Downloaded from orbit.dtu.dk on: May 10, 2024

DTU Library

Sparse Trust Data Mining

Nie, Pengli; Xu, Guangquan; Jiao, Litao; Liu, Shaoying; Liu, Jian; Meng, Weizhi; Wu, Hongyue; Feng, Meiqi; Wang, Weizhe; Jing, Zhengjun

Total number of authors:
11

Published in:
IEEE Transactions on Information Forensics and Security

Link to article, DOI:
10.1109/TIFS.2021.3109412

Publication date:
2021

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Nie, P., Xu, G., Jiao, L., Liu, S., Liu, J., Meng, W., Wu, H., Feng, M., Wang, W., Jing, Z., & Zheng, X. (2021). Sparse Trust Data Mining. *IEEE Transactions on Information Forensics and Security*, *16*(99), 4559-4573. https://doi.org/10.1109/TIFS.2021.3109412

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

Sparse Trust Data Mining

Pengli Nie, Guangquan Xu*, *Member, IEEE*, Litao Jiao, Shaoying Liu, *Fellow, IEEE*, Jian Liu*, Weizhi Meng, Hongyue Wu*, *Member, IEEE*, Meiqi Feng, Weizhe Wang, Zhengjun Jing, and Xi Zheng, *Member, IEEE*

Abstract—As recommendation systems continue to evolve, researchers are using trust data to improve the accuracy of recommendation prediction and help users find relevant information. However, large recommendation systems with trust data suffer from the sparse trust problem , which leads to grade inflation and severely affects the reliability of trust propagation. This paper presents a novel research on sparse trust data mining, which includes the new concept of sparse trust, a sparse trust model, and a trust mining framework. It lays a foundation for the trust-related research in large recommended systems. The new trust mining framework is based on customized normalization functions and a novel transitive gossip trust model, which discovers potential trust information between entities in a large-scale user network and applies it to a recommendation system. We conducts a comprehensive performance evaluation on both real-world and synthetic datasets. The results confirm that our framework mines new trust and effectively ameliorates sparse trust problem.

Index Terms—Anti-sparsification, Recommendation System, Sparse Trust, Trust Model

1 Introduction

With the accelerated development of the Internet, an increasing number of online business activities generate a large amount of data every day. Recommendation systems try to provide users with items they might be interested in. Typically, recommendation systems automatically predict users' interests by collecting, integrating, and analyzing entity rating data in a network. This technology is also called collaborative filtering. Interesting enough, relevant data on Taobao[1] shows that there may be only a few buyers for each product with few effective comments, which causes the problem of sparse evaluation. Although trust data has been widely used in recommendation systems to solve the problems of cold-start and sparse evaluation (little "users-items" evaluation information) inherent [1], [2], large-scale recommendation systems still have sparse trust problem (i.e., insufficient trust data available for inference) closely related to these two problems. The cold-start issue of a recommendation system refers to how to provide satisfactory recommendation service to newly registered users and how to recommend newly stored items to users who like it. Similarly, for newly developed systems, there are few initial users and few user behaviors, how to make the recommendation system works well is also a big issue.

The sparse trust problem means that there is not enough trust established between users in the system so that too little trust data is available for recommendation. Evaluating sparse trust is aimed to obtain the sparsity of the user's original trust. The sparsity of a user's trust reflects the fact how little information the user has about interacting with other users. Some existing trust models [3], [4], such as the model of personalized trust [5], swift trust model [6], and Semiring Trust Inference for Trust-Aware Social Recommenders [7], evaluate a business's performance on the basis of overall ratings and neglect some of the underlying information in textual reviews [8]. The Textual Reputation Model [9] calculates a merchant's overall reputation score based on user reviews to improve the traditional model [10]. However, these recommendations are not valid for new or inactive users. This is due to the very few correlation ratings between users and products [11].

Sparse trust problem is a challenge not only for effective large-scale recommendation systems, but also for calculation about huge matrices. As mentioned earlier, the scarcity of trust evaluation and the cold-start have largely contributed to the sparse trust problem, which further increases the difficulty of huge matrices' calculation. At the same time, solving these problems is challenging [12].

We propose a *sparse trust mining framework* based on the notion of trust transfer to achieve anti-sparsity (i.e., to alleviate the sparse problem. The words "anti-sparsification" and "anti-sparsified" used later in the paper all mean the same thing), which can improve the recommendation system to

- *Pengli Nie, Jian Liu, Meiqi Feng and Weizhe Wang are with Tianjin Key Laboratory of Advanced Networking (TANK), College of Intelligence and Computing, Tianjin University, Tianjin, China, 300350. E-mail: pretty_pl@sina.com, jianliu@tju.edu.cn, fengmeiqi@tju.edu.cn, will@tju.edu.cn*
- *Guangquan Xu is with the School of Big Data, Qingdao Huanghai University and Tianjin Key Laboratory of Advanced Networking (TANK), College of Intelligence and Computing, Tianjin University, Tianjin, China, 300350. E-mail: losin@tju.edu.cn*
- *Litao Jiao is with the School of Big Data, Qingdao Huanghai University, Qingdao, China. Email: jiaolitao_11@163.com*
- *Shaoying Liu is with Graduate School of Advanced Science and Engineering, Hiroshima University, Japan. E-mail: sliu@hiroshima-u.ac.jp*
- *Weizhi Meng is with Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark. E-mail: weme@dtu.dk*
- *Hongyue Wu is with the College of Intelligence and Computing, Tianjin University, Tianjin, China, 300350. Email: hongyue.wu@tju.edu.cn*
- *Zhengjun Jing is with School of Computer Engineering, Jiangsu University of Technology, China. Email: zhengjun_jing@163.com*
- *Xi Zheng is with Department of Computing, Macquarie University, Australia. E-mail: james.zheng@mq.edu.au*

Guangquan Xu, Jian Liu and Hongyue Wu are the corresponding authors (Email: losin@tju.edu.cn, jianliu@tju.edu.cn, hongyue.wu@tju.edu.cn).

1. Taobao is a popular online retail platform in China. (https://www.taobao.com/)

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

get higher accuracy. Our main contributions are threefold.

- **F**irstly, we define the concept of *sparse trust* and propose a *sparse trust model* based on the concept, which lay the foundation for the trust related research in large recommendation systems. New trustworthy recommendation techniques can be developed based on the concept and the model.
- **S**econdly, we propose a new *sparse degree* to measure the sparsity of trust data. We then describe two new challenges regarding the big data sparse trust matrix, analyse the main features of the sparse trust, and discuss how the sparse trust is evaluated.
- **T**hirdly, we propose a novel data mining framework for alleviating the problem of sparse trust. It is based on the customized softmax function and a novel transitive gossip trust model. The data mining framework can effectively tackle the hidden trust information problem and the sparse evaluation problem in large recommendation systems.

The rest of the paper is organized as follows: Section 2 introduces the related research work on the sparse problem. Section 3 explores the concept of *sparse trust* and describes our proposed trust model. Section 4 elaborates on our *sparse trust mining framework* used in large-scale recommendation systems. Section 5 introduces the relevant parts of the experiments. Finally, Section 6 summarizes this paper and considers future research directions.

2 Related Work

Learning recommendation system and data mining plays a very important role for trust mining and application. Recommender Systems were often introduced to solve the major problem of information overload. A survey has been done where different approaches and techniques implemented to make recommendation for tweet [13]. Katarya [14] proposed an algorithm which adopts a self-devised approach and a two-level filtering to predict ratings and know about the nearest neighbors respectively to improve collaborative filtering. Gupta et al. [15] used reinforcement learning to implement dynamic recommendation, where the use of the Markov Decision Process is the innovative core of the article. Different from the traditional recommendation with "interested" or "non-interested", Yadav et al. [16] introduced fuzzy logic to simulate human behavior's flexible inference. Due to the increasing volume of social network data, the selection of neighbors in collaborative filtering has become more and more difficult. A hybrid model based on movie recommender system was proposed [17]. Type division method in the model plays a key role in reducing the computational complexity, and K-Means is used to improve the performance of particle swarm optimization. Collaborative filtering is a classic idea commonly used in recommender system [18], a hybrid recommender system has been proposed which utilized k-means clustering algorithm with bio-inspired artificial bee colony (ABC) optimization technique [19]. It improves the traditional collaborative filtering recommendation systems.

Group recommendation systems designed for efficient prediction in emerging applications such as big data, forecasting, medical diagnosis [20]. "Median strategy", [21] a group modeling strategy, was proposed based on collaborative filtering and group recommendations, where the user's preferences are modeled by cluster them into a group.

Learning affective recommendation systems (ARS) is helpful for us. The research process of ARS was classified and discussed from the aspects of methods, properties, differences, data sets, experimental platforms and evaluation [22]. Gautam et al. [23] introduces the concept of "opinion leader" into online social network, and studies its identification and usage in online social network and data mining. In the era of big data, data mining is obviously an important technology to deal with large data sets. According to different classification technology, researchers have done relatively perfect research on data mining classifier [24]. "CapsMF" [25], a novel product recommender system applying the advanced neural network architecture Capsule Networks (Caps) for document representation, integrates the Deep Neural Network text analysis model with the Probabilistic Matrix Factorization to improve recommendations.

Generally, reputation-based frameworks are adopted by researchers to overcome the sparseness of data in recommendation systems. Guo et al. [26] added reputation-based trust to the Singular Value Decomposition plus (SVD++) model. In daily life, people are usually unwilling to show trust information out of the protection of privacy [27], [28], [29], [30], which results in the limitations in SVD++ model. Yao et al. [31] considered the impact of trust on recommendation effects from both sides of a user's trust relationship; that is, each user is both a trustor and a trustee.

Some trust spread frameworks have been created to alleviate the problem of sparse trust. Kostas [32] proposed a Random Walk algorithm with a restart function to deal with social and friendship comments. In addition, Li et al. [33] proposed the T-broker model, which is a multi-cloud collaborative service broker scheme based on trust. Spertus et al. [34] recommended online communities based on the membership of users. Chen [35] proposed a method recommended for communities based on techniques of Latent Dirichlet Allocation and association rule mining. Wen et al. [36] proposed an analysis model with difference equation by considering different choices made by people and studying negative and positive propagation. In addition, Wang et.al. [37] studied the propogation of negative and positive news in the Internet and provided ideas for the modeling of trust propagation that motivates this paper.

