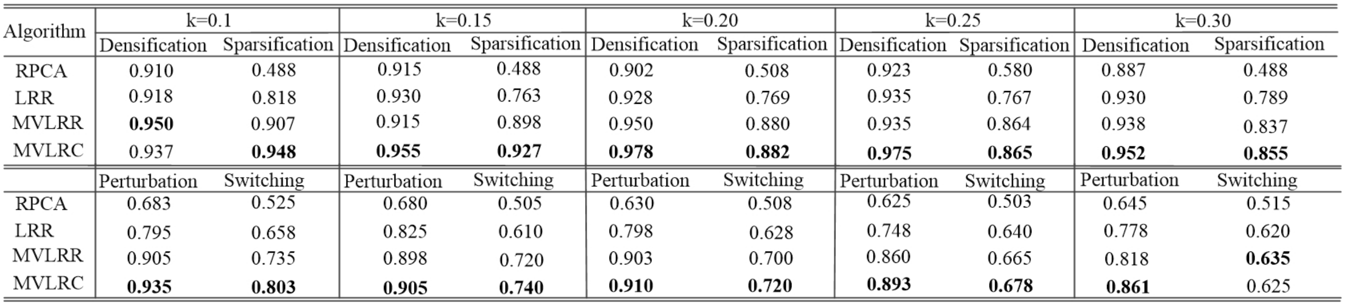
FIGURE 8. The visualization of the true and detected anonymized link set in Facebook network. The anonymized strategy of Facebook adopts *Perturbation* mechanism, anonymization coefficient k equals to 0.2, trade off parameter λ equals to 0.13 and network size equals to 1,000. In (b), the orange lines represent the detected anonymized links, and the blue lines represent the undetected links.

TABLE 11. Reliability of different algorithms on the Bitcoin database under various anonymization coefficient k.



Privacy	Densification				Sparsification				Perturbation				Switching			
	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC	RPCA	LRR	MVLRR	MVLRC
k=0.10	0.570	0.609	**0.647**	0.639	0.018	0.070	0.137	**0.466**	0.352	0.409	0.447	**0.685**	0.133	0.193	0.244	**0.416**
k=0.15	0.632	0.674	0.700	**0.738**	0.015	0.100	0.168	**0.440**	0.370	0.439	0.504	**0.616**	0.120	0.242	0.294	**0.449**
k=0.20	0.674	0.714	0.740	**0.775**	0.015	0.108	0.181	**0.392**	0.400	0.470	0.537	**0.690**	0.255	0.318	0.362	**0.477**
k=0.25	0.690	0.729	0.759	**0.793**	0.019	0.120	0.182	**0.347**	0.413	0.492	0.552	**0.665**	0.295	0.357	0.391	**0.456**
k=0.30	0.718	0.757	0.775	**0.812**	0.018	0.131	0.190	**0.335**	0.428	0.508	0.562	**0.711**	0.332	0.394	0.422	**0.467**

TABLE 12. AUC of network de-anonymization methods on the Bitcoin database under various anonymization coefficient k.



Algorithm	k=0.1		k=0.15		k=0.20		k=0.25		k=0.30	
	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification	Densification	Sparsification
RPCA	0.910	0.488	0.915	0.488	0.902	0.508	0.923	0.580	0.887	0.488
LRR	0.918	0.818	0.930	0.763	0.928	0.769	0.935	0.767	0.930	0.789
MVLRR	**0.950**	0.907	0.915	0.898	0.950	0.880	0.935	0.864	0.938	0.837
MVLRC	0.937	**0.948**	**0.955**	**0.927**	**0.978**	**0.882**	**0.975**	**0.865**	**0.952**	**0.855**
	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching	Perturbation	Switching
RPCA	0.683	0.525	0.680	0.505	0.630	0.508	0.625	0.503	0.645	0.515
LRR	0.795	0.658	0.825	0.610	0.798	0.628	0.748	0.640	0.778	0.620
MVLRR	0.905	0.735	0.898	0.720	0.903	0.700	0.860	0.665	0.818	**0.635**
MVLRC	**0.935**	**0.803**	**0.905**	**0.740**	**0.910**	**0.720**	**0.893**	**0.678**	**0.861**	0.625

