Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm

Shugang Li[a], Xuewei Song[a], Hanyu Lu[a], Linyi Zeng[a,b,*], Miaojing Shi[c], Fang Liu[a]

[a] *School of management, Shanghai University, Shanghai, 200444, PR China*
[b] *Software Development Center, Industrial and Commercial Bank of China, Shanghai, 201206, PR China*
[c] *Univ Rennes, Inria, CNRS, IRISA, 35042, France*

A B S T R A C T

Circle structure of online brand communities allows companies to conduct cross-marketing activities by the influence of friends in different circles and build strong and lasting relationships with customers. However, existing works on the friend recommendation in social network do not consider establishing friendships between users in different circles, which has the problems of network sparsity, neither do they study the adaptive generation of appropriate link prediction algorithms for different circle features. In order to fill the gaps in previous works, the intelligent attention allocation link prediction algorithm is proposed to adaptively build attention allocation index (AAI) according to the sparseness of the network and predict the possible friendships between users in different circles. The AAI reflects the amount of attention allocated to the user pair by their common friend in the triadic closure structure, which is decided by the friend count of the common friend. Specifically, for the purpose of overcoming the problem of network sparsity, the AAIs of both the direct common friends and indirect ones are developed. Next, the decision tree (DT) method is constructed to adaptively select the suitable AAIs for the circle structure based on the density of common friends and the dispersion level of common friends' attention. In addition, for the sake of further improving the accuracy of the selected AAI, its complementary AAIs are identified with support vector machine model according to their similarity in value, direction, and ranking. Finally, the mutually complementary indices are combined into a composite one to comprehensively portray the attention distribution of common friends of users in different circles and predict their possible friendships for cross-marketing activities. Experimental results on Twitter and Google+ show that the model has highly reliable prediction performance.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

With the rapid development of the Internet, the popularity of high-speed, stable Internet services and the increasing experiences of online shopping, 80% of the top 500 companies in the world have established online brand communities. The online brand community brings together the scattered target customers of the company accurately and has become a new platform for company to carry out marketing activities as well as build strong and lasting relationships with customers (John, Mochon, Emrich, & Schwartz, 2017).

A large number of works have already shown that friend groups can affect individual consumer decisions, for example, product evaluation, purchase possibility (Whittler & Spira, 2002), and actual purchase behavior (Li, Chou, & Lin, 2014). Moreover, when brand preference conflicts between group and individuals, individuals may hide their consumption behavior (Thomas, Jewell, & Jennifer, 2015). Some scholars in consumer behavior studies (Solomon, 2016) pointed out that the lower the user requirements for product, the higher the influence of the reference group. Aral (2013) studied the case of 1.4 million Facebook users downloading a movie application to find out the factors that affected their decision making. He randomly divided users who had downloaded the application into three groups: The first group had the right to invite their friends to try it on their own; friends of the users in the second group received an automatically generated message indicating that their friends were using the application;

friends of the users in the third group did not receive any messages. As a result, 6% of the friends who received the unsolicited invitation downloaded the application, compared with 2% of those who received automatic tips. In addition, he also compared users who actively sent invitations and successfully invited friends with users who sent automatic tips to friends and invited friends. In the long run, the former used the application for more time than the latter.

Friends in online brand communities share their passion for a specific brand as well as exchange information and knowledge, and these social interactions positively influence the member loyalty to the brand (Brogi, 2014). In addition, some members who share a common interest or prefer the same product are further gathered together to form different sub-groups, i.e. circles (Wang & Xue, 2010) in a brand community. Circle structure provides fertile ground for cross-marketing, where sales are amplified with the influence of friends though forming friendships between users in different circles.

Scoring link prediction algorithm (SLPA) is the main approach to predict whether there are links between node pairs in the social network and recommend friendships between user nodes based on network topology. So far, various SLPAs have been widely applied for the recommendation of friends, such as the Common Neighbor index (CN), the Hub Promoted Index (HPI), and the Resource Allocation (RA) (Lü & Zhou, 2011). However, existing works on prediction friendships in social network do not consider establishing friendships between users in different circles, which bears on the problem of network sparsity. What's worse, there is not yet an SLPA suitable for all network circle structures. So, constructing the suitable SLPA based on circle structure characteristics is of significant theoretical and practical value, which is a complex work that requires a lot of expert experience. Therefore, it is of great significance to develop method that relies solely on network data to build appropriate SLPA, so that inexperienced practitioners can easily use it to obtain highly reliable prediction results.

To address these problems, this study proposes the intelligent attention allocation link prediction algorithm (IAALP A) to predict the possible friendships between users in different circles by adaptively building attention allocation index (AAI) for specific circles, and the AAI represents the amount of attention assigned to the user pair by their common friend in the triadic closure structure, which is determined by the friend count of the common friend.

Specifically, for the purpose of overcoming the network sparsity problem in predicting friendships between users in different circles, the AAIs of both the direct common friends and indirect ones are developed based on the principle of attention allocation of common friends in triadic closure structure. Since it's difficult to build the AAI that fits all the structural characteristics of circles, in light of the characteristics of the density of common friends and the dispersion level of common friends' attention, the decision tree (DT) method (Wang, Wu, & Yao, 2017) is developed to adaptively select the AAI suitable for predicting possible links between user nodes in different circles. In addition, although the combination of single AAI can help improve its link prediction performance, the blind combination of AAIs cannot bring the expected results (Gomes, Barddal, Enembreck, & Bifet, 2017). To this end, IAALPA applies the support vector machine (SVM) (Shan, Kong, Zhang, Li et al., 2018) model to identify the complementary AAIs of the AAI selected by DT, which can improve the performance of the selected AAI by combining them as a composite indicator. Finally, the selected AAI and its complementary ones are used to design the composite mutually complementary AAI to comprehensively portray the attention allocated to user node pairs by their common friends and forecast the possible connections between user

nodes in different circles. Consequently, the friendships between users in different circles are recommended and successfully cross-marketing is achieved. Specifically, a group of users of one product circle are recommended to the target customers who are the users of another circle in online brand community. Accordingly, when the brand marketers are supported by the friend group's influence on individuals, they can significantly enhance the cross-marketing efficiency of the product.

The remainder of this paper is organized as follows: Section 2 introduces the link prediction; Section 3 offers friend group recommendation; Section 4 explains the IAALPA; Section 5 offers the experimental design and the results analysis; Section 6 gives a summary of this study.

2. Link prediction

Link prediction is a network-related problem, which consists of predicting new connections and detecting hidden links in a network. It is an important task applicable to a wide variety of areas, such as bibliographic domain, molecule biology, criminal investigations and recommending systems (Lü & Zhou, 2011; Martínez, Berzal, & Cubero, 2017; Xiang, 2008). The link prediction problem can be formally defined as follows. Given a snapshot of a social network at time t, we seek to accurately predict the edges that will be added to the network at time $t + \Delta$ by defining a similarity or a probability index (Liben-Nowell & Kleinberg, 2007; Martínez et al., 2017). The existing link prediction methods can be divided into four categories: Link prediction algorithms based on similarity measures, probabilistic and statistical methods, algorithmic approaches and preprocessing methods (Martínez et al., 2017). Among these approaches to treat the problem, the most widespread ones rely on the use of similarity measures between node pairs (Lü & Zhou, 2011; Martínez et al., 2017).