In addition to using other data sources, anti-sparsity processing frameworks have been studied in the collaborative filtering based recommendation systems. Smyth and Donovan regarded trust as one-way data that corresponds to a user's expertise or ability by analyzing user behavior, and the potential trust is explored to reduce the sparsity of trust [38]. Massa and Ansari used explicit trust as a measure of user similarity that allows trust to be spread over a trust network. The propagation of trust greatly relieves the sparsity of trust matrix and improves the accuracy of the recommendation system [39], [40]. Golbeck proposed a framework based on trust propagation to predict the rating for the target project [41]. Hu et al. [42] proposed a reliable trust-based platoon service recommendation scheme, REPLACE, to help users avoid selecting shift vehicles with

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

poor performance. Specifically, the main idea of REPLACE is to design a reputation system for vehicles by collecting and modeling the feedback information of vehicle users. In addition, REPLACE includes an algorithm based on iterative filtering for unreal feedback from vehicle users.

Although the above trust models have their own advantages, none of the existing research is able to completely solve the problem of data sparsity caused by the cold-start and sparse trust evaluation. This problem is even more severe and intense in the context of big data [43]. Furthermore, the accuracy of the prediction must be considered. Based on our prior work on a rudimentary roundtable algorithm to mitigate the sparse trust issue, we introduced a thorough sparse trust model and a complete set of algorithms to further resolve the sparse trust issue. Compared with our prior work in [44], we conducted a much more thorough experiment, and our new results indicate that, compared with the current trust frameworks, our new set of algorithms can achieve much better accuracy in trust prediction of anti-sparsification in cold-start and all users environments, especially for datasets with high sparsity. We have summarized some of these highly relevant work in Table 1.

3 SPARSE TRUST MODEL

It is essential to formally define the concept of sparse trust to better study the problems in relation to sparse trust. This section describes relevant concepts and introduces our trust model and sparse metrics.

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.79 & 0.80 & 0 \\ 0.52 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.00 \\ 0 & 0 & 0 & 0.55 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (1)$$

3.1 Concept of Sparse Trust

Definition 1 (trust): Trust describes an emotional tendency of one party (trustor) being willing to rely on the actions of another party (trustee). The trust degree (emotional tendency) is often described in terms of probability, called the trust value, which describes the trust between the trustor and the trustee. As shown in Equ. (1), the trust value information is represented by a matrix, where each row and column refers to a person (either trustor or trustee), and each element refers to the trust between the trustor and the trustee of the corresponding row and column. The non-zero elements in the matrix represent the trust value between different users. The higher the value is, the higher the trust degree is. When the value is 0, there is no direct trust relationship between the trustor and the trustee.

Definition 2 (sparse trust): Trustor and trustee have no directly related evidence (including fuzzy information or data noise shielding evidence) that leads to a lack of trust between them, which is called sparse trust. For example, A and B are classmates, which indicates a latent trust relationship between A and B, but A does not show trust in B. The sparse trust of trustor A in trustee B is expressed in ternary form as $[A, B, P_{AB}]$, where P_{AB} is the trust value representing the trust between A and B. The sparse trust of trustor B in trustee A in ternary form is expressed as $[B, A, P_{BA}]$. We use P_{AB} to indicate how much entity A trusts entity B. It is also known as the trust value, represented by a real number between 0 and 1.

The causal model of sparse trust is presented in Fig. 1. It shows that trust is described by trust evidence and related information between the trustors and trustees. A relationship that is not described by direct or expressive evidence is called "true" sparse trust. In addition, a relationship that contains data noise or fuzzy information with masking evidence is called "false" sparse trust. As shown in Equ. (1), matrix T has the following ternary form: $[2, 4, 0.79] \rightarrow [4, 6, 1.00]$. Take the ternary form $[2, 4, 0.79]$ for example, the number 2 denotes the second row and the number 4 represents the fourth column in the matrix. The same goes for the others. Therefore, it implies a trust relation between entities 2 and 6 through intermediate entity 4, i.e., there is "false" sparse trust from entities 2 in 6. On the contrary, if we cannot build trust or there is no benefit for A, we can call it "true" sparse trust. For matrix T, there is ternary form $[2, 4, 0.79]$ but $[4, 3, 0]$ and there is ternary form $[2, 5, 0.80]$ but $[5, 3, 0]$, so "true" sparse trust exists between entities 2 and 3.

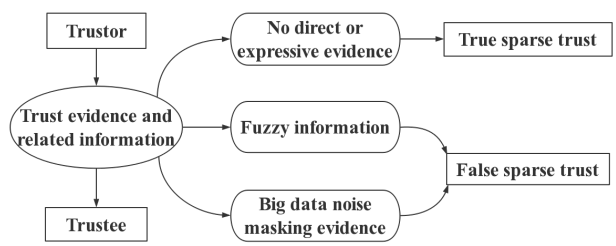


Fig. 1: Policy enforcement.

Trust is an emotional tendency and the degree of trust is described by a probability function and reflected by the emotional intensity, which is proportional to the probability description [45]. When the probability value is 1, the trustor has complete trust in the trustee. In contrast, when the probability value is 0, the trustor has no trust in the trustee. Sparse trust is described in ternary form because ternary form saves storage space, especially when the number of zero elements in the sparse matrix is much more than that of non-zero elements.

3.2 Trust Model

It is important to specify the type of model used as the reference of trust evaluation before studying the sparse trust problem. We construct a trust model targeting on large-scale recommendation systems (as shown in Fig. 2).

There are four elements in the trust model F_{X_i} (Entities mentioned below refer to the collection of all trustors and trustees):

- **C**entral entity X_i : X_i is the trustor of the Trust Model F_{X_i} , and the trustor in the trust relation;
- **T**rust track l : track l belongs to $[0, N]$, where N represents the infinite track. Input all elements on the i-th row of the trust matrix into the trust model F_{X_i} in order of the trust value from largest to smallest.

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

TABLE 1: Some of the typical related work.

Keywords	Core Idea	Year
Recommender system, Collaborative filtering, Capsule networks, Matrix factorization	CapsMF, a novel product recommender system applying the advanced neural network architecture Capsule Networks for document representation, integrates the Deep Neural Network text analysis model with the Probabilistic Matrix Factorization to improve recommendations.	2020
Recommender system, Collaborative filtering	An algorithm adopting a self-devised approach and a two-level filtering is aimed to predict ratings and know about the nearest neighbors respectively to improve collaborative filtering.	2018
Trust propagation	The paper studied the propogation of negative and positive news in the Internet and provided ideas for the modeling of trust propagation.	2018
Recommender systems, Reinforcement learning, Markov Decision	The paper used reinforcement learning to implement dynamic recommendation, where the use of the Markov Decision Process is the innovative core of the article.	2018
Recommendations, Group modeling	Median strategy, a group modeling strategy, was proposed based on collaborative filtering and group recommendations. The method with automatically detection implemented additive utilitarian, approval voting with threshold 1 and 2, least misery and most pleasure strategy from existing literature.	2017
Trust-based recommendation, Reputation system	A reliable trust-based platoon service recommendation scheme, REPLACE, aims to help users avoid selecting shift vehicles with poor performance, whose main idea is to design a reputation system for vehicles by collecting and modeling the feedback information of vehicle users.	2016
Recommender systems, Collaborative filtering, Fuzzy C-means	A hybrid model based on movie recommender system was proposed. Type division method in the model plays a key role in reducing the computational complexity, and K-Means is used to improve the performance of particle swarm optimization.	2016
Reputation-based frameworks, Trust, Matrix factorization	Trustsvd, a trust-based matrix factorization technique, added reputation-based trust to the Singular Value Decomposition plus (SVD++) model.	2015
Trust spread frameworks, Random Walk, Sparse trust	A Random Walk algorithm with restart function to deal with social and friendship comments to alleviate the problem of sparse trust.	2009
Trust propagation, Anti-sparsity	A framework based on trust propagation is proposed to predict the rating for the target project.	2006

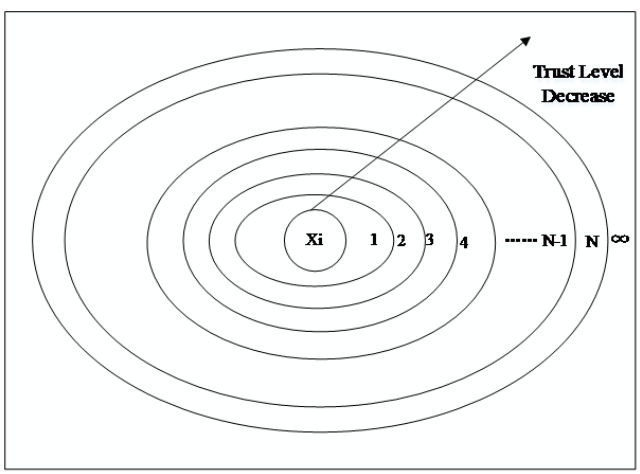


Fig. 2: Trust model F_{x_i}.

Non-zero elements with larger trust values are closer to internal track, and those with a value of zero are on the outermost track. The distance between the track and the central entity represents the trust of the trustor in the trustee, the closer the distance, the greater the trust value;

- **T**rack entity S_l : the entity S_l is the set of all the trustee entities which are located on track l in the trust model;
- **T**rust value T_{x_i,S_l} : trust value of the central entity X_i in the track entity S_l.

In Fig. 2, trustor X_i is the center of the trust model F_{X_i} and each ellipse represents a track l. T_{x_i,S_l} describes the trust value of central entity X_i to entity S_l, which is on track l. The same track may have zero, one, or multiple entities:

$$T_{x_i,S_0} = T_{x_i,x_i} = 1. \tag{2}$$

Equation (2) represents the self-confidence of central entity and that the default trust value is 1. It means the central entity on the 0 track.

$$0 \leq T_{x_i,S_{l+1}} < T_{x_i,S_l}, (l \in [0, N]) \tag{3}$$

Relation (3) describes that the distance between a track and the central entity represents the trust of trustor in trustee. The distance between trustee S_l and trustor X_i is less than the distance between trustee S_{l+1} and trustor X_i. Therefore, T_{x_i,S_l} is greater than $T_{x_i,S_{l+1}}$.

The left relation of (3) holds when $l \to \infty$ and the right relation holds when $l = 0$.

$$\lim T_{x_i,s_n} \tag{4}$$

The distance represents the level of trust between the track entity and the central entity. The closer the track entity to the central entity is, the greater the trust value of the track entity in the central entity will be. According to Equ. (4), the infinite track trust value is zero; that is, there is no trust relationship between the track entity and the central entity. In addition, due to the sparsity of data, the number of entities that are located on the infinite track may be much larger than the sum of the numbers of entities that

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

are located on other tracks for a central entity [44], [46], that is, the sparse trust is shown as below:

$$|S_{\infty}| >> \sum_{i \in \neg\infty} |S_i| \,. \tag{5}$$

A trust model of a central entity corresponds to a row of the trust matrix like in Equ. (1), and each value P_{ij} in the matrix represents the trust value of central entity i in entity j. Compared to the trust matrix, the trust model describes the trust-strength relationship more clearly.