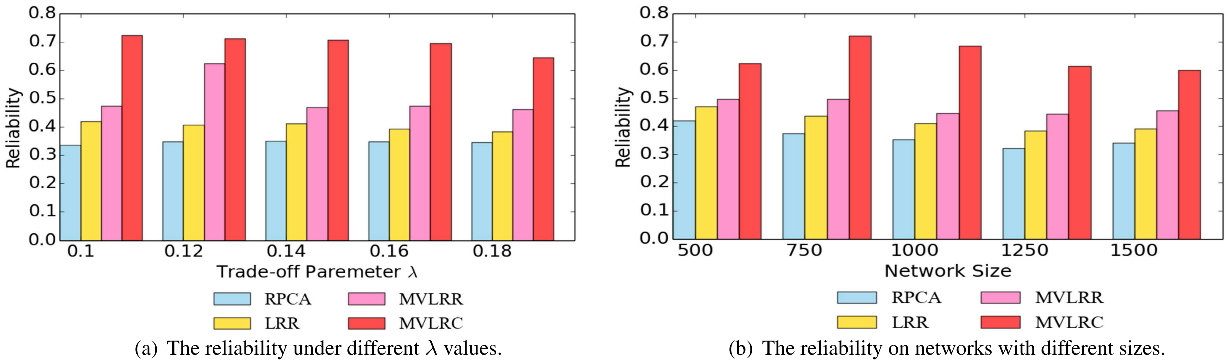
results verify the superiority of the MVLRC method under various conditions, which confirms the results obtained from the previous subsection.

After reconstructing the target network, we infer the anonymized links by comparing the recovered target network with the original target network. Then, we estimate the consistency between the inferred anonymized links and true anonymized links to measure the de-anonymization accuracy. The detailed results presented in Fig. 8 show that the MVLRC has excellent performance for network structure de-anonymization.

E. RESULTS ON BITCOIN-ALPHA DATABASE

To further verify the de-anonymization performance of MVLRC, we adopt the Bitcoin-Alpha dataset [82] for evaluation. The data is collected from Bitcoin Alpha, which is an online trust platform for users who make a deal by using Bitcoin. Since anonymity makes transactions risky, many researchers use Bitcoin-Alpha and Bitcoin-OTC to verify the effectiveness of de-anonymization algorithms. We transform the Bitcoin-Alpha dataset into an undirected graph with a weight value of 1, and then use the sampled subgraphs as experimental networks.

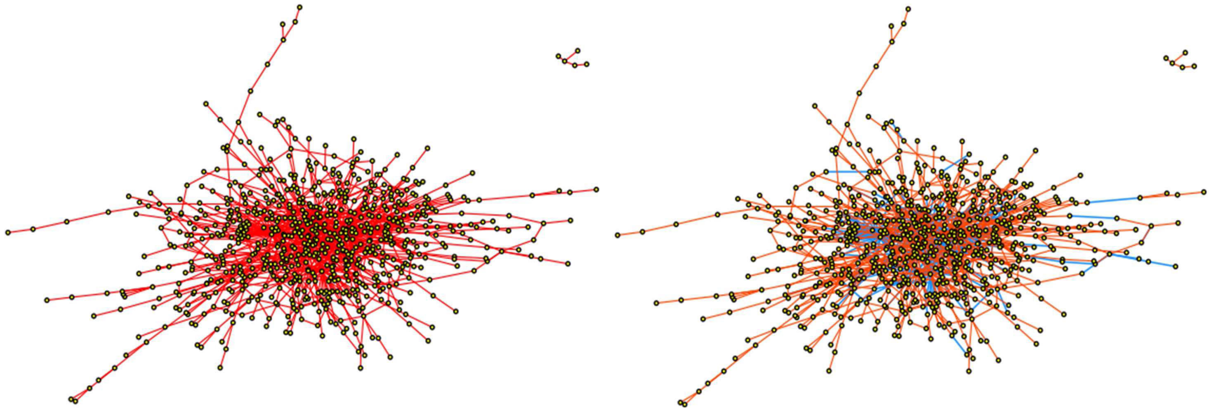
Considering the experimental results on EuAll-Email network and Facebook network, our algorithm performs better than other methods on the Bitcoin-Alpha network. Specifically, Table 11 and Table 12 show the evaluation results with different anonymization strategies and various anonymization coefficient on the size of 1000 nodes in terms of reliability and AUC. When compared with the single view methods, multi-view learning algorithms show significant performance improvements. Furthermore, for the multi-view



(a) The reliability under different λ values.