The similarity measures proposed and evaluated in previous literature can be broadly categorized into two groups: semantic and topological measures (Kaya & Poyraz, 2016). Semantic measures use the nodes' content to survey similarity. For instance, in a co-authorship network, the similarity between keywords extracted from published papers were applied to predict future interaction among the authors (Xiang, 2008). Different from the semantic measures, the topological measures consist of deploying the network structure to compute the similarity scores (e.g. the number of common neighbors that two nodes share). Topological measures are more commonly adopted in the literature since they are more general and do not require the definition of rich features to describe content. In fact, rich features are not always available and depend on the social network considered.

Several topological measures are proposed in existing literature and mainly categorized into neighborhood-based and path-based measures (Kaya & Poyraz, 2016). The neighborhood-based measures take the nodes' immediate neighbors into account. In general, these measures consider that two nodes are more likely to form a link if their sets of neighbors have a large overlap (Xiang, 2008). Among the neighborhood-based measures, Salton (Newman, 2001), Sorenson, HPI, Hub Depressed Index (HDI), Leicht–Holme–Newman (LHN), Preferential Attachment(PA) (Barabâsi et al., 2002), RA (Adamic & Adar, 2003) and Jaccard's coefficient (Martínez et al., 2017) can be mentioned. The path-based measures in turn define the similarity between nodes by considering the paths between them. The basic idea is that two nodes are more likely to form a link if there are more short paths between them. The path-based measures range from the ordinary path-distance measures to more sophisticated measures that consider ensembles of different paths, for instance, the Katz measure (Soares & Prudêncio, 2013). In comparative terms, the neighborhood-based methods are more widespread, due to both their computational efficiency and great

performance observed in experiments (Huang, 2006; Liben-Nowell & Kleinberg, 2007; Murata & Moriyasu, 2008). The measures proposed in this study can be categorized as neighborhood-based ones, since they use information about the connections around nodes to assign scores to them. But the uniqueness of our algorithm is that it considers not only the direct common neighbors of node pairs but also their indirect common neighbors. Moreover, the feature of attention allocation of common neighbors, namely the friend count of common neighbors, is taken into account.

Since a SLPA is not sufficient to fully characterize the network, the combined link prediction algorithms are developed. On the whole, there are many combined link prediction algorithms which are built with a high degree of versatility to suit a variety of networks based on multiple network characteristics. For example, Fan, Liu, Lu, Xiu, and Chen (2017) proposed a composite link forecasting index and took into account the node type effect and node structure similarity. Wu and Tang (2014) developed a directed social network link prediction method based on the topic model that integrated node attributes and network structures for link prediction. Muniz, Goldschmidt, and Choren (2018) used context (node and link attributes), temporal information (chronological interaction data) and topology information to calculate link weights between nodes, and then applied weighted similarity function to identify potential links. Xiao, Li, Wang, Xu, and Liu (2018) studied the internal and external factors that influenced the formation of links and developed a three-level hidden Bayesian link prediction model by combining the user behaviors and user relationships to link prediction.

Unlike existing works that focus on algorithm versatility, this study tends to build the suitable algorithms for specific circle structures. Obviously, network structures are diverse and a SLPA cannot perform well in all networks. Accordingly, this study proposes IAALPA to adaptively select AAIs suitable for the given circle structure according to the network features of common neighbor density and attention dispersing. Additionally, the complementary AAIs of the selected AAI are screen out based on their similarity in value, direction, and ranking. Subsequently, the selected AAI and its complementary ones are combined into one composite index to avoid the low prediction performance caused by the blind combination of AAIs.

In recent years, many scholars have done valuable works on friend recommendation in social network based on link prediction methods. Using the recommendation algorithm of data mining, Liu, Yu, Wei, and Ning (2018) proposed an improved algorithm to rank the recommended information with confidence interval, and recommended friends with the same interest for users in microblog. Yuan, Cheng, Zhang, Liu, and Lu (2015) designed a social influence propagation method to mine user's buddy (friends who had a great impact on user) and susceptibility (willingness to be affected), and developed a recommendation model based on the impact of social relations. He et al. (2017) integrated link and content information and developed the MapReduce distributed computing framework to implement the recommendation of friends in large-scale online community network. Zhu, Lu, and Ma (2015) mined user interests from short messages and proposed the neighbor-based friend recommendation to recommend users with similar interests.

Despite intensive research efforts, there is a distinct lack of methodology for recommending friends in different social circles in a social network, which is different from the traditional friend recommendation due to the problem of network sparseness. In this study, IAALPA is constructed to recommend the user group in one product circle to the target user in another product circle, and the influence of the friend group is used to affect the purchase decision of the target user, so as to realize cross-selling.

3. Friend group recommendation for cross-marketing

3.1. Cross-marketing in a brand community

The online brand community is defined as $D(V, E)$, where V is the set of nodes representing the users in the community and E is the set of edges representing the friend relationships between users. In a brand community, users will form a circle because they prefer the same type of product. The set of users in the circle of product A is defined as V_A and the set of users in the circle of product B is defined as V_B. In cross-marketing, if product A is sold to users in the circle of product B, all users in the product A circle can be recommended to users in the product B circle as the friend group, and vice versa.

Usually, the nodes in the same circle are closely connected (there are more common friends), while the nodes in different circles are sparsely connected (there are fewer common friends). Therefore, recommending users in different circles to become friends often faces the problem of network sparsity, which is characterized by the average degree of network being far less than the number of nodes (Lei & Rinaldo, 2015). Since the topological characteristics of social networks are all related to average degree, either directly or indirectly, sparsity will affect the performance of the existing SLPAs because these algorithms heavily rely on the network topology. Consequently, traditional SLPAs cannot guarantee high prediction accuracy of friend recommendation in different product circles. To address this problem, IAALPA is developed to fully describe the possibility of connection between user nodes in different circles and construct a suitable mutually complementary AAI for the specific circle structure.

Fig. 1 shows an example of friend group recommendation in different circles in brand community D, where the initial network is shown in Fig. 1(a), the network with predicted links is described in Fig. 1(b). In Fig. 1(a), users that belong to product A circle are 1, 2, 3, and 4, users that belong to product B circle are 5,6, and 7. Suppose that user 7 is the target customer, the purpose of friend group recommendation is to enable him to purchase product A. Through the link prediction method, it is found that user 7 may establish a friend relationship with users 1, 3, and 4 in product A circle. Then the online brand managers can recommend users 1, 3, and 4 to become friends with user 7, as shown in Fig. 1(b). When they become friends, the target customer 7 is encouraged to buy product A using the impact of the friend group of 1, 3, and 4.

3.2. AAI for friend group recommendation

In the triadic closure structure, the fewer friends a common neighbor of a node pair has, that is, the smaller its degree is, the more attention the common neighbor assigns to the node pair (Backstrom, Bakshy, Kleinberg, Lento, & Rosenn, 2011), accordingly, the more likely there is a link between the node pair. Based on this principle, AAIs are constructed from the point of view of microstructure (node pairs and their common neighbors), and macrostructure (node pairs, their common neighbors, and friends of common neighbors), as a result, the problem of network sparsity is overcome and the attention allocation between node pairs is comprehensively depicted.

