

Article The Research of "Products Rapidly Attracting Users" Based on the Fully Integrated Link Prediction Algorithm

Shugang Li ¹, Ziming Wang ¹, Beiyan Zhang ¹, Boyi Zhu ^{1,*}, Zhifang Wen ^{1,*} and Zhaoxu Yu ²

- ¹ School of Management, Shanghai University, Shanghai 200444, China; luck_li@shu.edu.cn (S.L.); shuwzm2021@shu.edu.cn (Z.W.); zhangbeiyan@shu.edu.cn (B.Z.)
- ² Department of Automation, East China University of Science and Technology, Shanghai 200237, China; yyzx@ecust.edu.cn
- * Correspondence: boyi_star@shu.edu.cn (B.Z.); 21720674wzf@shu.edu.cn (Z.W.)

Abstract: One of the main problems encountered by social networks is the cold start problem. The term "cold start problem" refers to the difficulty in predicting new users' friendships due to the limited number of links those users have with existing nodes. To fill the gap, this paper proposes a Fully Integrated Link Prediction Algorithm (FILPA) that describes the social distance of nodes by using "betweenness centrality," and develops a Social Distance Index (SDI) based on micro- and macro-network structure according to social distance. With the aim of constructing adaptive SDIs that are suitable for the characteristics of a network, a naive Bayes (NB) method is firstly adopted to select appropriate SDIs according to the density and social distance characteristics of common neighbors in the local network. To avoid the risk of algorithm accuracy reduction caused by blind combination of SDIs, the AdaBoost meta-learning strategy is applied to develop a Fully Integrated Social Distance Index (FISDI) composed of the best SDIs screened by NB. The possible friendships among nodes will then be comprehensively presented using high performance FISDI. Finally, in order to realize the "products rapidly attracting users" in new user marketing, FILPA is used to predict the possible friendship between new users in an online brand community and others in different product circles.



Citation: Li, S.; Wang, Z.; Zhang, B.; Zhu, B.; Wen, Z.; Yu, Z. The Research of "Products Rapidly Attracting Users" Based on the Fully Integrated Link Prediction Algorithm. *Mathematics* 2022, *10*, 2424. https:// doi.org/10.3390/math10142424

Academic Editor: Catalin Stoean

Received: 29 May 2022 Accepted: 7 July 2022 Published: 12 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Keywords: friend recommendation; link prediction; FILPA; cold start; rapidly gaining followers

MSC: 05C82

1. Introduction

Social networks can provide a low-cost, effective, and efficient marketing platform for enterprises. By leveraging social networks, marketing enterprises can swiftly grow their user base, accumulate user groups, and realize "products rapidly attracting users" among new users. An online brand community based on a social network is a relational network grouped by users with common interests and desires for brands. This group is characterized by close unity, highly efficient communication, and convergent actions [1]. In order to expand the marketing scale and benefits of an online brand community, marketers should attract new users to join the brand community and keep purchasing the brand's products. Therefore, it is necessary to quickly foster friendships between new users and those of the brand community, and then carry out marketing to new users by leveraging the influence of these friendships.

Based on the understanding, homogeneity, and trust between friends [2], users can get valuable, trustworthy [3], and high-quality [4] commodity information through frienddriven communication. At the same time, friendship leads users to be more inclined to believe that the sales operations conducted by marketers are genuinely based on the quality of the product [5]. As a consequence, users are more willing to accept information shared friends [6], making it easier to stimulate users' intent to purchase [4] and influence their purchasing decisions [7]. For example, the Florida government set up Students Work Against Tobacco to solve the problem of adolescent smoking. It took half of the "customers" from well-capitalized tobacco manufacturers and turned them into propagandists against adolescent smoking by leveraging the influence of friends. The influence of friend recommendations, the most effective of customer purchase motivations, is the key to its success. However, the extremely limited number of links between new users and other nodes will cause the cold start problem when attempting to rapidly extend user groups through the use of social networks.