(b) The reliability on networks with different sizes.

FIGURE 9. **Reliability of network de-anonymization methods on the Bitcoin database under various settings. The anonymized strategy of Bitcoin adopts *Perturbation* mechanism, anonymization coefficient *k* equals to 0.1. In (a), the network size equals to 1,000. In (b), the tradeoff parameter λ equals to 0.13.**



(a) The true anonymized link set.

(b) The detected anonymized link set.

FIGURE 10. **The visualization of true and detected anonymized link set in Bitcoin database. The anonymized strategy of Bitcoin adopts *Perturbation* mechanism, anonymization coefficient *k* equals to 0.2, trade off parameter λ equals to 0.13, and network size equals to 1,000. In (b), the orange lines represent the detected anonymized links, and the blue lines represent the undetected links.**

learning algorithms, MVLRC results in a better performance than MVLRR for network structure de-anonymization. Moreover, the experimental results in Fig. 9 under different parameter settings prove the robustness of our MVLRC approach. In addition, the results in Fig. 10 illustrate the effectiveness of our MVLRC method for anonymized network recovery.

F. TIME CONSUMPTION

We explore the average time consumption of the proposed approaches, RPCA and LRR over 10 runs on two synthetic networks and three real-world databases, as shown in Fig. 11. We can observe from the results that multi-view low rank learning methods MVLRR and MVLRC generally cost more time than single-view low rank learning methods RPCA and LRR. However, compared with single-view low rank learning methods, the multi-view low rank learning methods with higher computational cost have better performance for link inference. Moreover, about the multi-view low rank learning methods, we can see that the running time of MVLRC is much lower than that of MVLRR, which demonstrates the efficiency of the proposed multi-view network structural learning procedure.

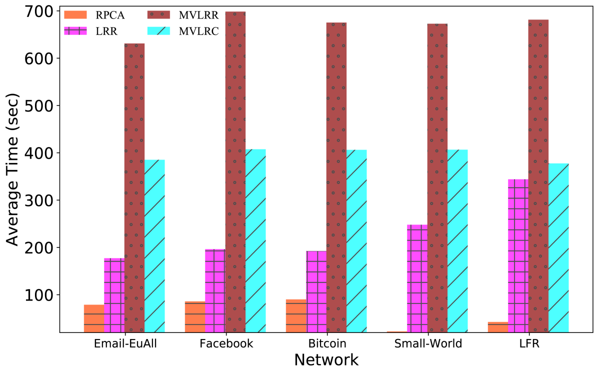


FIGURE 11. **Average running time of network de-anonymization methods on synthetic networks and real-world databases.**

VI. CONCLUSION AND DISCUSSION

Data publication has become a vital foundation for big data analysis and applications; however, inappropriate sharing and usage of data could threaten users' privacy. To the commonly existing multi-view network data, in this study, we propose the MVLRC algorithm for network structure de-anonymization. The method models the target network and auxiliary network together to learn the common structural patterns, thereby identifying the anonymized link set of target network. Our empirical results on real-world networks show highly promising improvements in accuracy of anonymized links inference compared with the methods that only utilize single-view data. Therefore, besides the target network, the auxiliary networks collected from different viewpoints could also be explored to strengthen privacy inference attacks, which challenges the traditional privacy protection methods.

Three problems for future research are worthy to be considered: 1) investigating the performance of network de-anonymization algorithms in the face of more sophisticated privacy preserving techniques, 2) exploring the influence of auxiliary network on de-anonymization accuracy theoretically, and 3) developing effective anonymization methods for multi-view network data.

REFERENCES

[1] F. Bonchi, C. Castillo, A. Gionis, and A. Jaimes, "Social network analysis and mining for business applications," *Acm Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–22, 2011.

[2] T. Wu, L. Chen, X. Xian, and Y. Guo, "Evolution prediction of multi-scale information diffusion dynamics," *Knowl.-Based Syst.*, vol. 113, pp. 186–198, Dec. 2016.