3.2.1. AAIs based on microstructure

Seven most commonly used indices, such as Salton, Sorenson, HPI, HDI, LHN, PA, RA, and resource allocation average (RAA) are selected (Lü & Zhou, 2011) and adopted as microstructure based AAIs, which consider the microstructure consisting of node pairs and their common neighbors. As shown in Table 1, the larger number of common neighbors or the less degree of common neighbor indicates the larger scores of attention allocation, vice versa. In
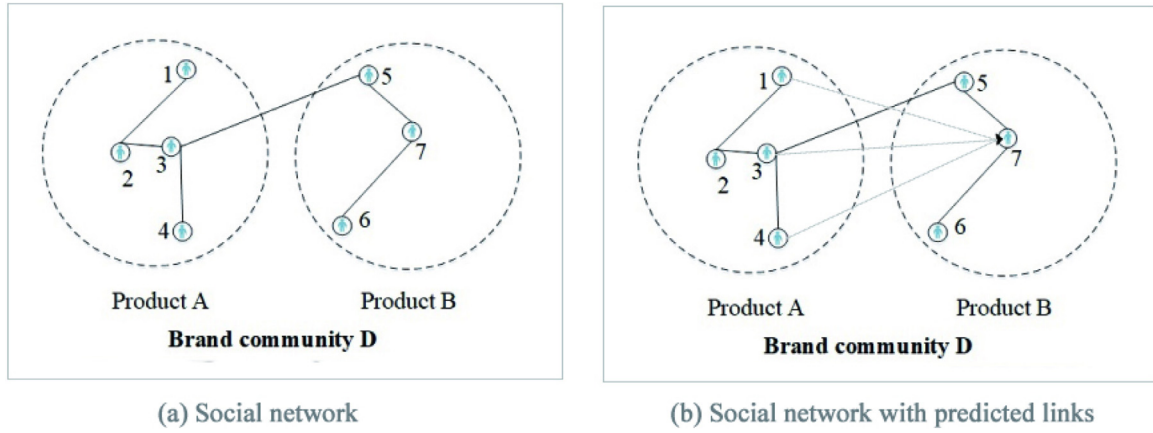
(a) Social network   (b) Social network with predicted links

Fig. 1. An example of friend group recommendation.

Table 1
AAIs based on microstructure.

AAI	Formula
Salton	$S_{xy}^{Salton} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{k(x) * k(y)}}$
Sorenson	$S_{xy}^{Sorenson} = \frac{2|\Gamma(x) \cap \Gamma(y)|}{k(x) + k(y)}$
HPI	$S_{xy}^{HPI} = \frac{|\Gamma(x) \cap \Gamma(y)|}{min\{k(x), k(y)\}}$
HDI	$S_{xy}^{HDI} = \frac{|\Gamma(x) \cap \Gamma(y)|}{max\{k(x).k(y)\}}$
LHN	$S_{xy}^{LHN} = \frac{|\Gamma(x) \cap \Gamma(y)|}{k(x) * k(y)}$
PA	$S_{xy}^{PA} = k(x) * k(y)$
RA	$S_{xy}^{RA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z)}$
RAA	$S_{xy}^{RAA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{max(k(x).k(y))}{k(z)}$

Table 1, $\Gamma(\cdot)$ represents the neighbor set of a node, $k(\cdot)$ denotes the degree of a node.

3.2.2. AAIs based on macrostructure

Considering the sparseness of local networks formed by nodes in different circles, we innovatively develop a wider range of AAIs based on the macrostructure of network, which consider the effect of indirect common neighbors, i.e., friends of common neighbors. Specifically, the attention allocated to common neighbors by their friends can change the connections of common neighbors and then indirectly affect the attention that is allocated to target node pairs by common neighbors. These innovatively proposed AAIs are WA1, WA2, WA3, and RWA.

(a) WA1

WA1 represents the attention allocated to node pairs by the indirect common neighbors. The larger degree of indirect common neighbor indicates a smaller score of attention allocation, as shown in formula (1).

$$S_{xy}^{WA1} = \sum_{\Gamma(z), z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(\Gamma(z))} \tag{1}$$

where $k(\Gamma(z))$ represents the degree of friend of common neighbor z between node x and y.

(b) WA2

WA2 represents the attention allocated to node pairs by direct and indirect common neighbors, where the larger clustering coefficient represents the less attention allocation of indirect common neighbors and the larger degree of direct common neighbors indicates the less attention allocation, as shown in formula (2).

$$S_{xy}^{WA2} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z) * (\rho * c(z) + \varphi * (1 - c(z)))} \tag{2}$$

where $c(z)$ is the clustering coefficient of node z, namely $c(z) = 2 * n/(k * (k - 1))$, n represents the number of links between all k neighbors of node z, and ρ and φ are constant parameters.

(c) WA3

Obviously, the larger degree of node indicates the wider attention dispersion of the node. In WA3, the attention dispersion of the nodes themselves is combined with the attention dispersion of the common neighbor nodes, as shown in formula (3).

$$S_{xy}^{WA3} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \left(\frac{1}{k(z) * (1 - c(z))} + \frac{1}{k(x) * c(x)} + \frac{1}{k(y) * c(y)} \right) \tag{3}$$

(d) RWA

In fact, the connection of node pairs is often affected by the joint influence of the attention allocation of direct and indirect common neighbors as well as the nodes contained in the node pairs. Accordingly, a combined index RWA is obtained by integrating AAIs based on microstructure and macrostructure so that the characteristics of attention distribution among nodes are adequately depicted and the shortcomings of low prediction accuracy caused by circle structure sparsity are overcome. Among AAIs based on microstructure, since RA is highly effective and widely used, it is adopted in RWA. In AAIs based on macrostructure, because WA1 is included in $c(z)$, we just consider WA2 and WA3 in RWA. Consequently, RWA is constructed, as shown in formula (4).

$$S_{xy}^{RWA} = \alpha * S_{xy}^{RA} + \beta * S_{xy}^{WA2} + \gamma * S_{xy}^{WA3} \tag{4}$$

where α, β, and γ are the weight parameters of S_{xy}^{RA}, S_{xy}^{WA2}, S_{xy}^{WA3}, respectively.

4. IAALPA

IAALPA considers the relationship between the joint effect of various features of the network and the prediction performance of the algorithm. In IAALPA, DT model is developed to select the appropriate AAI based on the network characteristics related to common neighbors. Moreover, because a single AAI often makes overestimation or underestimation and the random combination of AAIs does not guarantee excellent results every time, SVM is constructed to recognize the complementary AAIs to generate an effective combined prediction model. Fig. 2 shows the structure of IAALPA.
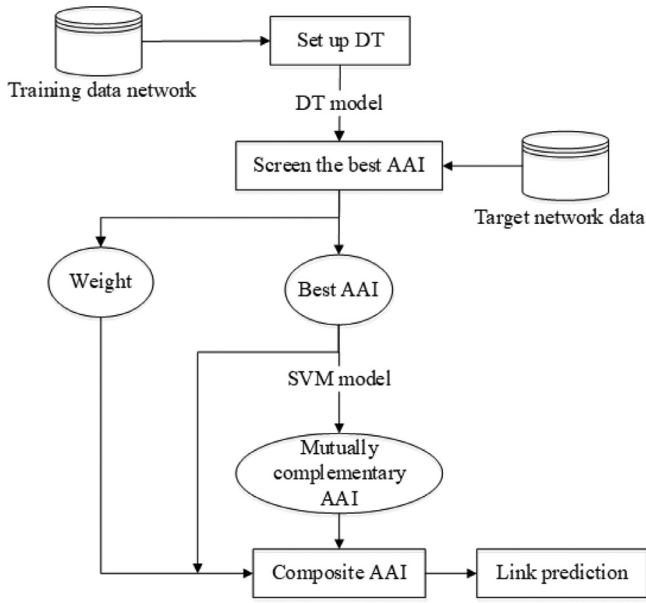
Fig. 2. The structure of IAALPA.