For example, a trust model F_{X_2} is established based on the central entity X_2, which corresponds to the second row of the matrix in Equ. (1). Its associated entity X_4 and entity X_5 scatter in different trust tracks, while other entities' trust relations are located on the infinite track N. Entity 5 is closer to the central entity than entity 4 in the trust track, because $T_{2,5} > T_{2,4}$. The trust model F_{X_2} is shown in Fig. 3.

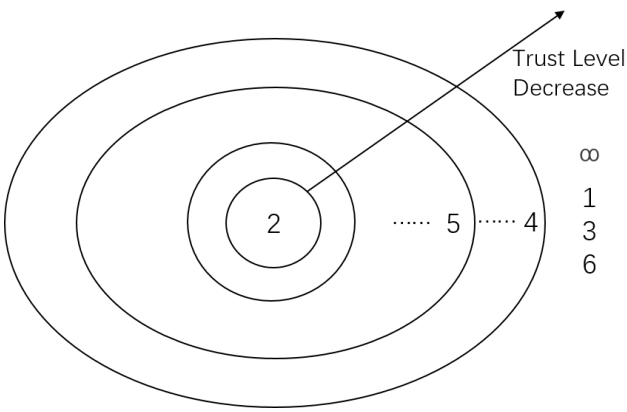


Fig. 3: Trust model F_{x_2}.

The above defined trust model has the following characteristics:

- **T**he trust is not symmetric. Suppose there are two entities, which are denoted as A and B. The trust of A in B is expressed as $[A, B, P_{AB}]$ and the trust of B in A is expressed as $[B, A, P_{BA}]$. P_{AB} is not necessarily equal to P_{BA};
- **E**ach trust model has only one central entity. All the entities in the dictionary are located on the trust tracks and most entities are located on the infinite track in sparse trust;
- **E**ach entity must be located on a unique track in the trust model. The same track may have multiple or no entities. The entities on the infinite track have no trust with the central entity (as shown in Fig. 2. by the vertical line "|");
- **T**he track of each trust model is divided into 0, 1, 2 ... N. The central entity is located at the 0 track and the trust value decreases from inner to outer until the outermost infinite track trust value is 0. The central entity has 100% trust in itself;
- **E**ach entity may be in multiple trust models at the same time and plays different roles in different trust models. Therefore, the trust values in different trust models are also different.

3.3 Sparsity degree

The more non-zero elements the trust matrix contains, the more latent associative relationships between the trustor and trustee exist. The number of zero elements and non-zero elements can be obtained by traversing the trust matrix. Then, the sparsity degree here is expressed by Equ. (6):

$$\text{Sparse}_v = \frac{\sum_{j \in n} \text{Num}_{nz}^{j}}{m \times n}, \tag{6}$$

where m is the number of matrix rows, n is the number of matrix columns, and Num_{nz}^{j} represents the number of nonzero elements in column j.

Note that there are far fewer non-zero elements than zero elements, and the cost is statistically less in a sparse trust matrix. Therefore, sparsity degree is measured by the number of nonzero elements. The smaller the value of $Sparse_v$ is, the more sparse the matrix becomes.

3.4 Challenges

Big data computation refers to using all the data in a large volume of data for analysis and processing, rather than using the method of random analysis (sample survey) as a shortcut [47].Therefore, big data sparse trust matrix has two main challenges:

High Sparsity: In most cases, the rating sparsity is higher than 85% under the cold-start condition [44]. Therefore, the sparse matrix provides little evidence for recommendation due to its sparsity, which is difficult for ordinary experimental methods. Our work is based on a trust transfer mechanism to achieve anti-sparsity. However, in practice, the sparsity degree of many recommendation systems is much less than 0.0001% (sparsity is higher than 99.9999%).

$$G_1 = \begin{bmatrix} 0&0&0&0&0&0 \\ 0&0&0&1&1&0 \\ 1&0&0&0&0&0 \\ 0&0&0&0&0&1 \\ 0&0&0&1&0&0 \\ 0&0&0&0&0&0 \end{bmatrix} \quad G_2 = \begin{bmatrix} 0&0&0&0&0&0 \\ 0&0&0&1&1&0 \\ 1&0&0&0&0&1 \\ 0&0&0&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&0&0 \end{bmatrix} \tag{7}$$

Irregularly Available Data: Different locations of data have different influences on the matrix. The entities need to rely on intermediate entities for trust transfer to build relationships. The distribution of non-zero elements will directly affect the transfer of trust relations. For example, the sparsity degrees of the two matrices G_1 and G_2 in (7) are both 13.89% . Matrix G_1 has the following ternary form: $[2, 4, 1] \rightarrow [4, 6, 1]$ and $[5, 4, 1] \rightarrow [4, 6, 1]$. Therefore, it will establish a trust relation between entities 2 and 6 through intermediate entity 4, and between entities 5 and 6 through intermediate entity 4. However, matrix G_2 has ternary form $[4, 6, 0]$ rather than $[4, 6, 1]$ and it cannot establish a trust relation between entities 2 and 6, or entities 5 and 6. This implies that even in two matrices with the same size and the same sparsity degree, different node positions will cause different anti-sparsity effects.

With the explosion of big data, the above two challenges have become increasingly severe.

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

4 SPARSE TRUST DATA MINING FRAMEWORK

In this section, we describe our sparse trust data mining framework.

4.1 Sparse Trust Mining Model

Assume that two people (entities) without a trust relationship can transfer the trust through a mutual friend (intermediate entity). For example, suppose A purchased B's product and gave a positive review, and C trusts A or has the same hobby as A. We can infer that C has a high probability of trusting (purchasing) B's product. This is very common in real life.

Nowadays, the traditional recommendation systems are affected by the problem such as the cold-start. The trust-based recommendation systems can effectively alleviate the problem, but the trust data is even more sparse. Our framework can overcome the sparse trust problem in recommendation system, and can estimate and establish the sparse trust relationship more effectively.

Our sparse trust mining model is depicted in Fig. 4.

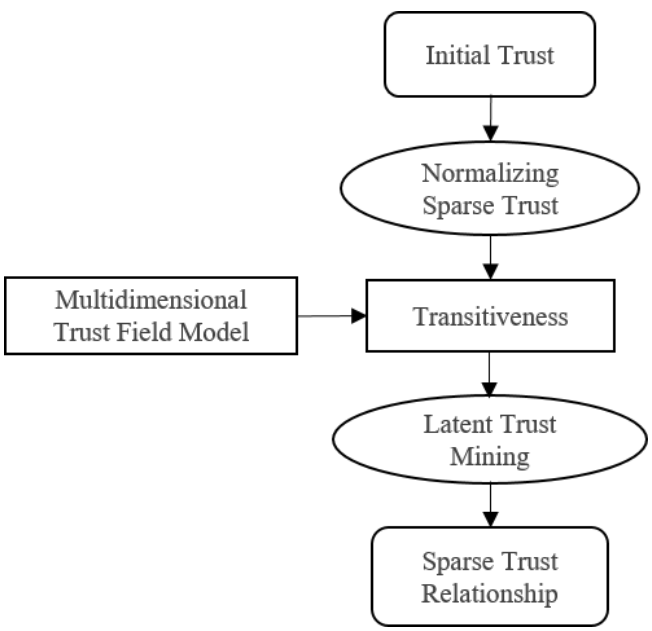


Fig. 4: Sparse trust mining model.

The framework is divided into five parts. Part 1: input the initial trust value into the system in a ternary form. The system converts the received trust values into an adjacency list for calculation. The sparse trust matrix is anti-sparsified by our data mining framework. Part 2: Sparse trust normalization is performed to ensure all entities can fit on a round table, reduce the calculation cost and eliminate the Grade Inflation [48]. Part 3: Transitive gossip trust. Before transferring trust, the system constructs the multidimensional trust model, which lays the foundation for latent trust mining. Part 4: Latent trust mining is carried out. The system makes full use of the trust model to mine multi-modal social relationships in different communities. Part 5: the sparse trust mining model outputs the sparse trust.

4.2 Normalizing Sparse Trust

The roundtable algorithm[2] was used in a combat game called World of Warcraft. The algorithm in this game can be regarded as the origin of the round table algorithm, where all possible states and event sets are abstracted as a round table. In essence, it puts all the possible events on the round table according to priority until all events are placed or the round table is filled. In contrast, the waterfall algorithm uses a top-down approach to in-game item selection, which is the decision algorithm used in most games. Compared with the waterfall algorithm, the roundtable algorithm has a lower computational overhead (it only needs to make a decision once, rather than step by step) and its flow can be summarized as follows:

- **S**tep 1: First, abstract all the entities in the same community into a round table.
- **S**tep 2: Put the entities on the table according to the trust value until all entities are placed or the round table is completely filled..
- **S**tep 3: If there is an entity that is still not placed when the round table is filled, the entity will be eliminated.

The advantages and disadvantages of the roundtable algorithm are clear at a glance:

Based on Step 1 and Step 2, an abstract modeling process can mine the latent relationship of all entities in the same community at a low cost. Additionally, at the round table, the trust value of each entity has not changed, and it still truly reflects the trustor's emotional tendencies towards the trustee. However, it is very unfair to eliminate a low-priority entity while ensuring that entities with high trust values have high priority and is not conducive to the discovery of intent relations. Therefore, we propose our data mining algorithms.