[3] T. Wu, Y. Guo, L. Chen, and Y. Liu, "Integrated structure investigation in complex networks by label propagation," *Phys. A, Stat. Mech. Appl.*, vol. 448, pp. 68–80, Apr. 2016.

[4] S. Qiao, N. Han, Y. Gao, R.-H. Li, J. Huang, J. Guo, L. A. Gutierrez, and X. Wu, "A fast parallel community discovery model on complex networks through approximate optimization," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 9, pp. 1638–1651, Sep. 2018.

[5] S. Qiao, N. Han, J. Zhou, R.-H. Li, C. Jin, and L. A. Gutierrez, "SocialMix: A familiarity-based and preference-aware location suggestion approach," *Eng. Appl. Artif. Intell.*, vol. 68, pp. 192–204, Feb. 2018.

[6] L. Wu, S. Shah, S. Choi, M. Tiwari, and C. Posse, "The browsemaps: Collaborative filtering at linkedin," in *Proc. 6th Workshop Recommender Syst. Social Web*, 2014, pp. 1–8.

[7] W. Tan, K. Xu, and D. Wang, "An anti-tracking source-location privacy protection protocol in WSNs based on path extension," *IEEE Internet Things J.*, vol. 1, no. 5, pp. 461–471, Oct. 2014.

[8] S.-F. Tzeng, S.-J. Horng, T. Li, X. Wang, P.-H. Huang, and M. Khurram Khan, "Enhancing security and privacy for identity-based batch verification scheme in VANETs," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3235–3248, Apr. 2017.

[9] P. Wang, W. He, and J. Zhao, "A tale of three social networks: User activity comparisons across Facebook, Twitter, and foursquare," *IEEE Internet Comput.*, vol. 18, no. 2, pp. 10–15, Mar. 2014.

[10] X. Yang, R. Lu, K.-K.-R. Choo, F. Yin, and X. Tang, "Achieving efficient and privacy-preserving cross-domain big data deduplication in cloud," *IEEE Trans. Big Data*, early access, Jun. 29, 2017, doi: 10.1109/TBDATA.2017.2721444.

[11] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey of recent developments," *ACM Comput. Surv.*, vol. 42, no. 4, pp. 1–53, Jun. 2010.

[12] J. H. Abawajy, M. I. H. Ninggal, and T. Herawan, "Privacy preserving social network data publication," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1974–1997, 3rd Quart., 2016.

[13] J. Casas-Roma, J. Herrera-Joancomartí, and V. Torra, "A survey of graph-modification techniques for privacy-preserving on networks," *Artif. Intell. Rev.*, vol. 47, no. 3, pp. 1–26, 2016.

[14] A. M. Fard, K. Wang, and P. S. Yu, "Limiting link disclosure in social network analysis through subgraph-wise perturbation," in *Proc. 15th Int. Conf. Extending Database Technol. (EDBT)*, 2012, pp. 109–119.

[15] A. Milani Fard and K. Wang, "Neighborhood randomization for link privacy in social network analysis," *World Wide Web*, vol. 18, no. 1, pp. 9–32, Jan. 2015.

[16] J. Cheng, W. C. Fu, and J. Liu, "K-isomorphism: Privacy preserving network publication against structural attacks," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, Indianapolis, IN, USA, Jun. 2010, pp. 459–470.

[17] L. Zhang and W. Zhang, "Edge anonymity in social network graphs," in *Proc. Int. Conf. Comput. Sci. Eng.*, 2009, pp. 1–8.

[18] E. Zheleva and L. Getoor, "Preserving the privacy of sensitive relationships in graph data," in *Proc. Int. Workshop Privacy, Secur., Trust (KDD)*, 2007, pp. 153–171.

[19] T. Tassa and D. J. Cohen, "Anonymization of centralized and distributed social networks by sequential clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 2, pp. 311–324, Feb. 2013.

[20] X. Ying, K. Pan, X. Wu, and L. Guo, "Comparisons of randomization and K-degree anonymization schemes for privacy preserving social network publishing," in *Proc. 3rd Workshop Social Netw. Mining Anal. (SNA-KDD)*, 2009, pp. 1–10.

[21] G. Beigi, K. Shu, Y. Zhang, and H. Liu, "Securing social media user data: An adversarial approach," in *Proc. 29th Hypertext Social Media*, Jul. 2018, pp. 165–173.