4.1. Algorithm evaluation and complementary AAI

The area under the curve (AUC) is the most common standard metric for measuring the accuracy of SLPAs. AUC randomly selects the connected node pairs and unconnected ones in the test set, and compares their score values obtained by AAI. In m independent comparisons, if the connected node pairs have a higher score of $m1$ times, then the AUC is shown in formula (5) (Liben-Nowell & Kleinberg, 2007).

$$AUC = \frac{m1 + 0.5(m - m1)}{m} \tag{5}$$

When the network size is large, the AUC value obtained by this random sampling method can reduce the computation complexity and improve the efficiency. It is obvious that the larger the AUC value, the higher the accuracy of the algorithm.

The complementary AAI of any index B is: If the combination of B and candidate AAI has better AUC than that B has, then the candidate AAI is considered to be the complementary index of B.

4.2. DT for screening AAI

Although AUC is the most common method for evaluating AAI, the AUC method does not analyze the characteristics of the specific circle structure. In practice, it is often necessary to try all AAIs before finding the AAI that is most suitable for a given network. Since DTs are good at solving multi-class problems and can implicitly perform variables screening or features selection while requiring relatively little effort for data preparation (Mantas, Abellán, & Castellano, 2016), DT is developed to adaptively select the appropriate AAIs for the specific circle structure.

The features of the network according to the idea of attention allocation of common neighbors in the triadic closure structure are fully described in the following two dimensions. First of all, the density of common neighbors is considered since more common neighbors indicate more attention allocated, such as the average degree and the average clustering coefficient. Secondly, the dispersion level of common neighbors' attention is employed. Because more short paths connecting common neighbors mean their more attention dispersion, we consider the short path related features, for instance the average shortest path, the average node betweenness, and the average link betweenness.

In DT, the independent variables of the learning sample are the network characteristic indices, and the dependent variable is the AAI with the maximum AUC value. Fig. 3 shows an example of DT. In this study, the DT algorithm is designed based on the idea of C4.5 algorithm (Mantas et al., 2016), because C4.5 classification tree is the most popular algorithm and has been proved by many studies to be the simple and practical learning algorithm. Assuming that the network training set is T, in each sample, the network is labeled by the AAI with the largest AUC. Given that there are k types of AAIs, a division of T is obtained as $\{S_1, S_2, \cdots, S_k\}$. The prior probability of division is $P_i = |S_i|/|T|$. Then the information entropy used for classifying T is $Info(T) = -\sum_{i=1}^{k} P_i log_2 P_i$.

The networks in T are divided by feature A (such as the average clustering coefficient), and the sequence $\{A_1, A_2, \cdots, A_J\}$ is obtained by arranging the values of feature A in ascending order. Defining the ith $(1 \leq i \leq J - 1)$ partition point as $a_i = (A_i + A_{(i+1)})/2$, T is divided into 2 subsets $\{T_1, T_2\}$, where the value of feature A of the networks in T_1 is $V(A, T_1) \in [A_1, a_i]$, and similarly $V(A, T_2) \in (a_i, A_J]$. Corresponding to this kind of division, the information gain of feature A is $Gain(A) = Info(T) - Info_A(T)$, where $Info_A(T) = \sum_{i=1}^{2} \frac{|T_i|}{|T|} Info(T_i)$. Corresponding to partition point a_i, the information gain rate of A is shown in formula (6).

$$Gain\_Ratio(A, a_i) = \frac{Gain(A)}{Split(A)} \tag{6}$$

where $Split(A) = -\sum_{i=1}^{2} \frac{|T_i|}{|T|} log_2 \frac{|T_i|}{|T|}$.

Subsequently, the information gain rate for each partition point in sequence $\{A_1, A_2, \cdots, A_J\}$ is calculated according to formula (6), and the partition point with the maximum gain rate is selected as the best branch threshold of the feature A, namely, $Threshold(A) = \max_{1 \leq i \leq J-1} \{Gain\_Ratio(A, a_i)\}$.

The main procedure of the DT proposed in this study is shown as follows:

Step1. The features of common friend density and the dispersion level of common friends' attention are calculated, the feature with the maximum information gain rate is selected as the root node, and it is branched based on its best branch threshold;

Step 2. For the subset of data corresponding to the branch of different features, branches of the tree are recursively established by the same method as Step1, and this procedure is repeated until all data samples of each branch belong to the same class AAI;

Step 3. The simplified DT is obtained by pruning the initial DT to eliminate the influence of random factors such as noise and isolated nodes;

Step 4. The decision rules are extracted. For the DT generated by Step 3, the decision rules can be obtained directly, that is, the best AAI suitable for the circle structure is selected based on the structural features of the community network.

It should be noted that in the process of training the DT model, the AUC of AAI for each network in training set is calculated, and the average AUC value w of each AAI will be used as its weight in the combined model in Section 4.4.

4.3. Selecting complementary AAIs based on SVM

Although it is possible to select the most suitable AAI for the network through DT method, the combination of the selected AAI and its complementary indices can fully capture the attention allocation characteristics of nodes in different circles. The sparsity of the network circle is sufficiently diverse, and that may lead to the overfitting of the model to identify the mutually complementary AAIs, namely the model adapts too exactly to the particular circles, and fails to fit additional circles reliably. SVM shows many unique advantages in avoiding overfitting and solving small sample, non-linear and high-dimensional pattern recognition problems, so it is
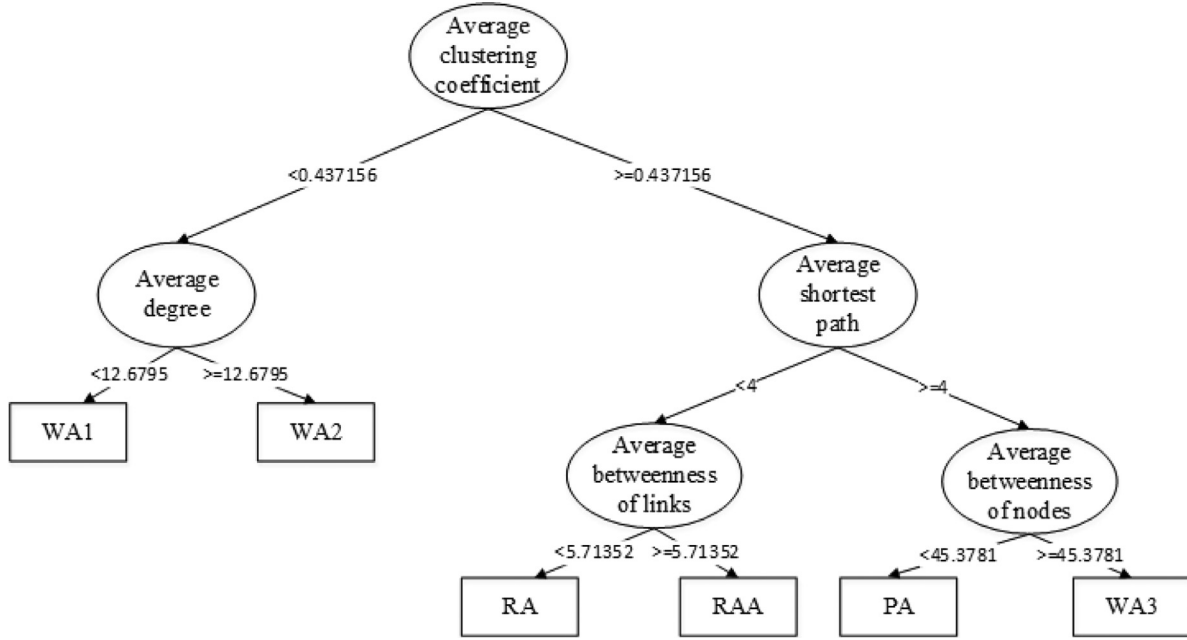
Fig. 3. An example of DT.

adopted to identify AAIs that are complementary to the best AAI selected by the DT model (Wang & Xing, 2019). Subsequently, the mutually complementary indices are combined into a composite one.