There are different levels of trust for the entities used for trust mining in the same network community. What's more, different users do not adopt a uniform standard to rate items. In this case, some people will give a score that is far from the truth, while others are used to grading within a small range. This phenomenon is often referred to as Grade Inflation and can be frequently encountered in the field of data mining [48]. In order to solve the problem of Grade Inflation, we adopt a normalized approach to deal with all entities in a round table. We utilize the softmax function [49] (G_{ij}) to normalize sparse trust:

$$G_{ij} = \frac{e^{p_{ij}}}{\sum_{j \in I} e^{p_{ij}}}, \tag{8}$$

where variable I stands for the network community. The function makes sure that the sum of the normalized trust values is 1. It is important for the execution of several steps in our algorithms. Notice that if $e^{p_{ij}} = 1$, the trust value will be 0, which indicates distrust. For example, if an entity does not interact with any other entities, the value of trust between it and other entities is represented by 0. However, the softmax function also computes a value for it. This is not realistic. This situation is common in practice, so the handling of these entities is particularly important. In our work, we define the value of G_{ij} in this case as zero, which

2. https://blog.csdn.net/qq_35361004/article/details/80989781

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

is used to mine the trust value. Furthermore, only those nodes that have a trust value are normalized. The improved softmax function is shown in (9).

$$G_{ij} = \begin{cases} \frac{e^{p_{ij}}}{\sum_{j \in I} e^{p_{ij}}}, \text{ if } p_{ij} \neq 0, \\ 0, \text{ otherwise .} \end{cases} \tag{9}$$

The normalized sparse trust values do not distinguish between a trustee with whom a trustor did not interact and a trustee with whom a trustor had a good intersection. Moreover, if $G_{ij} = G_{ik}$, we can conclude that both entity j and k are trusted by entity i with the same degree of sparsity, but we don't know if they have the same interaction, whether they are all highly trustworthy or whether they are all untrustworthy. In addition, the two trust values may come from online communities with different evaluation criteria. In other words, these normalized sparse trust values are relative rather than absolute. After the trust value is normalized, the relativity will not be changed, and the result is still satisfactory. A very important advantage of this approach to normalize sparse trust values is that we do not need to re-normalize each iteration (which is prohibitively costly in the trust transitive process), which saves us relevant expenses and is also the reason why we choose this approach. This method solves the problem of Grade Inflation and plays an important role in the whole model. The calculation process is shown in Algorithm 1.

Algorithm 1 Normalizing Sparse Trust

Input: Sparse Trust Matrix, P_{ij} ; the Number of Entities m.
Output: Normalized Sparse Trust Matrix: G_{ij}.

```
for i ← 1 to m do
    for j ← 1 to m do
        if e^{p_ij} ≠ 1 then
            p_{i.} ← e^{p_ij} ;
for i ← 1 to m do
    for j ← 1 to m do
        if e^{p_ij} ≠ 1 then
            g_ij ← e^{p_ij} / p_{i.} ;
        else g_ij ← 0 ;
```

4.3 Transitive Gossip Trust

4.3.1 Multidimensional Trust Model

The abovementioned normalized sparse trust effectively overcomes the unfairness of the elimination mechanism of the original round table algorithm. Next, we structure the multidimensional trust model of entities in the same community in a specified way, which is of great significance for subsequent trust transfer mining and potential association mining. The trust value on the diagonal of the square trust matrix represents the self-confidence of the entity. We specify that each entity fully trusts itself, and the central entity has full self-confidence in the trust model that is defined above. The default trust value is 1 and the initial trust value of each entity is between 0 and 1; thus, the value on the diagonal is the largest in the normalized trust matrix.

The trust model describes the trust of the central entity in other entities. The trustee's distance from the central entity reflects the trust strength. Furthermore, the track on which the trustee is located also reflects the extent of its trust. Each trust model, as shown in Fig. 2, corresponds to a row of data in the trust matrix, which is similar to the multi-dimensional superposition trust model:

- **E**ntity i as the center of trust model F_i : Entity j is on the track $a(a \ll \infty)$, and the trust value of entity i in j is g_{ij} $(0 < g_{ij} < 1)$; Entity k is on the track $l(l = \infty)$, and the trust value of entity i in k is $g_{ik} = 0$ (k does not have a trust with central entity i), as shown in Fig. 5(a);
- **E**ntity j as the center of trust model F_j : Entity k on track $b(b \ll \infty)$; the trust value of the entity j in k is g_{ik} $(0 < g_{ik} < 1)$, as shown in Fig. 5(b);
- **S**uperposition of trust model F_j onto trust model F_i , the position on the round table conforms to the relationship of multiple trust models, and the two-dimensional superposition trust model is shown in Fig. 5(c). By analogy, when there is more than one entity in the same community, we structure the multidimensional trust model of entities in this way.

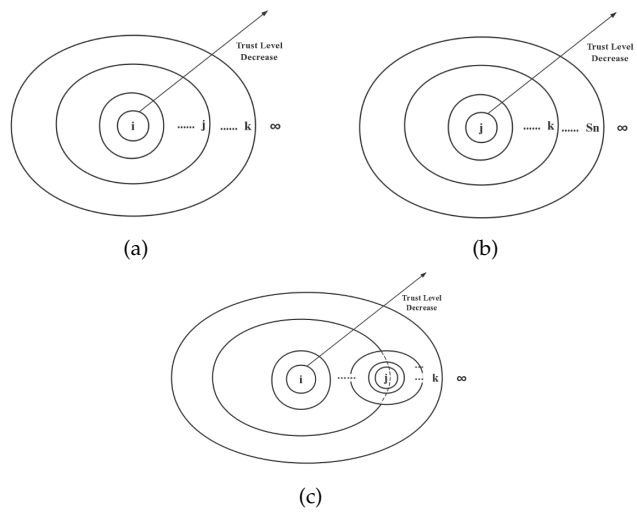


Fig. 5: Multidimensional trust model.

4.3.2 Transitiveness

We aim to mine and build sparse trust for more entities, which is not so obvious, but still similar to mining the classic relationship between diapers and beer. Our data mining algorithm is based on a trust transitive mechanism and aims at finding acquaintances. In other words, "acquaintances" here refers to common neighbors. It makes sense to weigh their opinions according to the trust that the trustor places on them:

$$t_{ik} = \sum_{j \in I} g_{ij} g_{jk}, \tag{10}$$

where t_{ik} represents the trust that entity i places in entity k, which is determined by polling their acquaintances.

In the example of Section 4.3.1, trust values g_{ij} and g_{ik} are both less than 1, so when j is constant, the calculated t_{ik} is less than g_{ij} and g_{ik} . Thus, the trust value of the transmission is attenuated, and the more intermediate entities there are, the greater the attenuation is. Because g_{ij} and g_{ik} are independent of each other, cross-field indirect trust can be considered to have the Markov property. The sparse

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

trust value is described by the stationary distribution of the Markov chain that is defined by the normalized trust matrix G.

Gossip trust is, as the name suggests, the principle that is consistent with the gossip algorithm, inspired by the gossip of the office. As long as one person goes about a gossip, all people will know the information about the gossip in a limited time. Trust relationship transfers like a virus, spreading on the round table. It should be noted that the framework of the gossip is similar to the random-surfer model based on the user's behavior of browsing the Web, which is a direct probabilistic explanation of it [50].

In the above example, according to the transitivity of trust, trust models F_i and F_j are superimposed on the round table, and the infinite track entity k infects the trust with the central entity i. When an intermediate entity is seeking trustworthy entities, it can obtain gossip trust values in the network with the following manner: at each entity i, it can gossip with entity j with probability g_{ij}. After transferring for a while in this manner, the intermediate entity is more likely to be a trustworthy entity than an untrustworthy entity, and entity j is the only information source and the intermediate entity. Intermediate entities, as the name suggests, exist between two entities that do not trust each other directly, "bridging" them and passing trust values. In our algorithm, each entity variable contains two quantities of information: an adjacency list of nodes that has the trust with the entity node, which is denoted as $Trustlist(i)$, and a set of their trust values, which is denoted as $Data_set\,(g_{ij})$. Intermediate entities are found using a depth-first search, marked, and stored in the stack. Furthermore, the searched effective path is stored in matrix T_{ij}. If it is still not a valid path after searching the last entity in $Trustlist(i)$, it will return to the previous parent node to search for the next path. The search ends when all valid paths have been explored. This calculation method is shown in Algorithm 2.

Algorithm 2 Transitive Gossip Trust Model

Input: Normalized Sparse Trust Matrix: G_{ij}.

Definitions: The number of nonzero elements of matrix $t^{(r)}$: $\mathrm{Num}_{nz}^{t^{(r)}}$.

Output: Trust value of entity i in entity k through r intermediate entities : $t_{ik}^{(r)}$.

Initial: $t_{ik}^{(0)} = g_{ik}$;
repeat
  $t_{ik}^{(r+1)} \leftarrow g_{i1}t_{1k}^{(r)} + g_{i2}t_{2k}^{(r)} + \cdots + g_{ij}t_{jk}^{(r)}$;
until $\left|\mathrm{Num}_{nz}^{t^{(r+1)}} - \mathrm{Num}_{nz}^{t^{(r)}}\right| < \varepsilon$.

4.4 Latent Trust Mining

When it comes to trust, we always think about social relationships, which can be divided into single and multiple modes. Many researchers have therefore resorted to addressing trust scarcity by referring to "sparse" information in social networks. In the above example, j is the only intermediate entity. In general, the trust mining between entities in the same community is called isomorphic association rule mining, and correspondingly, the trust mining between entities in different communities is heterogeneous association rule mining. In practice, physical entity users typically cannot interact across all communities. By mining the potential information of one community, their trust can be used and gathered in other communities, which is of great significance for the anti-sparsity of sparse trust matrix. Therefore, we considered the issue of tapping into potential trust in multiple online communities.

Section 4.3.2 explains how to transfer trust based on the trust model. In the subsequent sections, we will discuss ways to incorporate trust from different communities. First, we need to explain how the trust relationship is calculated. We can calculate the sparse trust value of any complete path between two entities (such as from entity i to entity k) according to Equ. (10) and save the result to the corresponding matrix $T_{ik}^{(r)}$, where r represents the number of intermediate entities. The calculation method of latent trust mining is shown in Algorithm 3.

Algorithm 3 Aggregate the mining trust

Input: Normalized Sparse Trust Matrix: G_{ij} ; entity i to entity k gossip through r intermediate entities: $t_{ik}^{(r)}$.

Output: Sparse Trust Matrix: $T_{ik}^{(r)}$.

for $i \leftarrow 1$ to $r-1$ **do**
  $t_{ik} \leftarrow t_{ik}^{(r)}$;
$T_{ik} \leftarrow t_{ik}/r$;
return T_{ik}.

4.5 Overview of our data mining scheme

In the example in Section 4.3.1, j is the only intermediate entity. Entity j is directly trusted by entity i, entity k is directly trusted by entity j, and entity i has cross-field trust in entity k based on entity information source j. However, the number of intermediate entities is greater than the one in practice. Here we describe the anti-sparsification scheme to compute the sparse trust values. The abovementioned two-dimensional superposition trust model with single information source shows the simplest cross-field indirect trust. The problem of data sparsity in multi-domain social relationships is solved by using heterogeneous entity association schemes between trustors and trustees. In some cases, an entity can have an inter-community association with another entity that is in a different network community. The algorithm proposed in this paper aggregates multiple trust paths, each involving the passing of trust across multiple entities. A general overview of our data mining scheme is shown in Algorithm 4. Fig. 6 depicts the detailed flow of this work. Before normalizing the sparse trust to mine more trusts (refer to Section 4.3.1), we initialize the self-trust values of the entities in step 1 (line 3).