[22] S. Ji, W. Li, M. Srivatsa, and R. Beyah, "Structural data de-anonymization: Quantification, practice, and implications," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2014, pp. 1040–1053.

[23] X. Li, J. Smith, T. Pan, T. N. Dinh, and M. T. Thai, "Quantifying privacy vulnerability to socialbot attacks: An adaptive non-submodular model," *IEEE Trans. Emerg. Topics Comput.*, early access, May 24, 2018, doi: 10.1109/TETC.2018.2840433.

[24] S. Ji, P. Mittal, and R. Beyah, "Graph data anonymization, de-anonymization attacks, and de-anonymizability quantification: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1305–1326, 2nd Quart., 2017.

[25] K. J. L. Backstrom and C. Dwork, "Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography," in *Proc. 16th Int. Conf. World Wide Web*, 2007, pp. 181–190.

[26] M. Backes, M. Humbert, J. Pang, and Y. Zhang, "walk2friends: Inferring social links from mobility profiles," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2017, pp. 1943–1957.

[27] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, 2007.

[28] Y. Zhang, M. Humbert, B. Surma, P. Manoharan, J. Vreeken, and M. Backes, "Towards plausible graph anonymization," 2017, *arXiv:1711.05441*. [Online]. Available: http://arxiv.org/abs/1711.05441

[29] M. Fire, G. Katz, L. Rokach, and Y. Elovici, "Links reconstruction attack," in *Security and Privacy in Social Networks*, Y. Altshuler, Eds. New York, NY, USA: Springer, 2013, pp. 181–196.

[30] L. Wu, X. Ying, and X. Wu, "Reconstruction from randomized graph via low rank approximation," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2010, pp. 60–71.

[31] N. Vuokko and E. Terzi, "Reconstructing randomized social networks," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2010, pp. 49–59.

[32] J. Li, Y. Wu, J. Zhao, and K. Lu, "Low-rank discriminant embedding for multiview learning," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3516–3529, Nov. 2017.

[33] X. Xian, T. Wu, S. Qiao, X. Wang, W. Wang, and Y. Liu, "NetSRE: Link predictability measuring and regulating," *Knowl.-Based Syst.*, vol. 196, pp. 1–16, May 2020.

[34] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 663–670.

[35] S. Yu, "Big privacy: Challenges and opportunities of privacy study in the age of big data," *IEEE Access*, vol. 4, pp. 2751–2763, 2016.

[36] S. Ji, W. Li, P. Mittal, X. Hu, and R. Beyah, "SecGraph: A uniform and open-source evaluation system for graph data anonymization and de-anonymization," in *Proc. 24th USENIX Secur. Symp. (USENIX Secur.)*, 2015, pp. 303–318.

[37] S. Nilizadeh, A. Kapadia, and Y.-Y. Ahn, "Community-enhanced de-anonymization of online social networks," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2014, pp. 537–548.

[38] H. Fu, A. Zhang, and X. Xie, "Effective social graph deanonymization based on graph structure and descriptive information," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 4, pp. 1–29, Aug. 2015.

[39] N. Matatov, L. Rokach, and O. Maimon, "Privacy-preserving data mining: A feature set partitioning approach," *Inf. Sci.*, vol. 180, no. 14, pp. 2696–2720, Jul. 2010.

[40] B. Zhou and J. Pei, "The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks," *Knowl. Inf. Syst.*, vol. 28, no. 1, pp. 47–77, Jul. 2011.

[41] J. Yang, B. Wang, X. Yang, H. Zhang, and G. Xiang, "A secure K-automorphism privacy preserving approach with high data utility in social networks," *Secur. Commun. Netw.*, vol. 7, no. 9, pp. 1399–1411, Sep. 2014.

[42] T. Zhu, G. Li, W. Zhou, and P. S. Yu, "Differentially private data publishing and analysis: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 8, pp. 1619–1638, Aug. 2017.

[43] X. Meng, H. Li, and J. Cui, "Different strategies for differentially private histogram publication," *J. Commun. Inf. Netw.*, vol. 2, no. 3, pp. 68–77, Sep. 2017.

[44] S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava, "Class-based graph anonymization for social network data," *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 766–777, Aug. 2009.