In order to fully describe complementary relationship between indices, the score similarity between AAIs is described from three aspects: firstly, the distance between the scores of the two indices, such as Euclidean distance, standardized Euclidean distance, Manhattan distance, and Chebyshev distance; secondly, the similarity in direction, such as Cosine distance and Pearson correlation coefficient; thirdly, the difference in performance ranking, for example, Spearman distance.

Given a set of training samples $[I_i, M_i]$, $i = 1, 2, \cdots, l$, input vector I_i represents the above 7 similarity indices for the scores of two AAIs in sample i, and $M_i = 1$ if two AAIs are mutually complementary, else $M_i = -1$. The main idea of SVM is to map the input vector to a high dimensional eigenvector space, and to construct the optimal classification surface in the eigenvector space. In order to improve the efficiency of the algorithm, Radial Basis Function (RBF) $K(I_i, V) = exp(-||V - I_i||^2/\delta^2)$ is selected as the kernel function, where V is the input vector and δ is constant. Accordingly, an optimal classification function can be obtained in formula (7).

$$f(V) = sgn\{(W \cdot \Phi(V)) + b\} = sgn\left(\sum_{i=1}^{l} a_i M_i K(I_i, V) + b\right) \quad (7)$$

where a and b are constants, W is a normal vector, which determines the direction of hyperplane, b is a displacement term, which indicates the distance between hyperplane and origin. $\Phi(V)$ represents the eigenvector after mapping V.

4.4. Composite AAI

In order to comprehensively portray the attention distribution of common friends of users in different circles, IAALPA designs a composite mutually complementary AAI. The composite index model is mainly composed of the best index B selected by DT and the complementary AAIs of B identified by SVM (i.e. $E_1, E_2, \cdots, E_i$). Aiming at making the AAI with good performance account for a large proportion in the composite AAI, average AUC value w of

each AAI, which is obtained in training DT model, is applied as its weight, as mentioned in Section 4.2. In addition, h excellent complementary AAIs with larger weights are selected to build the composite AAI to further improve the accuracy of the algorithm, as shown in formula (8).

$$S = w * (B, E_1, E_2, \cdots, E_h)^T \quad (8)$$

5. Experiments and results analysis

5.1. Experimental design

971 Ego-net data sets in Twitter (http://snap.stanford.edu/data/egonets-TwittQQer.html) and 132 Ego-net data sets in Google+ (http://snap.stanford.edu/data/ego-Google+.html) were adopted to verify the effectiveness of IAALPA. Each Ego-net is a brand community network, where center node represents the brand company and users are divided into different product circles. 760 networks with product circles were selected from Twitter, and 100 networks with product circles were selected from Google+. Table 2 shows the mean, minimum and maximum of the statistical indicators of the selected network samples in Twitter and Google+.

To test the accuracy of the proposed algorithm, the networks in Twitter and Google+ were randomly divided into two parts, respectively: the training set, which contained 80 percent of networks, and the test set, which contained the remaining 20 percent of networks.

To describe briefly, S1-S25 are used to represent 25 algorithms, as shown in Tables 3 and 4. Table 3 gives AAIs without parameters, and Table 4 gives AAIs with specific parameter values. These values were determined according to a large number of experiment results and those values chosen for the parameters have been proven to work correctly on a wide range of problems.

In the DT experiments, 25 AAIs were screened based on training set data, and 15 optimal AAIs were selected, which were S6, S7, S8, S9, S10, S11, S13, S15, S16, S17, S19, S20, S23, S24, and S25. For each optimal AAI, train set was used to train SVM, and then its complementary AAIs were found, so that 15 SVMs were obtained corresponding to 15 AAIs. Then, the composite mutually complementary AAI was designed based on formula (8).

Table 2
Statistical information of network samples in Twitter and Google+.

Statistical indicators	Twitter			Google+		
	Minimum	Mean	Maximum	Minimum	Mean	Maximum
Number of nodes	4	40.35	230	4	222.11	962
Number of edges	4	398.06	5137	8	1.23	56330
Average degree	1.66	13.55	64.50	2	37.89	143.11
Average shortest path	0.44	1.64	3.65	0.34	1.81	3.20
Average node betweenness	1	33.38	330.04	1	205.73	1.15
Average link betweenness	0.44	1.64	3.65	0.34	1.81	3.20
Average clustering coefficient	0.12	0.66	1	0.24	0.66	0.87

Table 3
AAIs without parameters.

AAI abbreviation	S1	S2	S3	S4	S5	S6	S7	S8	S12	S13
AAI name	Salton	Sorenson	HPI	HDI	LHN	PA	RA	WA1	RAA	WA3

Table 4
AAIs with parameters.

Algorithm	AAI	Parameter	Algorithm	AAI	Parameter
S9	WA2	$\rho = 1, \phi = 4$	S19	RWA	$\rho = 1, \phi = 1, \alpha = 4, \beta = 1, \gamma = 1$
S10	WA2	$\rho = 1, \phi = 1$	S20	RWA	$\rho = 1, \phi = 1, \alpha = 1, \beta = 4, \gamma = 1$
S11	WA2	$\rho = 4, \phi = 1$	S21	RWA	$\rho = 1, \phi = 1, \alpha = 1, \beta = 1, \gamma = 4$
S14	RWA	$\rho = 1, \phi = 4, \alpha = 1, \beta = 1, \gamma = 1$	S22	RWA	$\rho = 4, \phi = 1, \alpha = 1, \beta = 1, \gamma = 1$
S15	RWA	$\rho = 1, \phi = 4, \alpha = 4, \beta = 1, \gamma = 1$	S23	RWA	$\rho = 4, \phi = 1, \alpha = 4, \beta = 1, \gamma = 1$
S16	RWA	$\rho = 1, \phi = 4, \alpha = 1, \beta = 4, \gamma = 1$	S24	RWA	$\rho = 4, \phi = 1, \alpha = 1, \beta = 4, \gamma = 1$
S17	RWA	$\rho = 1, \phi = 4, \alpha = 1, \beta = 1, \gamma = 4$	S25	RWA	$\rho = 4, \phi = 1, \alpha = 1, \beta = 1, \gamma = 4$
S18	RWA	$\rho = 1, \phi = 1, \alpha = 1, \beta = 1, \gamma = 1$			

Table 5
The performance of all algorithms in the Twitter experiments.