4.6 An Example

Assuming that there is an original sparse matrix S:

$$S = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix},$$

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

Algorithm 4 The Sparse Trust Data Mining Algorithm

Input: Ternary Form $P : [i, j, P_{ij}]$.
Definitions: Sparse Matrix: S_Matrix ; Number of nonzero elements: Num_{nz} ; Number of Entities: m; Trust value of entity i in entity k when gossiping through r intermediate entities: $t_{ik}^{(r)}$.
Output: Trust Matrix T .

Initial: $S_\ \mathrm{Matrix}(i, j) \leftarrow P_{ij}$;
for $i \leftarrow 0$ to m **do**
  $S_$ Matrix $(i, i) \leftarrow 1$ //Initialize Self Confidence
  for $j \leftarrow 1$ to m **do**
    if $e^{p_{ij}} \neq 1$ **then**
      $p_{i.} \leftarrow e^{p_{ij}}$;
for $i \leftarrow 1$ to m **do**
  for $j \leftarrow 1$ to m **do**
    if $e^{p_{ij}} \neq 1$ **then**
      $g_{ij} \leftarrow \frac{e^{p_{ij}}}{p_{i.}}$;
    else $g_{ij} \leftarrow 0$;
$t_{ik}^{(0)} = g_{ik}$;
repeat
  $t_{ik}^{(r+1)} \leftarrow g_{i1}t_{1k}^{(r)} + g_{i2}t_{2k}^{(r)} + \cdots + g_{ij}t_{jk}^{(r)}$;
until $\left|\mathrm{Num}_{nz}^{t^{(r+1)}} - \mathrm{Num}_{nz}^{t^{(r)}}\right| < \varepsilon$;
for $i \leftarrow 1$ to $r - 1$ **do**
  $t_{ik} \leftarrow t_{ik}^{(r)}$; //Aggregate the Mining Trust
$T_{ik} \leftarrow t_{ik}/r$;
return T_{ik}.

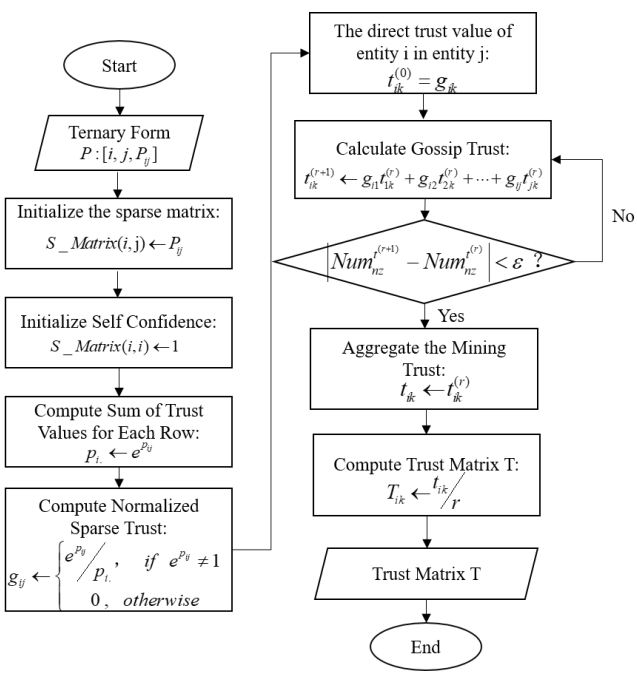


Fig. 6: Detailed flowchart of *sparse trust data mining*.

after normalization according to the idea of this study, matrix G can be obtained:

$$G = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix},$$

then,

$$t^{(0)} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix},$$

the number of nonzero elements of matrix $t^{(0)}$: $Num_{nz}^{t^{(0)}} = 6$.

According to the Transitive Gossip Trust Model(i.e., algorithm 2), we calculate Gossip Trust as follows:

$$\begin{aligned}
t_{11}^{(1)} &= g_{11}t_{11}^{(0)} + g_{12}t_{21}^{(0)} + g_{13}t_{31}^{(0)} = \frac{1}{2}, \\
t_{12}^{(1)} &= g_{11}t_{12}^{(0)} + g_{12}t_{22}^{(0)} + g_{13}t_{32}^{(0)} = 0, \\
t_{13}^{(1)} &= g_{11}t_{13}^{(0)} + g_{12}t_{23}^{(0)} + g_{13}t_{33}^{(0)} = \frac{1}{2}, \\
t_{21}^{(1)} &= g_{21}t_{11}^{(0)} + g_{22}t_{21}^{(0)} + g_{23}t_{31}^{(0)} = \frac{1}{4}, \\
t_{22}^{(1)} &= g_{21}t_{12}^{(0)} + g_{22}t_{22}^{(0)} + g_{23}t_{32}^{(0)} = \frac{1}{4}, \\
t_{23}^{(1)} &= g_{21}t_{13}^{(0)} + g_{22}t_{23}^{(0)} + g_{23}t_{33}^{(0)} = \frac{1}{2}, \\
t_{31}^{(1)} &= g_{31}t_{11}^{(0)} + g_{32}t_{21}^{(0)} + g_{33}t_{31}^{(0)} = \frac{1}{2}, \\
t_{32}^{(1)} &= g_{31}t_{12}^{(0)} + g_{32}t_{22}^{(0)} + g_{33}t_{32}^{(0)} = 0, \\
t_{33}^{(1)} &= g_{31}t_{13}^{(0)} + g_{32}t_{23}^{(0)} + g_{33}t_{33}^{(0)} = \frac{1}{2}.
\end{aligned}$$

The number of nonzero elements of matrix $t^{(1)}$: $Num_{nz}^{t^{(1)}} = 7$.

Similarly, we can get: $t^{(2)} = G \cdot t^{(1)}$, thus,

$$t^{(2)} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

The number of nonzero elements of matrix $t^{(2)}$: $Num_{nz}^{t^{(1)}} = 7$.

Set $\varepsilon = 0.1$, $\left|Num_{nz}^{t^{(2)}} - Num_{nz}^{t^{(1)}}\right| < \varepsilon$. It's just going through an intermediate entity to get the mining trust value, so the final trust matrix $T = t^{(1)}/1$.

This example aims to help readers better understand the idea of the algorithm. The data in the example is not a large-scale sparse matrix, so the calculation results are not of reference significance.

5 EXPERIMENTS

In this section, we aim to assess the performance of our framework in modeling and measuring mining trust. We conduct comprehensive experiments with real-world and synthetic datasets. By calculating the trust matrix of different sparsity degrees, the anti-sparsity ability of the proposed algorithm is evaluated.

5.1 Dataset

The experimental datasets are introduced as follows (for the datasets, our last visit was in December 2017):

Part I: To sufficiently validate the performance of our proposed framework, we choose three representative datasets that are related to propagation framework for our experiments, which are taken from three

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

TABLE 2: Datasets statistics from popular social networking websites.

Statistics	Flixster	Epinions	Douban	CiaoDVDs
Num.Users	147,612	49,289	129,490	7,375
Num.Items	48,794	139,738	58,541	99,746
Num.Ratings	8,196,077	664,823	16,830,939	278,483
(Sparsity degree)	0.1138%	0.0097%	0.2220%	0.0379%
Friends/User	17.20	9.88	13.22	15.16

popular social networking websites respectively, including Douban[3] (www.douban.com/) [51], Epinions[4] (https://www.epinions.com/) [52] and Flixster[5] (https://www.flixster.com/) [53].

Douban: Douban, founded on March 6, 2005, is a Chinese 2.0 site offering user reviews and recommendations of film,book and music. It is also one of the largest online communities in China. This anonymized Douban dataset accommodates 58,541 unique movie items, 129,490 unique users and 16,830,839 movie ratings. For this social network, there are a total of 1,692,952 claimed social relationships.

Epinions: Epinions is a website where people can review products. One feature of Epinions is that users are paid based on the number of reviews (revenue sharing schemes) that are deemed useful. The dataset contains 487,181 social relationships, 139,738 items, 49,289 users and 664,823 ratings.

Flixster: Flixster is a social movie site that allows users to share movie ratings to meet others with similar movie tastes and discover new movies. Most users in this dataset do not rate any items, so we removed users and items that were not meaningful to the experimental evaluation and did not involve any ratings. Finally, the dataset that we obtained contains 2,538,746 social relations, 8,196,077 ratings, 48,794 items and 147,612 users [54].

Part II: The experimental dataset comes from CiaoDVDs [55], and we use the latest update of the dataset from December 2017. In this dataset, users can rate products that they have bought or used in the past. Other users can rate how helpful the comments are on a rating scale. This dataset reflects the typical social evaluation network and is often used by previous studies. The statistics of these datasets are presented in Table 2.

We apply our anti-sparsity trust mining algorithm to these datasets. Entity relevance values are contained in a file in a tab-delimited format. Each line of this file represents the relevance of entities, and the experimental data are represented in ternary form (as defined in Section 3.1). Relevance values in Part I are on a continuous 0-1 scale. A value of 1 indicates that the trustee is strongly relevant to the trustor and 0 indicates no relevance.

In order to achieve a good simulation effect, we need to process the original trust matrix in some non-sparse data sets to obtain the sparse trust matrix with the required degree of sparsity. We add a random function in the process to determine whether the non-zero trust value of the current position of the matrix is valid by comparing the random value with the preset threshold value. If it is valid, it will remain unchanged; if not, the original trust value of the current position will be set to zero. The trust matrix of the target sparsity is obtained through continuous iteration. In other words, we have determined the value of sparsity in advance. We consider simulating different big data matrices, where a big data matrix is used to infer the trust. Moreover, we fully use the self-confidence of the entity to reasonably update the existing trust (performance analysis results are shown in Section 5.4). We need to change the matrix to verify the performance of the algorithm. Therefore, it is important to obtain matrices with different sparsity degrees. In the data processing, random function values for eight sparsity degrees were obtained through repeated experimental comparisons. Algorithm 5 shows the general process of processing the matrix in the original data set using a random function. Furthermore, the number of non-zero elements in the initial dataset is relative (the error range of sparsity degree is 0.0005%).

Algorithm 5 Process the data set CiaoDVDs

Input: Original Trust Matrix, P_{ij} ; the Number of Entities m; Random number comparison threshold x ; Expected sparsity t.
Output: Sparse Trust Matrix: P'_{ij}.

```
repeat
    for i ← 1 to m do
        for j ← 1 to m do
            p'_ij ← p_ij ;
            if x > rand() then
                p'_ij ← 0 ;
    Calculate the sparsity degree of p'_ij ; Sparse_P' ;
    Change threshold x;
until Sparse_P' = t;
return P'_ij.
```

5.2 Evaluation Metrics

Our experiment selects two evaluation indicators to measure the prediction quality of the trust communication framework of the evaluated method (the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE)) [56], which are used to measure the deviation between the predicted value and the original value. The smaller the deviation is, the better the approach becomes.