[45] L. Zou and L. Chen, "K-automorphism: A general framework for privacy preserving network publication," *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 946–957, 2009.

[46] X. Ying and X. Wu, "Randomizing social networks: A spectrum preserving approach," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2008, pp. 739–750.

[47] P. Mittal, C. Papamanthou, and D. Song, "Preserving link privacy in social network based systems," *CoRR*, vol. abs/1208.6189, pp. 1–19, Apr. 2013.

[48] L. Lian, W. Jie, J. Liu, and J. Zhang, "Privacy preserving in social networks against sensitive edge disclosure," Dept. Comput. Sci., Univ. Kentucky, Lexington, KY, USA, Tech. Rep. CMIDA-HiPSCCS 006-08, 2008.

[49] Y. Wang and X. Wu, "Preserving differential privacy in degree-correlation based graph generation," *Trans. Data Privacy*, vol. 6, no. 2, p. 127, 2013.

[50] D. Proserpio, S. Goldberg, and F. McSherry, "Calibrating data to sensitivity in private data analysis: A platform for differentially-private analysis of weighted datasets," *Proc. VLDB Endowment*, vol. 7, no. 8, pp. 637–648, Apr. 2014.

[51] Q. Xiao, R. Chen, and K.-L. Tan, "Differentially private network data release via structural inference," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. New York, NY, USA: ACM, 2014, pp. 911–920.

[52] M. Korayem and D. J. Crandall, "De-anonymizing users across heterogeneous social computing platforms," in *Proc. AAAI Conf. Weblogs Social Media*, 2013, pp. 689–692.

[53] K. Shu, S. Wang, J. Tang, R. Zafarani, and H. Liu, "User identity linkage across online social networks: A review," *ACM SIGKDD Explor. Newslett.*, vol. 18, no. 2, pp. 5–17, Mar. 2017.

[54] W.-H. Lee, C. Liu, S. Ji, P. Mittal, and R. B. Lee, "Blind de-anonymization attacks using social networks," in *Proc. Workshop Privacy Electron. Soc. (WPES)*, 2017, pp. 1–4.

[55] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in *Proc. IEEE Symp. Secur. Privacy*, May 2009, pp. 173–187.

[56] S. Ji, W. Li, M. Srivatsa, J. S. He, and R. Beyah, "General graph data de-anonymization: From mobility traces to social networks," *ACM Trans. Inf. Syst. Secur.*, vol. 18, no. 4, pp. 1–29, 2016.

[57] S. Ji, W. Li, N. Z. Gong, P. Mittal, and R. Beyah, "Seed-based de-anonymizability quantification of social networks," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 7, pp. 1398–1411, Jul. 2016.

[58] X. Zhou, X. Liang, X. Du, and J. Zhao, "Structure based user identification across social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 6, pp. 1178–1191, Jun. 2018.

[59] X. Ying and X. Wu, "On link privacy in randomizing social networks," *Knowl. Inf. Syst.*, vol. 28, no. 3, pp. 645–663, Sep. 2011.

[60] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Inf. Fusion*, vol. 38, pp. 43–54, Nov. 2017.

[61] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1413–1421.

[62] M. White, X. Zhang, D. Schuurmans, and Y. Yu, "Convex multi-view subspace learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1673–1681.

[63] D. Dou and S. Coulondre, "Detecting privacy violations in multiple views publishing," in *Proc. Int. Conf. Database Expert Syst. Appl.* Berlin, Germany: Springer, 2012, pp. 506–513.

[64] C. Yao, X. S. Wang, and S. Jajodia, "Checking for k-anonymity violation by views," in *Proc. 31st Int. Conf. Very Large Data Bases*. Los Angeles, CA, USA: VLDB Endowment, 2005, pp. 910–921.

[65] M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava, "Anonymizing social networks," Univ. Massachusetts, Boston, MA, USA, Tech. Rep. 07-19, 2007, pp. 1–18.

[66] Z. Zhang, Q. Gu, T. Yue, and S. Su, "Identifying the same person across two similar social networks in a unified way: Globally and locally," *Inf. Sci.*, vols. 394–395, pp. 53–67, Jul. 2017.

[67] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.