The Scoring Link Prediction Algorithm (SLPA) [8], which predicts whether there are links between node pairs in social network on the basis of network topology, is a major method used in the friend recommendation process. To date, many kinds of SLPAs, such as Common Neighbors (CN), Preferential Attachment (PA), and Resource Allocation (RA) have been used for friend recommendation. However, the above traditional algorithm based on a local triadic closure structure (users with common neighbors are more likely to become friends) cannot solve the problems of sparse network and cold start due to the scarcity of available information when forecasting the connections of new users. Although some scholars have considered that a large range of local networks would increase the amount of information [9], this information has little reference value, due to the lack of deep excavation capabilities. In addition, although the combination of single indexes can help improve their link-prediction performance, the blind combination of indexes does not produce the expected results [10], therefore this method cannot adequately solve the cold start problem.

Although some scholars have proposed global and quasi-local measures [11], there is no research on link prediction from the perspective of network connection strength. Nicholas and James put forward the Three Degrees of Influence Rule, pointing out that close social distance was a strong connection that could trigger behaviors, while estranged social distance could affect the construction of friendship. In order to enrich the available information of link prediction in the network and overcome the cold start problem faced by new users' friendship prediction, this paper designs a Social Distance Index (SDI) to predict the friendship of users, and uses the social distance of related nodes in a large range to predict the possibility of friendship. We then propose a Fully Integrated Link Prediction Algorithm (FILPA) which comprehensively describes the possible links from different aspects.

Centrality is an important index used to describe the characteristics of a network. Therefore, this paper describes the distance of nodes by using the betweenness centrality of global variables, and proposes a measure of social distance. Then, according to the principle of social distance, SDIs based on micro- and macro-network structures are developed. In order to construct adaptive SDIs which are suitable for the characteristics of network structure, FILPA firstly uses a naive Bayes (NB) method to select appropriate SDIs according to network maturity characteristics such as the density and social distance in the local network. Then, in order to ensure that the above indicators can comprehensively and accurately describe the possibility of establishing friendship and avoid the combination indicators performance degradation that is caused by SDIs with weak performance, a NB model is constructed to screen out suitable SDIs according to different network maturities. Subsequently, FILPA applies AdaBoost meta-learning strategy to develop the Fully Integrated Social Distance Indexes (FISDIs), which are composed of SDIs suitable for different network maturities. FISDIs are used to predict the possibility of friendships between users in the online brand community and new users, thereby recommending friendships in order to realize the "products rapidly attracting users" for new user marketing. Through the actual data of the Twitter network, it is proven that the FILPA designed in this paper can overcome the cold start problem and achieve excellent results in new user marketing.

This paper is organized as follows: Section 2 presents the link prediction algorithm; Section 3 introduces new user marketing and SDI; Section 4 provides FILPA; Section 5 verifies the effectiveness of FILPA; Section 6 gives a summary of the paper.

2. Link Prediction

The link prediction problem can be formally defined as follows: Given the snapshot of social network at time t, we can infer the missing link subset at current time or predict the link subset that will be added to the network at that time $t + \Delta$ by defining a similarity function or probability function or by using supervised learning or optimization technology [13]. The research idea of link prediction based on complex networks has attracted the attention of researchers from many fields, such as community structure detection [14], rumor spreading, and security monitoring [15]. The existing link prediction methods can be classified into similarity-based methods, probabilistic statistical techniques, algorithmic methods, and preprocessing approaches [13]. Among these methods, due to its low computational complexity and potential for application in large-scale networks, the similarity method has become the most studied method [13,16].

In the light of the amount of information in the network topology, the similarity method can be divided into two categories: local method and global method [17]. The method based on local similarity considers the structure information related to the neighborhood of nodes. If two nodes are similar, they are more likely to be linked together. The disadvantage of the method based on local similarity is that there is little information about common neighbors, which cannot solve the problems of network sparsity and cold start. Some scholars consider the global similarity method and use the topological information of the whole network to score the unlinked node pairs [18]. For example, Rafiee S et al. [19] considered clustering coefficient and the neighbors of shared neighbors, which led to better performance. Zareie A et al. [20] considered direct and indirect common neighbors to predict potential relationships. While this increased the amount of information, it did not provide a lot of effective information. Nicholas and James emphasized the Three Degrees of Influence Rule [12], that is, social distance could affect the relationship between users in a social network. In addition, existing literature found that betweenness centrality best quantifies the path connectivity [21]. For example, scholars divided a network into communities and used the centrality of nodes [22] or community centrality [23] to describe the similarity of nodes for prediction, thereby obtaining good prediction results. Therefore, in order to enrich the available information of link prediction in the network, this paper uses the social distance of related nodes, which is described by betweenness centrality in a wide range, to solve the problems of network sparsity and cold start.