(1) The metric MAE is defined as:

$$MAE = \frac{\sum_{ij} \left| t_{ij} - \hat{t}_{ij} \right|}{N}, \tag{11}$$

(2) The metric RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_{ij} \left(t_{ij} - \hat{t}_{ij} \right)}{N}}, \tag{12}$$

where t_{ij} denotes the original trust value that entity i gave to entity j, $\hat{t}_{ij}$ denotes the predict trust value that entity

3. From an anonymized Douban dataset. Two files are included in this Douban dataset, the user-item rating file and the user social friend network file. (https://www.cse.cuhk.edu.hk/irwin.king.new/pub/data/douban)

4. From a website where we can download the data set directly. (http://www.trustlet.org/wiki/Epinions_datasets)

5. From the first interactive network repository with visual analytic tools, which is the largest network data repository with thousands of network data sets. (http://networkrepository.com/soc-flixster.php)

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

i gives to entity j, as predicted by the framework, and N denotes the number of tested trust values.

5.3 Comparative Experimental Studies

In this section, we evaluate the performance of our proposed framework by comparing it with relevant frameworks.

5.3.1 Comparison frameworks

We choose the following three typical trust propagation frameworks for comparison: Tidal Trust (TT, 2005) [56], Mole Trust (MT, 2007) [57], and Multi-Faceted Trust (MFT, 2015) [58]. All these frameworks predict trust leveraging the trust transfer mechanism. MFT is the latest and most relevant to our work in the three benchmark algorithms. We conduct comparative experiments under all users and cold-start environments to verify the anti-sparsity accuracy of our data mining algorithm. In the following, our sparse trust data mining algorithm is called "Our Solution".

In order to ensure fair comparison and accurate verification, we set the optimal parameters for all the compared frames according to their respective references or on the basis of our experiments:

- **T**idal Trust: max $= 0.008$;
- **M**ole Trust: $mpd = Num.$iterations ;
- **M**ulti-Faced Trust: $\varepsilon = 0.1$ and $\theta = 0.1$.

5.3.2 Compared-Validation

We make predictions and tests through comparative verification methods. Randomly dividing the trust data into two equal parts, we use one part as the predict set (50 percent of the sparse trust data) and the other part as the test set (the remaining 50 percent of the sparse data) in each experiment. In addition, we conduct each experiment five times and take the mean as the final result for each experiment.

5.4 Results and Analysis

The anti-sparsification accuracy of the proposed data mining algorithm is verified by comparing the experimental results on the three datasets with those of the competing frameworks. Table 3 and Table 4 respectively show the results of MAE and RMSE on testing of all users and cold-start, which were computed based on the user's predictions. In addition, Fig. 7 and Fig. 8 show the performance comparison histogram of the experiments (In order to improve the display effect of the data (the numerical gap is large), the truncation diagram is adopted in Fig. 7(b) and Fig. 8(b)).

5.4.1 Validation on All Users

Our algorithm outperforms the other approaches in both MAE and RMSE on all three datasets obviously. Overall, MFT framework achieves the second-best performance on the three datasets in MAE, except for dataset Epinions. Because the trust data of Epinions is extremely sparse (sparsity degree of 0.0097%), the traditional framework of TT could perform much better than the frameworks of MFT and MT.

However, when applied to experimental data with relatively dense trust (such as 0.2220% sparsity of Douban), the performance of MFT is relatively significant. While in the Flixster (sparsity degree of 0.1138%) dataset, MT performs better. Finally, TT is better suited to Epinions containing directed trust networks than MFT or MT because it performs more accurately. On Flixster and Douban, which include undirected social networks, MT has similar performance to TT. Therefore, the combined effect of these methods is not satisfactory for improving the comprehensive quality of the recommendation system. In general, the performance of MFT is optimal among the baseline methods.

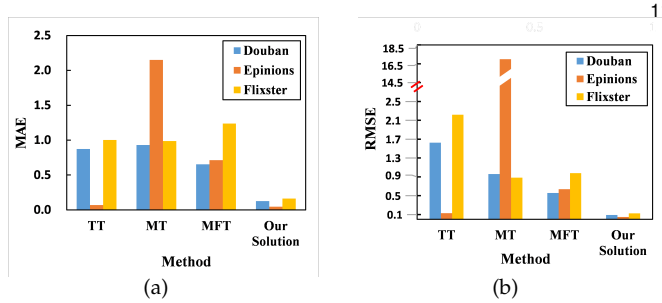


Fig. 7: Performance comparison on testing of all users. (a) MAE comparison on testing of all users; (b) RMSE comparison on testing of all users.

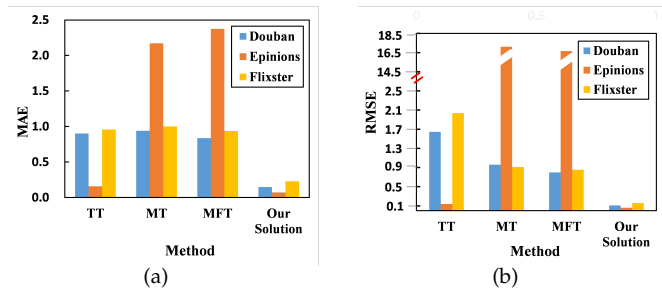


Fig. 8: Performance comparison on testing of cold-start. (a) MAE comparison on testing of cold-start; (b) RMSE comparison on testing of cold-start.

At the same time, it is clear that our framework stands out in terms of improving recommendation quality compared to the baseline approaches, as demonstrated by testing all users (as shown in Table 3). Especially for Epinions, a highly sparse data set, the anti-sparsity effect of our framework is even more significant. In the experiment of testing all users, our method can realize anti-sparsity well, and it is obviously better than the comparison methods. Therefore, it can be concluded that our method is an effective and reliable anti-sparsification framework, and it has certain advantages in stability.

5.4.2 Validation on cold-start Users

We also evaluate the accuracy of the algorithm against trust sparsity in a cold-start environment. Typically, we refer to users who trust fewer than five as cold-start users. Compared-validation is still used in the test but we only care about the accuracy of predictions for cold-start users (with five or fewer trusts) at this moment. Table 4 shows that our algorithm still has the best performance on all datasets, indicating that our algorithm can be implemented with accuracy of anti-sparsification on the testing of cold-start.

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

TABLE 3: Experimental results on testing of all users.

Datasets	Measure	MT	MFT	TT	Our solution
Douban	MAE	0.9309	0.6531	0.8703	0.1267
	(Improve)	86.390%	80.60%	85.442%	-
	RMSE	0.9582	0.5579	1.6258	0.0922
	(Improve)	90.378%	83.47%	94.329%	-
Epinions	MAE	2.1507	0.7124	0.0700	0.0474
	(Improve)	97.796%	93.35%	32.286%	-
	RMSE	16.9959	0.6395	0.1286	0.0480
	(Improve)	99.718%	92.49%	62.675%	-
Flixster	MAE	0.9873	1.2379	1.0025	0.1636
	(Improve)	83.430%	86.78%	83.681%	-
	RMSE	0.8854	0.9784	2.2191	0.1257
	(Improve)	85.803%	87.15%	94.336%	-

TABLE 4: Experimental results on testing of cold-start users.

Datasets	Measure	MT	MFT	TT	Our solution
Douban	MAE	0.9412	0.837	0.9016	0.1495
	(Improve)	84.12%	82.14%	83.42%	-
	RMSE	0.9603	0.7979	1.6473	0.1103
	(Improve)	88.51%	86.18%	93.30%	-
Epinions	MAE	2.1705	2.3741	0.0918	0.0714
	(Improve)	96.71%	96.99%	22.22%	-
	RMSE	17.1202	16.6594	0.1416	0.0606
	(Improve)	99.65%	99.64%	57.20%	-
Flixster	MAE	1.0006	0.9375	1.0173	0.2305
	(Improve)	76.96%	75.41%	77.34%	-
	RMSE	0.9107	0.856	2.0386	0.1576
	(Improve)	82.69%	81.59%	92.27%	-

The experiment indicates that, compared with current trust frameworks, our data mining algorithm's trust prediction of anti-sparsification is more accurate for all users and the cold-start environment.

5.5 Performance Analysis

We use the dataset CiaoDVDs for our experiments and compare the effectiveness, stability and updateability of the algorithm in three different cases.

Effectiveness: In this part, two experiments are conducted to investigate the effectiveness of the proposed algorithm, which is designed according to the two challenges mentioned in Section 3.4.

In the first experiment, eight kinds of sparsity degree trust data from CiaoDVDs are used as experimental data. These sparsity degrees of eight data are as follows: 0.009%, 0.018%, 0.028%, 0.039%, 0.048%, 0.052%, 0.064%, and 0.072%. Obviously, these are sparse matrices [47]. Fig. 9 gives a histogram that illustrates the anti-sparsification situation and each dataset corresponds to one of the eight trust data mentioned above. The experimental results in Fig. 9 are the average values over many experiments, because each experimental result is influenced by many factors and single experiments are unreproducible. As shown in Fig.8, the algorithm can achieve anti-sparsification. In particular, the more the evidence is provided initially, the greater the effect of the anti-sparsity is.

The second experiment is based on another challenge that the distribution of non-zero elements. For different trust matrices with the same sparsity, we conduct several

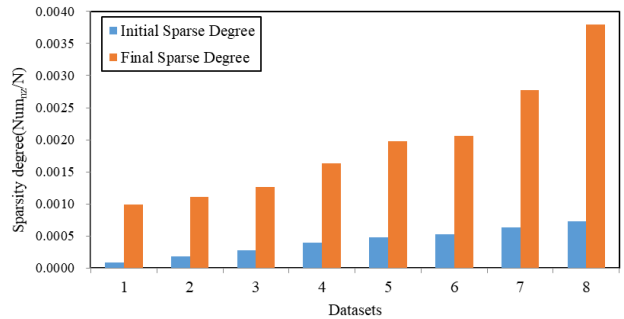


Fig. 9: Column chart of datasets CiaoDVDs of eight different sparsity degrees, compared after anti-sparsification operations.

experiments respectively, and select 10 representative experimental data from them, all of which have a sparsity of 0.018%. As shown in Fig. 10, each column represents the residual $|\text{Sparse}'_v - \text{Sparse}_v|$ of the sparsity degree. Different positions of non-zero elements lead to different anti-sparsity results, so the results of the ten groups of experimental data are not the same. The location of non-zero elements determines the number of intermediate entities and the number of mining paths. Nevertheless, for any initial sparse matrix, our method can improve its anti-sparsity to the maximum extent.