Algorithm	Average AUC	Algorithm	Average AUC	Algorithm	Average AUC
S1	0.78543	S10	0.819112	S19	0.81563
S2	0.78078	S11	0.815698	S20	0.815624
S3	0.737871	S12	0.822598	S21	0.800107
S4	0.767094	S13	0.793013	S22	0.80673
S5	0.622253	S14	0.806359	S23	0.814921
S6	0.775256	S15	0.814763	S24	0.811521
S7	0.819129	S16	0.810858	S25	0.798743
S8	0.799544	S17	0.798534	IAALPAa	0.822624
S9	0.820377	S18	0.809194	IAALPAb	0.8158
Single DT	0.811517			SVM	0.656711

Table 6
The performance of all algorithms in the Google+ experiments.

Algorithm	Average AUC	Algorithm	Average AUC	Algorithm	Average AUC
S1	0.66928	S10	0.84928	S19	0.841492
S2	0.644073	S11	0.845608	S20	0.841461
S3	0.727259	S12	0.882082	S21	0.813375
S4	0.622012	S13	0.803694	S22	0.825926
S5	0.387874	S14	0.824811	S23	0.840444
S6	0.873019	S15	0.840224	S24	0.835089
S7	0.849222	S16	0.832744	S25	0.811044
S8	0.884159	S17	0.810624	IAALPAa	0.892791
S9	0.85096	S18	0.830051	IAALPAb	0.87925
Single DT	0.857105			SVM	0.688149

Since quite a few composite link prediction algorithms were constructed based on SVM, another composite algorithm was developed with SVM to evaluate whether IAALPA outperformed the recently developed other composite algorithms. In the SVM based composite algorithm, the links between node pairs were predicted using 8 AAIs in Table 1. All algorithms were applied in MATLAB software with default settings.

The average AUCs for Twitter and Google+ dataset in 100 random experiments are shown in Tables 5 and 6, respectively, where IAALPAa represents the combination of the optimal AAI selected by the DT model and two complementary AAIs with the largest weights. IAALPAb represents the combination of the optimal AAI selected by the DT model and its all complementary AAIs.

Figs. 4 and 5 show performance comparison of various AAIs in the Twitter and Google+ experiments, respectively. Figs. 6 and 7 show the performance comparison of single DT and non-combination AAIs in the Twitter and Google+ experiments, respectively.
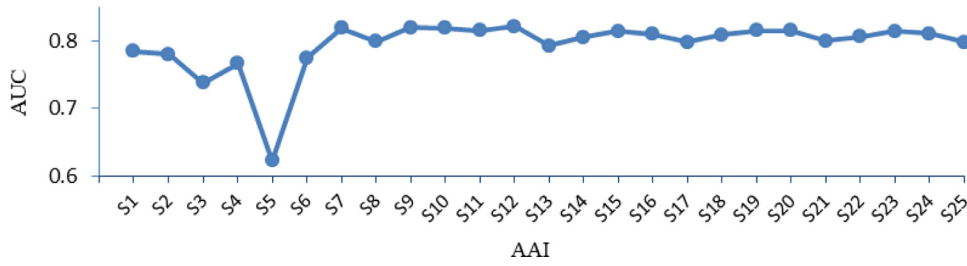
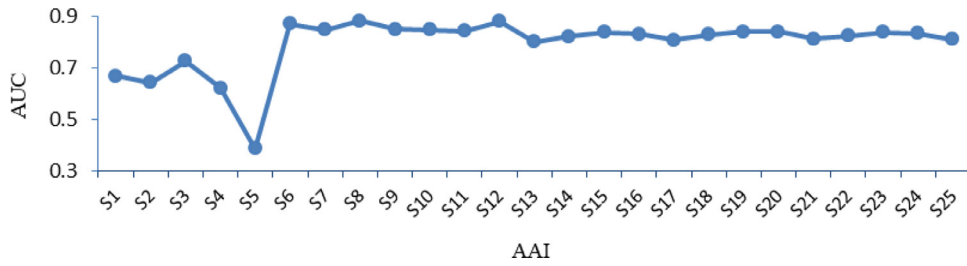**Fig. 4.** Performance comparison of AAIs in the Twitter experiments.

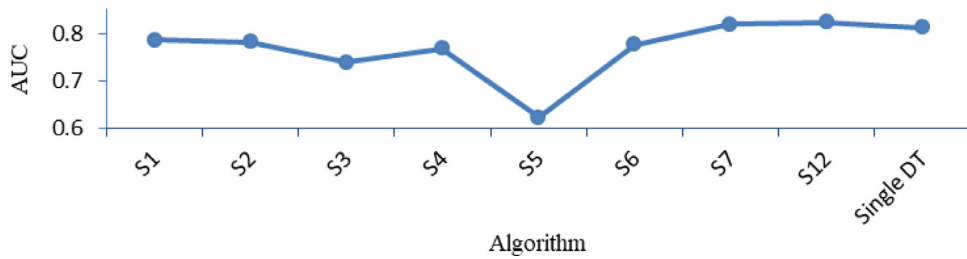**Fig. 5.** Performance comparison of AAIs in the Google+ experiments.

Fig. 6. Performance comparison of single DT and non-combination AAIs in the Twitter experiments.
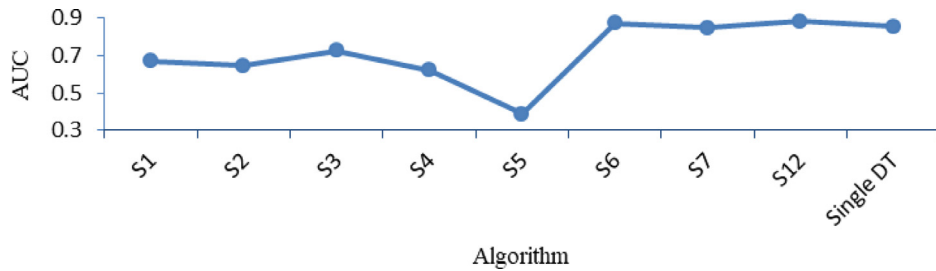

Fig. 7. Performance comparison of single DT and non-combination AAIs in the Google+ experiments.

5.2. Performance analysis of algorithms

Tables 5 and 6 show that compared with existing outstanding algorithms, IAALPA has a significant improvement effect, that is, IAALPA proposed in this study can accurately recommend friendships between users in different circles when conducting cross-marketing activities in online community. It can also be observed from Tables 5 and 6 that performance of IAALPAa is better than that of IAALPAb integrating all complementary AAIs, and this indicates that combining complementary AAIs with larger weights is an effective means to improve the accuracy of IAALPA.

From Figs. 4 and 5, it can be seen that the accuracy of the 19 algorithms, ranging from S7 to S25, is significantly higher than that of the other 7 AAIs, and these optimal AAIs are all new proposed except for S7. This shows that focusing on macroscopic network structure and adopting the principle of attention allocation of com-mon friends in the triadic closure structure can effectively over-come the problem of network sparsity in predicting friendships between users in different circles. It is noted that the macroscopic network structure consists of node pairs, their common neighbors and the friends of the common neighbors.

At the same time, the performance of IAALPAa and IAALPAb is better than that of DT and other algorithms, which shows that the mechanism to select complementary AAIs by SVM is effective. That is to say, using the SVM model to identify the complementary AAIs and integrating them together can deliver better prediction results than the random combination of AAIs.

It can also be seen from Tables 5 and 6 that IAALPAa is significantly better than SVM. These results show that the IAALPA framework proposed in this study is far superior to the SVM framework.

Additionally, from Figs. 6 and 7, it can be concluded that the accuracy of DT is higher than that of other non-combination AAIs,

Table 7
Performance comparison under various networks in the Twitter experiments.