Because a single SDI cannot fully describe the network characteristics and cannot be applied to the cold start environment faced by the prediction of new users' friends, it is necessary to develop a combined link prediction algorithm. In the literature, a lot of combined link prediction algorithms have been developed using complex network characteristics. For example, Lu et al. [24] combined rich semantic and time information, fully mining the meta-path in the network, and proposed a unified link prediction framework (UniLPF) to predict academic relationships in academic networks based on the similarity method. Ozcan et al. [25] combined time information, correlation between link evolution and multi-type relationships, and the link information of nodes, and using local and global similarity methods, put forward a new method called multivariate time series link prediction to predict links in dynamic, undirected, weighted, or unweighted heterogeneous social networks. Hristova et al. [26] constructed a multi-layered online social network using Twitter and location-based social networks, and used a similarity method to predict links across social networks according to their structure and interaction characteristics. Ozcan et al. [27] used a nonlinear autoregressive exogenous model (NARX) and proposed a multivariate method to predict links, which combined the correlation between different link types and the influence of different local and global topological similarity methods in different time periods.

The above combined algorithms cannot consider the complementarity between single indexes. Therefore, it cannot guarantee the comprehensive description of the connection possibilites between nodes. To fill the gap, this paper innovatively proposes a method of selecting the appropriate SDI according to the maturity of the network. Due to the diversity of network structures, a single index can only describe the characteristics of social distance from a single point of view, and combining single indexes randomly can cause the repeated description of network characteristics [10], which cannot enrich the network characteristics described by the integrated index. For example, Lü L et al. [28] applied the weighted indices to the co-authorship network and the US air transportation network; they found that the weighted indices could not provide better performance. Liben-Nowell and Kleinberg [29] reported a similar observation for weighted Katz index and unweighted Katz index, which used all paths between two nodes in a network to measure node similarity. In order to fully find the characteristics of potential social distance, this paper uses the AdaBoost meta-learning strategy to identify FISDI composed of the best SDIs that are screened by NB, and adopts high-performance FISDI to comprehensively describe the possibility of network users becoming friends.

3. New User Marketing and SDI

3.1. Product Marketing in Brand Community

Define the online brand community as P (V,E), where V is the node set, representing users in the community, and E is the edge set, representing the friendship among users. In the brand community, users will form a circle because they like the same kind of products. Assume that, in the online community P, the user set in the circle of product D is defined as V_D. In the process of "products rapidly attracting users" for marketing to new users, users in the circle of D are recommended to new users as friends, and product marketing is realized with the influence of friendship.

Figure 1 shows the schematic diagram of "products rapidly attracting users" based on the users' friend recommendations in the circle of product P. Figure 1a represents the network before evolution, Figure 1b shows the network after evolution. In Figure 1a, the users that belong to online community P are 1, 2, 3, 4, and 5. Assuming that the new user is 6, the purpose of friend recommendation is to enable user 6 to establish friendship with the users in circle P, so as to realize "products rapidly attracting users". By making use of the influence of friendship, we can help products accumulate a customer base and expand word-of-mouth quickly with a high conversion rate to achieve successful product marketing.



Figure 1. Schematic diagram of "products rapidly attracting users". (**a**) represents the network before evolution, the users that belong to online community P are 1, 2, 3, 4, and 5, (**b**) shows the network after evolution.

3.2. SDI for Friend Recommendation

In general, new users have few friends; the cold start problem is faced when attempting to recommend friends to new users for the purpose of product marketing [30]. In this case,

traditional SLPA cannot guarantee accurate friend recommendations for new users. To solve this problem, this paper proposes a new link prediction index based on the Three Degrees of Influence Rule (TDIR) [12]. TDIR states that, if the social distance is within three degrees, it can be called a strong connection and can trigger actions; while weak connections will occur when social distance is more than three degrees, prompting only the transmission of information between users. Generally speaking, whatever we do or say, it will have impacts on our friends (first degree), our friends' friends (second degree), and even the friends of our friends' friends (third degree). If the relationship is beyond the third degree, it will reduce our influence. According to TDIR, this paper presents the social intensity of users and predicts the possibility of connection between nodes by the shortest social distance between users instead of being confined to a social distance within three degrees.