[68] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, and H. E. Stanley, "Toward link predictability of complex networks," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 8, pp. 2325–2330, Feb. 2015.

[69] D. Koutra, U. Kang, J. Vreeken, and C. Faloutsos, "Summarizing and understanding large graphs," *Stat. Anal. Data Mining, ASA Data Sci. J.*, vol. 8, no. 3, pp. 183–202, Jun. 2015.

[70] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, Jan. 2010.

[71] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*. [Online]. Available: http://arxiv.org/abs/1009.5055

[72] Z. Ding, M. Shao, and Y. Fu, "Missing modality transfer learning via latent low-rank constraint," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4322–4334, Nov. 2015.

[73] A. Narayanan, E. Shi, and B. I. P. Rubinstein, "Link prediction by de-anonymization: How we won the kaggle social network challenge," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2011, pp. 1825–1834.

[74] P. Pedarsani, D. R. Figueiredo, and M. Grossglauser, "A Bayesian method for matching two similar graphs without seeds," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2013, pp. 1598–1607.

[75] S. Ji, W. Li, M. Srivatsa, J. S. He, and R. Beyah, "Structure based data de-anonymization of social networks and mobility traces," in *Proc. Int. Conf. Inf. Secur.*, 2014, pp. 237–254.

[76] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.

[77] R. Pech, D. Hao, L. Pan, H. Cheng, and T. Zhou, "Link prediction via matrix completion," *EPL (Europhys. Lett.)*, vol. 117, no. 3, 2016, Art. no. 38002.

[78] M. E. J. Newman and D. J. Watts, "Renormalization group analysis of the small-world network model," *Phys. Lett. A*, vol. 263, nos. 4–6, pp. 341–346, Dec. 1999.

[79] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 78, no. 4, Oct. 2008, Art. no. 046110.

[80] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *Acm Trans. Knowl. Discovery Data*, vol. 1, no. 1, pp. 1–41, 2007.

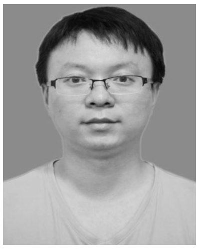
[81] J. Leskovec and J. J. Mcauley, "Learning to discover social circles in ego networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 539–547.

[82] S. Kumar, F. Spezzano, V. S. Subrahmanian, and C. Faloutsos, "Edge weight prediction in weighted signed networks," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 221–230.

XINGPING XIAN received the M.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, where she is currently pursuing the Ph.D. degree with the Department of Computer Science and Technology. She has published more than ten scientific articles in international journals and conferences. Her current research interests include machine learning, social network analysis, and privacy-preserving.

TAO WU (Member, IEEE) received the Ph.D. degree from the University of Electronic Science and Technology of China, Chengdu, China. He is currently an Assistant Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. He has published more than 20 scientific articles in international journals and conferences. His research interests include complex networks, machine learning, edge computing, and privacy-preserving. He is currently a Reviewer of *Knowledge-Based Systems*, IEEE ACCESS, *Physica A*, *Computers & Electrical Engineering*, and *International Journal of Modern Physics B*.

SHAOJIE QIAO (Member, IEEE) received the B.S. and Ph.D. degrees from Sichuan University, Chengdu, China, in 2004 and 2009, respectively. From 2007 to 2008, he worked as a Visiting Scholar with the School of Computing, National University of Singapore. He is currently a Professor with the School of Software Engineering, Chengdu University of Information Technology, Chengdu. He has led several research projects in the areas of databases and data mining. He has authored more than 40 high quality articles and coauthored more than 90 articles. His research interests include complex networks and trajectory data mining.



WEI WANG received the Ph.D. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2017. He is currently an Associate Professor with Sichuan University, Chengdu, China. His recent studies have focused on investigating the spreading mechanisms of information, epidemic, rumor, and associated critical phenomena in complex networks. He has published more than 40 articles in the field of network science and spreading dynamics.

YANBING LIU received the Ph.D. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2007. He is currently a Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include network analysis and network security.

NAN HAN received the M.S. and Ph.D. degrees from the Chengdu University of Traditional Chinese Medicine, Chengdu, China. She is currently an Associate Professor with the School of Management, Chengdu University of Information Technology. Her research interests include trajectory prediction and data mining.

• • •