Network ID	Node count	Circle count	Node count per circle	S7	S12	IAALPAa
200214366	59	12	4.9167	0.8262	0.8454	0.8267
29514951	39	7	5.5714	0.8327	0.8306	0.834
98633794	33	5	6.6	0.8603	0.8563	0.8577
7888452	80	9	8.8889	0.8906	0.8894	0.8862
35012277	20	2	10	0.9361	0.9164	0.9394
356963	114	11	10.3636	0.7532	0.7473	0.7673
15070932	21	2	10.5	0.82	0.8323	0.8283
351092905	55	4	13.75	0.9172	0.9066	0.9139
18886852	70	5	14	0.8917	0.8893	0.8959
16652550	70	5	14	0.8914	0.8894	0.8908
134943586	209	11	19	0.8877	0.886	0.89
Average AUC				0.8643	0.8626	0.8664

Table 8
Performance comparison under various networks in the Google+ experiments.

Network ID	Node count	Circle count	Node count per circle	S7	S12	IAALPAa
100715738096376666180	46	3	15.3333	0.7308	0.8833	0.9189
104672614700283598130	32	2	16	0.8281	0.9218	0.9275
114122960748905067938	222	6	37	0.8658	0.8752	0.8661
104607825525972194062	80	2	40	0.8645	0.9113	0.9172
113356364521839061717	116	2	58	0.8354	0.9077	0.9102
103503116383846951534	351	5	70.2	0.8269	0.8581	0.8648
107362628080904735459	168	2	84	0.8298	0.8492	0.8424
107203023379915799071	172	2	86	0.8493	0.9305	0.9347
115516333681138986628	305	3	101.6667	0.915	0.9249	0.9238
110971010308065250763	521	4	130.25	0.8637	0.9025	0.9096
101499880233887429402	514	2	257	0.9557	0.9625	0.9631
Average AUC				0.8513	0.9025	0.9071
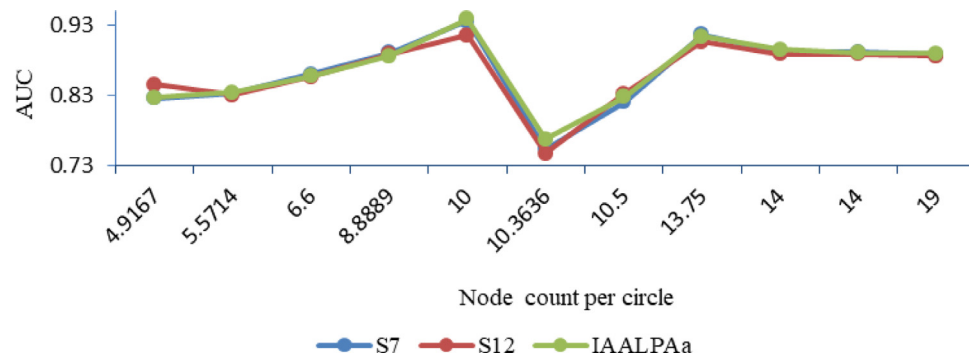
Fig. 8. Performance comparison in different node density in Twitter.
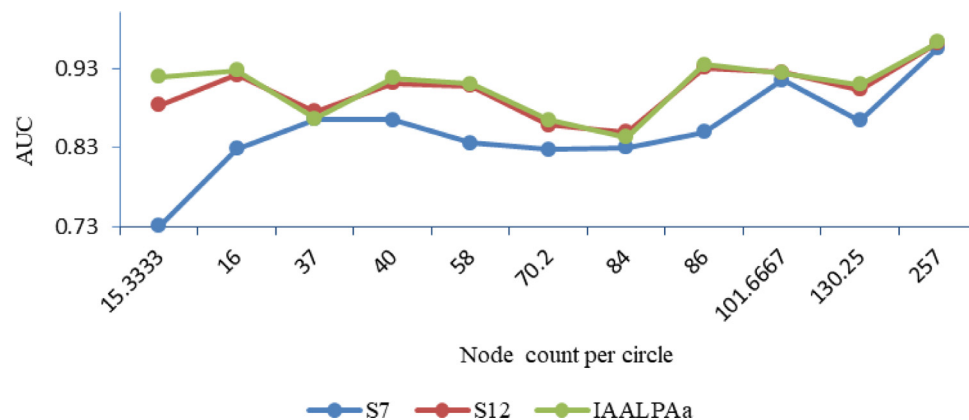
Fig. 9. Performance comparison in different node density in Google+.

which means the mechanism to select the appropriate AAIs based on network characteristics is effective.

Finally, paired samples of any two algorithms' AUC in 100 experiments were collected and Student t-test was adopted to determine whether the means of the two samples were statistically equal. Consequently, all the determined P-Values are less than the significant level 5%, which shows that there are significant differences among algorithms. These results show that IAALPA can provide reliable and accurate prediction of friendships between users in various product circles in the brand community, and at the same time, it also offers a promising method to recommend friends for cross-marketing in online brand community.

5.3. Analysis of IAALPA performance under various networks

In order to further analyze the accuracy of friend group recommendation of IAALPA in different circle densities, in this study, networks with the number of product circles ranging from a much smaller value at 2 to a larger value at 12 in the Twitter and fluctuating from a much smaller value at 2 to a larger value at 6 in Google+ data set were selected. In addition to IAALPAa, we separately selected AAIs with better performance in Tables 5 and 6, namely S7 and S12, to analyze their performance under various network densities. Tables 7, 8, Figs. 8 and 9 give the prediction accuracy of IAALPA, S7, and S12 in different networks.

It can be seen from Tables 7 and 8 that IAALPA has high recommendation accuracy in both networks with more product circles and networks with fewer circles and acquires the best average AUC 0.8664 in the Twitter experiments and 0.9071 in the Google+ experiments. It can be seen from Figs. 8 and 9 that IAALPAa produces the superior performance in both circles with large number of nodes (i.e. node density) and circles with small number of nodes (i.e. node sparseness). However, S12 achieves the worst performance when the node density is intermediate in Twitter, and S7 obtains the worst performance when the node density is small in Google+. These indicate that IAALPA is good at recommending friend group in both dense and sparse networks.

6. Conclusion

The group of friends has a great influence on the attitudes, behaviors and even feelings of individuals. In this study, the users in one product circle who are likely to be friends with the target customer in another product circle are identified and recommended by IAALPA. By friend group's influence, cross-marketing of products can be achieved successfully, that is, new target customer will purchase products preferred by friend group.

To overcome the problem of network sparsity being faced when predicting the friendships between users in different circles, based on the principle of attention allocation of common friends in the triadic closure structure, IAALPA extracts the AAIs based on not only the microscopic network structure, but also macroscopic network structure. Consequently, IAALPA can comprehensively describe the possibility of links between node pairs. Furthermore, instead of using fixed AAI to predict links in various networks, DT model is developed to select the suitable AAI for the given network based on the network characteristics of the common friend density and the dispersion level of common friends' attention. Subsequently, based on the value, direction, and ranking similarities of AAIs, SVM is designed to identify AAIs which are complementary to the optimal AAI selected by DT and the ideal composite AAI by integrating these mutually complementary AAIs is constructed. As a result, the problems of performance degradation in combination prediction model, which is caused by randomly integrating SLPAs, are overcome once and for all.