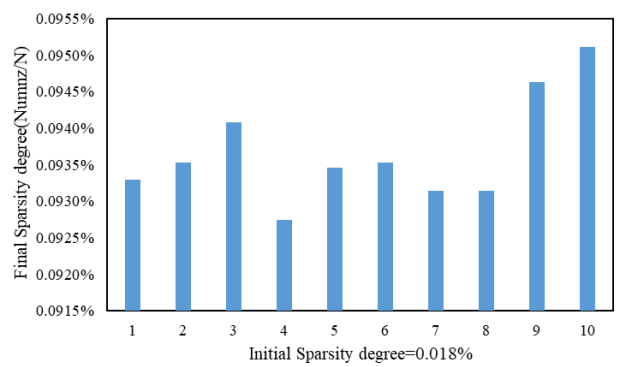


Fig. 10: Column chart of 10 datasets' sparsity degrees after anti-sparsification operations, which have the same initial sparsity degree of 0.018%.

Stability: Challenge 2 will affect the algorithm effectiveness as mentioned above. To solve this problem, we design corresponding comparative experiments. We choose two data sets, where each data set contains trust data of 28 different sparsity degrees. The experimental results are shown in Fig. 11. The two curves in the figure represent two groups of comparison experiments, and each group contains 28 sparse residuals, which are targeted at data with different sparsity. The results show that the location of sparse nodes affects the effectiveness of the algorithm, but has little influence on the overall trend. It follows that our method is stable.

Updateability: In our scheme, we assume the entity has a self-confidence relationship and the value is equal to 1 ($P_{ii} = 1$). Obviously, the beneficial effect of self-confidence

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

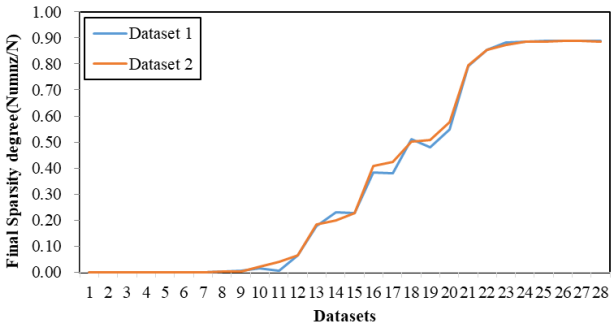


Fig. 11: Residuals of anti-sparsity.

initialization on anti-sparsification cannot be ignored. Moreover, even if the non-zero elements in the matrix reveal the relationship between the interacting entities, the role of the zero elements is not to be underestimated. Zero elements usually represent two cases: no interaction between entities, which means there is no trust, that is, the interaction between entities with a trust value of zero, which indicates no trust. In either case, our approach allows new relationships to be established, so as to update the existing trust without further association relationship mining to provide more information.

5.6 Validity Discussion

Computational complexity. The volume of big data is relatively large. In terms of the complexity of running time, in addition to improving the algorithm, the influence of hardware environment (such as computing soft environment, performance, number of computers, etc.) cannot be ignored. All of these limit the speed of the algorithm. In the future, we will focus on improving the running time complexity of the algorithm by combining the various influencing factors. Furthermore, under the influence of the big data environment, the number of users is increasing, and the load of sparse trust computing will also increase, which may affect the effect of our model. Therefore, it is necessary to consider cross-domain trust modeling and maintenance between different communities to reduce computational complexity.

The non-uniqueness of the evaluation model. In addition, the calculation methods, statistical dimensions and experimental environment used by many existing models are not the same. There is no unified evaluation model system recognized by scholars to compare the differences of different trust models objectively and fairly.

6 CONCLUSION

This paper proposes a sparse trust data mining scheme, and a novel framework for mining and predicting sparse trusts in large-scale recommendation systems. Our scheme alleviates the sparse problem of trust and improves the accuracy of the recommendation system in three aspects. First, we define the representation of sparse trust through analysis of existing sparse trust models. Second, we provide a unified formal description of sparse trust, define the sparsity degree, and declare two main challenges for the big data sparse trust matrix. Third, we propose an efficient and simple anti-sparsification method. It is based on the customized normalization function and the method of the transitive gossip trust model. The former is used to avoid the disadvantages of Grade Inflation and the latter is used to overcome random trust relationship propagation.

In the future, we plan to extend our framework to dealing with mining the hidden semantics to improve the prediction accuracy and effectiveness.

ACKNOWLEDGMENTS

This work is partially sponsored by National Key R&D Program of China (No.2019YFB2101700), National Science Foundation of China (62172297, 61902276), the Key Research and Development Project of Sichuan Province (No.21SYSX0082), Tianjin intelligent manufacturing special fund project(20201159), Natural Science Foundation of Tianjin City grant (No.19JCQNJC00200).

REFERENCES

[1] Z. Zhang, P.-H. Ho, and F. Naït-Abdesselam, "Radar: A reputation-driven anomaly detection system for wireless mesh networks," *Wireless Networks*, vol. 16, no. 8, pp. 2221–2236, 2010.

[2] R. Saini, A. Jindal, and S. De, "Jammer-assisted resource allocation in secure ofdma with untrusted users," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 1055–1070, 2016.

[3] X. Xie, T. Yuan, X. Zhou, and X. Cheng, "Research on trust model in container-based cloud service," *Computers, Materials and Continua*, vol. 56, no. 2, pp. 273–283, 2018.

[4] X. Xie, R. Liu, X. Cheng, X. Hu, and J. Ni, "Trust-driven and pso-sfla based job scheduling algorithm on cloud," *Intelligent Automation & Soft Computing*, vol. 22, no. 4, pp. 561–566, 2016.

[5] J. Zhang and R. Cohen, "Evaluating the trustworthiness of advice about seller agents in e-marketplaces: A personalized approach," *Electronic Commerce Research and Applications*, vol. 7, no. 3, pp. 330–340, 2008.

[6] G. Xu, Z. Feng, H. Wu, and D. Zhao, "Swift trust in a virtual temporary system: A model based on the dempster-shafer theory of belief functions," *International Journal of Electronic Commerce*, vol. 12, no. 1, pp. 93–126, 2007.

[7] P. Gao, H. Miao, J. S. Baras, and J. Golbeck, "Star: Semiring trust inference for trust-aware social recommenders," in *Proceedings of the 10th ACM conference on Recommender systems*, pp. 301–308, 2016.

[8] G. Guo, J. Zhang, D. Thalmann, A. Basu, and N. Yorke-Smith, "From ratings to trust: an empirical study of implicit trust in recommender systems," in *Proceedings of the 29th annual acm symposium on applied computing*, pp. 248–253, 2014.

[9] G. Xu, Y. Cao, Y. Zhang, G. Zhang, X. Li, and Z. Feng, "Trm: computing reputation score by mining reviews," in *AAAI Workshop: Incentives and Trust in Electronic Communities*, 2016.

[10] C. Zhu, H. Nicanfar, V. C. Leung, and L. T. Yang, "An authenticated trust and reputation calculation and management system for cloud and sensor networks integration," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 118–131, 2014.

[11] G. Guo, J. Zhang, and N. Yorke-Smith, "A novel recommendation model regularized with user trust and item ratings," *ieee transactions on knowledge and data engineering*, vol. 28, no. 7, pp. 1607–1620, 2016.

[12] W. Jiang, G. Wang, M. Z. A. Bhuiyan, and J. Wu, "Understanding graph-based trust evaluation in online social networks: Methodologies and challenges," *ACM Computing Surveys (CSUR)*, vol. 49, no. 1, pp. 1–35, 2016.

[13] R. Katarya and Y. Arora, "A survey of recommendation systems in twitter," in *2018 4th International Conference on Computational Intelligence & Communication Technology (CICT)*, pp. 1–5, IEEE, 2018.

[14] R. Katarya, "Reliable recommender system using improved collaborative filtering technique," *System Reliability Management: Solutions and Technologies*, vol. 113, 2018.

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

[15] G. Gupta and R. Katarya, "A study of recommender systems using markov decision process," in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1279–1283, IEEE, 2018.

[16] D. K. Yadav and R. Katarya, "Study on recommender system using fuzzy logic," in *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 50–54, IEEE, 2018.

[17] R. Katarya and O. P. Verma, "A collaborative recommender system enhanced with particle swarm optimization technique," *Multimedia Tools and Applications*, vol. 75, no. 15, pp. 9225–9239, 2016.

[18] G. Gupta and R. Katarya, "Recommendation analysis on item-based and user-based collaborative filtering," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2019.

[19] R. Katarya, "Movie recommender system with metaheuristic artificial bee," *Neural Computing and Applications*, vol. 30, no. 6, pp. 1983–1990, 2018.

[20] R. Katarya, "A systematic review of group recommender systems techniques," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 425–428, IEEE, 2017.

[21] R. Katarya and N. Verma, "Automatically detection and recommendation in collaborative groups," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 218–222, IEEE, 2017.

[22] R. Katarya and O. P. Verma, "Recent developments in affective recommender systems," *Physica A: Statistical Mechanics and its Applications*, vol. 461, pp. 182–190, 2016.

[23] R. Katarya and D. Gautam, "Survey on opinion leader in social network using data mining," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pp. 506–509, IEEE, 2019.

[24] R. Katarya, V. Gangwar, and I. Jaisia, "A study on different data mining classifiers," in *2018 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–6, IEEE, 2018.

[25] R. Katarya and Y. Arora, "Capsmf: a novel product recommender system using deep learning based text analysis model," *Multimedia Tools and Applications*, vol. 79, no. 47, pp. 35927–35948, 2020.

[26] G. Guo, J. Zhang, and N. Yorke-Smith, "Trustsvd: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings.," in *Aaai*, vol. 15, pp. 123–125, 2015.

[27] L. Guo, C. Zhang, and Y. Fang, "A trust-based privacy-preserving friend recommendation scheme for online social networks," *ieee transactions on dependable and secure computing*, vol. 12, no. 4, pp. 413–427, 2014.

[28] X. Liu, J. Liu, S. Zhu, W. Wang, and X. Zhang, "Privacy risk analysis and mitigation of analytics libraries in the android ecosystem," *IEEE Transactions on Mobile Computing*, vol. 19, no. 5, pp. 1184–1199, 2019.

[29] L. Li, J. Liu, L. Cheng, S. Qiu, W. Wang, X. Zhang, and Z. Zhang, "Creditcoin: A privacy-preserving blockchain-based incentive announcement network for communications of smart vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 7, pp. 2204–2220, 2018.

[30] G. Xu, W. Wang, L. Jiao, X. Li, K. Liang, X. Zheng, W. Lian, H. Xian, and H. Gao, "Soprotector: safeguard privacy for native so files in evolving mobile iot applications," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2539–2552, 2019.

[31] W. Yao, J. He, G. Huang, and Y. Zhang, "Modeling dual role preferences for trust-aware recommendation," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 975–978, 2014.

[32] I. Konstas, V. Stathopoulos, and J. M. Jose, "On social networks and collaborative recommendation," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 195–202, 2009.

[33] X. Li, H. Ma, F. Zhou, and W. Yao, "T-broker: A trust-aware service brokering scheme for multiple cloud collaborative services," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1402–1415, 2015.

[34] E. Spertus, M. Sahami, and O. Buyukkokten, "Evaluating similarity measures: a large-scale study in the orkut social network," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 678–684, 2005.

[35] W.-Y. Chen, J.-C. Chu, J. Luan, H. Bai, Y. Wang, and E. Y. Chang, "Collaborative filtering for orkut communities: discovery of user latent behavior," in *Proceedings of the 18th international conference on World wide web*, pp. 681–690, 2009.

[36] S. Wen, M. S. Haghighi, C. Chen, Y. Xiang, W. Zhou, and W. Jia, "A sword with two edges: Propagation studies on both positive and negative information in online social networks," *IEEE Transactions on Computers*, vol. 64, no. 3, pp. 640–653, 2014.

[37] D. Wang, W. Zhou, J. X. Zheng, S. Wen, J. Zhang, and Y. Xiang, "Who spread to whom? inferring online social networks with user features," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2018.

[38] J. O'Donovan and B. Smyth, "Trust in recommender systems," in *Proceedings of the 10th international conference on Intelligent user interfaces*, pp. 167–174, 2005.

[39] X. Li, G. Xu, X. Zheng, K. Liang, E. Panaousis, T. Li, W. Wang, and C. Shen, "Using sparse representation to detect anomalies in complex wsns," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 6, pp. 1–18, 2019.

[40] G. Zhou, G. Xu, J. Hao, S. Chen, J. Xu, and X. Zheng, "Generalized centered 2-d principal component analysis," *IEEE Transactions on Cybernetics*, 2019.

[41] J. Golbeck, "Generating predictive movie recommendations from trust in social networks," in *International Conference on Trust Management*, pp. 93–104, Springer, 2006.

[42] H. Hu, R. Lu, Z. Zhang, and J. Shao, "Replace: A reliable trust-based platoon service recommendation scheme in vanet," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1786–1797, 2016.

[43] H.-N. Dai, H. Wang, G. Xu, J. Wan, and M. Imran, "Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies," *Enterprise Information Systems*, pp. 1–25, 2019.

[44] M. Liu, G. Xu, J. Zhang, R. Shankaran, G. Luo, X. Zheng, and Z. Zhang, "Roundtable gossip algorithm: A novel sparse trust mining method for large-scale recommendation systems," in *International Conference on Algorithms and Architectures for Parallel Processing*, pp. 495–510, Springer, 2018.

[45] D. Zhang, C.-H. Hsu, M. Chen, Q. Chen, N. Xiong, and J. Lloret, "Cold-start recommendation using bi-clustering and fusion for large-scale social recommender systems," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 2, pp. 239–250, 2013.

[46] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in p2p networks," in *Proceedings of the 12th international conference on World Wide Web*, pp. 640–651, 2003.

[47] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative filtering and deep learning based recommendation system for cold start items," *Expert Systems with Applications*, vol. 69, pp. 29–39, 2017.

[48] R. Zacharski, *A Programmer% s Guide to Data Mining: The Ancient Art of the Numerati*. 2015.

[49] G. Ling, M. R. Lyu, and I. King, "Ratings meet reviews, a combined approach to recommend," in *Proceedings of the 8th ACM Conference on Recommender systems*, pp. 105–112, 2014.

[50] D. Zhao, Z. Ma, and D. Zhang, "A distributed and adaptive trust evaluation algorithm for manet," in *Proceedings of the 12th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, pp. 47–54, 2016.

[51] G. Guo, J. Zhang, D. Thalmann, and N. Yorke-Smith, "Etaf: An extended trust antecedents framework for trust prediction," in *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pp. 540–547, IEEE, 2014.

[52] L. Mostarda, C. Dong, and N. Dulay, "Place and time authentication of cultural assets," in *IFIP International Conference on Trust Management*, pp. 279–294, Springer, 2008.

[53] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 287–296, 2011.

[54] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of the fourth ACM conference on Recommender systems*, pp. 135–142, 2010.

[55] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 203–210, 2009.

[56] J. A. Golbeck, *Computing and applying trust in web-based social networks*. PhD thesis, 2005.

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

[57] P. Massa and P. Avesani, "Trust-aware recommender systems," in *Proceedings of the 2007 ACM conference on Recommender systems*, pp. 17–24, 2007.

[58] H. Fang, G. Guo, and J. Zhang, "Multi-faceted trust and distrust prediction for recommender systems," *Decision Support Systems*, vol. 71, pp. 37–47, 2015.

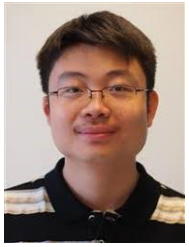
Pengli Nie is a master student at College of Intelligence and Computing, Tianjin University, China. She received her Bachelor degree from Tianjin University of Science & Technology in July 2019. Her research interests include cyber security.

Guangquan Xu is a Ph.D. and full professor at the Tianjin Key Laboratory of Advanced Networking (TANK), College of Intelligence and Computing, Tianjin University, China. He received his Ph.D. degree from Tianjin University in March 2008. His research interests include cyber security and trust management. He is the director of Network Security Joint Lab and the Network Attack & Defense Joint Lab. He has published 100+ papers in reputable international journals and conferences, including IEEE Transactions on Cybernetics, IEEE Internet of Things Journal, ACM Transactions on Internet Technology, ACM Transactions on Intelligent Systems and Technology, IEEE Transactions on Industrial Informatics, Information Sciences, IEEE Wireless Communications, IEEE Network, Computers & Security, and so on. He served as a TPC member for IEEE UIC 2018, SPNCE2019, IEEE UIC2015, IEEE ICECCS 2014 and reviewers for journals such as IEEE access, ACM TIST, JPDC, IEEE TITS, soft computing, FGCS, and Computational Intelligence, and so on. He is a Fellow of IET, IEEE member, senior member of China Computer Society. Contact him at losin@tju.edu.cn.

Litao Jiao received his MBA degree in 2016 from Shandong University of Science and Technology. He is now an associate professor in Qingdao Huanghai University, China. He has been awarded the prize of Provincial Educational Achievement in year 2018, participated in 5 major provincial and municipal research projects, and published more than 10 papers. His research interests include: HR management and information security.

Shaoying Liu [F'18] is a Professor of Software Engineering in the Dependable Systems Lab, Graduate School of Advanced Science and Engineering, School of Informatics and Data Science at Hiroshima University. He holds the Ph.D in Computer Science from the University of Manchester, U.K. His research interests include Software Engineering, Formal Engineering Methods, Software Testing and Verification, and Trustworthy systems. He has developed the SOFL specification language and the related Agile-SOFL method for developing dependable systems. He has authored a book entitled "Formal Engineering for Industrial Software Development" published by Springer in 2004 and over 200 refereed publications in journals and international conferences. He is a IEEE Fellow , BCS Fellow, and a member of IPSJ.

Jian Liu received the B.S. and Ph.D. degrees from the School of Mathematical Sciences at Nankai University, Tianjin, China, in 2009, and 2015, respectively.She was a visiting Ph.D. student at the Department of Mathematics, University of Paris VIII, Paris, France. She is currently an associate professor with the School of Cyber security, College of Intelligence and Computing, Tianjin University, Tianjin, China. Her research interests include cryptography and coding theory.



Weizhi Meng is currently an associate professor in the Cyber Security Section, Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Denmark. He obtained his Ph.D. degree in Computer Science from the City University of Hong Kong (CityU), Hong Kong. Prior to joining DTU, he worked as a research scientist in Infocomm Security (ICS) Department, Institute for Infocomm Research, A*Star, Singapore, and as a senior research associate in CS Department, CityU. He won the Outstanding Academic Performance Award during his doctoral study, and is a recipient of the Hong Kong Institution of Engineers (HKIE) Outstanding Paper Award for Young Engineers/Researchers in both 2014 and 2017. He is also a recipient of Best Paper Award from ISPEC 2018, and Best Student Paper Award from NSS 2016 and Inscrypt 2019. His primary research interests are cyber security and intelligent technology in security, including intrusion detection, smartphone security, biometric authentication, HCI security, trust computing, blockchain in security, and malware analysis. He served as program committee members for 50+ international conferences. He is a senior member of IEEE.

Hongyue Wu is currently an assistant professor at the College of Intelligence and Computing, Tianjin University, China. He received his Ph.D. degree from Zhejiang University, China, in 2018. From 2017 to 2018, he was a joint Ph.D. student at the University of Sydney, Australia. His research interests include service computing, mobile edge computing, and cloud computing. He has published more than 20 papers and won the Best Paper Award at the ICSOC 2017. He is a member of the IEEE.

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2021.3109412, IEEE Transactions on Information Forensics and Security

Meiqi Feng is a master student at the College of Intelligence and Computing, Tianjin University, China. She graduated from Tianjin University with a bachelor's degree in computer science and technology in 2019. Her main research direction is cyberspace security. She is interested in web security and the combination of artificial intelligence and security.

Weizhe Wang is a master student in the College of Intelligence and Computing, Tianjin University. He received his Bachelor's degree from Changchun University of Science and Technology in 2018. His research interests include cyber security.

Zhengjun Jing, Ph.D., associate professor of the School of Computer Engineering, Jiangsu University of Technology, China. His current research interests include public key cryptography scheme design and application, cloud computing and security, data security and privacy protection, and intelligent information systems.

Xi Zheng, received the Ph.D. in Software Engineering from UT Austin in 2015. From 2005 to 2012, he was the Chief Solution Architect for Menulog Australia. He is currently the Director of Intelligent Systems Research Group (ITSEG.ORG), Senior Lecturer (aka Associate Professor US) and Deputy Program Leader in Software Engineering, Macquarie University, Sydney, Australia. His research interests include CPS Verification, Machine Learning Security, Human Vehicle Interaction, Edge Intelligence and Intelligent Software Engineering. He has a number of highly cited papers and serves as PC members for PerCom (CORE A*) and TrustCom (CORE A).

1556-6013 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on September 02,2021 at 05:51:57 UTC from IEEE Xplore. Restrictions apply.