Experimental results on 971 online communities in the Twitter network and 132 online communities in the Google+ network show that the IAALPA proposed in this study achieves more accurate and reliable link prediction performance. And AAIs constructed based on the idea of attention allocation of common friends in both microscopic and macroscopic network structure are superior in the harsh scenario where network sparsity is faced when predicting friendships between users in different circles. Therefore, IAALPA provides strong support for marketers to use online brand communities to achieve profitable cross-marketing.

In the future, we will propose more link prediction algorithms and integrate them into the proposed prediction framework to further improve prediction accuracy. Furthermore, we will explore the performance of the framework in different types of network structures, such as the dynamic network and the network integrating different social networks.

Declaration of Competing Interest

The authors declare that we have no conflicts of interest to this work.

Credit authorship contribution statement

Shugang Li: Conceptualization, Formal analysis, Funding acquisition, Methodology. **Xuewei Song:** Validation. **Hanyu Lu:** Writing - original draft. **Linyi Zeng:** Data curation, Investigation. **Miaojing Shi:** Writing - review & editing. **Fang Liu:** Writing - review & editing.

Acknowlgedgments

References

Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the web. *Social Networks, 25*(3), 211–230. doi:10.1016/S0378-8733(03)00009-1.

Aral, S. (2013). What would Ashton do–and does it matter? New research reveals the power and limits of "influencers". (Idea Watch). *Harvard Business Review, 91*(5), 25–27.

Backstrom, L., Bakshy, E., Kleinberg, J., Lento, T. M., & Rosenn, I. (2011). Center of attention: How Facebook users allocate attention across friends. In *In Proceedings of the fifth international AAAAI conference on weblogs and social media* (pp. 34–41). AAAI. doi:10.1016/S0378-4371(02)00736-7.

Barabâsi, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications, 311*(3–4), 590–614. doi:10.1016/S0378- 4371(02)00736-7.

Brogi, S. (2014). Online brand communities: A literature review. *Procedia-Social and Behavioral Sciences, 109*, 385–389. doi:10.1016/j.sbspro.2013.12.477.

Fan, C., Liu, Z., Lu, X., Xiu, B., & Chen, Q. (2017). An efficient link prediction index for complex military organization. *Physica A: Statistical Mechanics and its Applications, 469*, 572–587. doi:10.1016/j.physa.2016.11.097.

Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). A survey on ensemble learning for data stream classification. *ACM Computing Surveys, 50*(2), 23:1–23:36. doi:10.1145/3054925.

He, C., Li, H., Fei, X., Yang, A., Tang, Y., & Zhu, J. (2017). A topic community-based method for friend recommendation in large-scale online social networks. *Concurrency and Computation: Practice and Experience, 29*(6), 1–20. doi:10.1002/cpe.3924.

Huang, Z. (2006). *Link prediction based on graph topology: The predictive value of generalized clustering coefficient.*

John, L. K., Mochon, D., Emrich, O., & Schwartz, J. (2017). What's the value of a like? Social media endorsements don't work the way you might think. *Harvard Business Review, 95*(2), 108–115.

Kaya, B., & Poyraz, M. (2016). Unsupervised link prediction in evolving abnormal medical parameter networks. *International Journal of Machine Learning and Cybernetics, 7*(1), 145–155. doi:10.1007/s13042-015-0405-y.

Lei, J., & Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics, 43*(1), 215–237. doi:10.1214/14-AOS1274.

Li, Y. M., Chou, C. L., & Lin, L. F. (2014). A social recommender mechanism for location-based group commerce. *Information Sciences, 274*, 125–142. doi:10.1016/j.ins.2014.02.079.

Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology, 58*(7), 1019–1031. doi:10.1002/asi.20591.

Liu, L., Yu, S., Wei, X., & Ning, Z. (2018). An improved Apriori–based algorithm for friends recommendation in microblog. *International Journal of Communication Systems, 31*(2), 1–11. doi:10.1002/dac.3453.

Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications, 390*(6), 1150–1170. doi:10.1016/j.physa.2010.11.027.

Mantas, C. J., Abellán, J., & Castellano, J. G. (2016). Analysis of Credal-C4. 5 for classification in noisy domains. *Expert Systems with Applications, 61*, 314–326. doi:10.1016/j.eswa.2016.05.035.

Martínez, V., Berzal, F., & Cubero, J. C. (2017). A survey of link prediction in complex networks. *ACM Computing Surveys, 49*(4), 1–33. doi:10.1145/3012704.

Muniz, C. P., Goldschmidt, R., & Choren, R. (2018). Combining contextual, temporal and topological information for unsupervised link prediction in social networks. *Knowledge-based Systems, 156*, 129–137. doi:10.1016/j.knosys.2018.05.027.

Murata, T., & Moriyasu, S. (2008). Link prediction based on structural properties of online social networks. *New Generation Computing, 26*(3), 245–257. doi:10.1007/s00354-008-0043-y.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 404–409. doi:10.1073/pnas.98.2.404.

Shan, Z., Kong, J., Zhang, Y., Li, H., et al. (2018). Remote sensing investigation method of object-oriented crops with special characteristics. *Journal of Geo-information Science, 20*(10), 1509–1519. doi:10.12082/dqxxkx.2018.180116.

Soares, P. R. S., & Prudêncio, R. B. C. (2013). Proximity measures for link prediction based on temporal events. *Expert Systems with Applications, 40*(16), 6652–6660. doi:10.1016/j.eswa.2013.06.016.

Solomon, M. R. (2016). Consumer behaviour: Buying, having, and being(12th Edition). *Pearson*.

Thomas, V. L., Jewell, R. D., & Jennifer, J. W. (2015). Hidden consumption behaviour: An alternative response to social group influence. *European Journal of Marketing, 49*(3/4), 512–531. doi:10.1108/EJM-06-2013-0336.

Wang, Q., Wu, Y., & Yao, Y. (2017). Classification of heart disease based on ultrasound data of transesophageal echocardiography. *Journal of Computer Applications, 37*(S1), 221.

Wang, X., & Xing, Y. (2019). An online support vector machine for the open-ended environment. *Expert Systems with Applications, 120*, 72–86. doi:10.1016/j.eswa.2018.10.027.

Wang, X., & Xue, H. (2010). Brand community social capital, perceived value and brand loyalty. *Journal of Management Science, 23*(6), 53–63.

Whittler, T. E., & Spira, J. S. (2002). Model's race: A peripheral cue in advertising messages? *Journal of Consumer Psychology, 12*(4), 291–301. doi:10.1207/15327660260382333.

Wu, M., & Tang, Y. (2014). Topic model based link prediction for directed social network. *Journal of Computational Information Systems, 36*(1), 152–158. doi:10.13718/j.cnki.xdzk.2014.01.023.

Xiang, E. W. (2008). *A survey on link prediction models for social network data*.

Xiao, Y., Li, X., Wang, H., Xu, M., & Liu, Y. (2018). 3-HBP: A three-level hidden Bayesian link prediction model in social networks. *IEEE Transactions on Computational Social Systems, 5*(2), 430–443. doi:10.1109/TCSS.2018.2812721.

Yuan, T., Cheng, J., Zhang, X., Liu, Q., & Lu, H. (2015). How friends affect user behaviors? An exploration of social relation analysis for recommendation. *Knowledge-based Systems, 88*, 70–84. doi:10.1016/j.knosys.2015.08.005.

Zhu, J.-Q., Lu, L., & Ma, C.-M. (2015). From interest to location: Neighbor-based friend recommendation in social media. *Journal of Computer Science and Technology, 30*(6), 1188–1200. doi:10.1007/s11390-015-1593-3.