Similarly, if the social distance between a user's friends and other users is close, we can assume the user potentially has close social distance to the other users (it will arise in the evolved network). It will be easier to become friends with two nodes when their common friends have a closer social distance to other users.

Firstly, the betweenness centrality is used to represent the social distance between any node *z* and other nodes (i.e., $C_b(z)^{\theta} = \frac{1}{(|V|-1)*(|V|-2)} * \sum_{i,j,z \in V, i \neq j} \frac{n_{ij}(z)}{n_{ij}} * \theta + 1$, where θ represents the constant parameter, |V| is the number of nodes, n_{ij} means the total number of shortest paths between node pairs (i, j), and $n_{ij}(z)$ indicates the number of shortest paths through node *z*). s(x) expresses the potential social distance of node *x*, namely, $s(x) = \sum_{z \in \Gamma(x)} C_b(z)^{\theta}$, where $\Gamma(x)$ represents the neighbor set of node *x*. In each node pair (x, y) to be predicted, one is a new node, the other is a known node in the circle.

On the basis of the principle of the potential social distance for constructing friendship, this paper builds SDIs based on a micro- and macro-network to comprehensively describe the implicit social distance information in an extremely sparse network.

3.2.1. SDI Based on Microstructure

Firstly, based on the principle that the closer the social distance between the common neighbor of two nodes and other nodes, the more likely it is that those two nodes will become friends, this paper proposes five indicators. In detail, CND0 is the direct influence of the distance between the pair of nodes, which is caused by the social distance between the common neighbor and other nodes. CND1–CND4 are the social distance influence coefficients of the common neighbor (the ratio of the social distance between a common neighbor's direct neighbors and other nodes to the social distance between their indirect neighbor and other nodes), and a larger coefficient will help spread the influence of social distance in the local network better.

In this case, the closer the social distance between the direct neighbors of the common neighbor and other nodes, the easier it is for this node pair to become friends. The details are shown in Table 1.

SDI	Formula
CND0	$S^{CND0}_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} C_b(z)^{ heta}$
CND1	$S_{xy}^{CND1} = rac{\sum_{z \in \Gamma(x) \cap \Gamma(y)} C_b(z)^ heta}{\min\{s(x), s(y)\}}$
CND2	$S_{xy}^{CND2} = rac{\sum_{z \in \Gamma(x) \cap \Gamma(y)} C_b(z)^ heta}{\max\{s(x), s(y)\}}$
CND3	$S_{xy}^{CND3} = rac{\sum_{z \in \Gamma(x) \cap \Gamma(y)} C_b(z)^ heta}{s(x) * s(y)}$
CND4	$S_{xy}^{CND4} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{C_b(z)^{ heta}}{s(z)}$

Table 1. SDI based on microstructure.

3.2.2. SDI Based on Macrostructure

The cold start problem is normally caused by the small number of links belonging to a new node., This section innovatively proposes more SDIs, which mainly take the role of indirect common neighbors (i.e., friends of common neighbors) into account. Specifically, the social distance of common neighbors' friends can change the connection of common neighbors and can have indirect effects on whether the target node pairs can establish links. These newly proposed SDIs are CCND1, CCND 2, CCND 3, and CCND4.

(a) CCND1

CCND1 is the social distance influence coefficient, that is, the ratio of the social distance between the direct neighbors of a common neighbor and other nodes to the social distance between the indirect neighbors of that common neighbor and other nodes. The larger the coefficient, the better the conductivity of the influence of social distance in the local network. In this case, the closer the social distance between the direct neighbor of the common neighbor and other nodes, the easier it is for this node pair to become friends.

$$S_{xy}^{CCND1} = \sum_{\Gamma(z), z \in \Gamma(x) \cap \Gamma(y)} \frac{C_b(\Gamma(z))^{\nu}}{s(\Gamma(z))}$$
(1)

Λ

(b) CCND2

Furthermore, this paper considers the influence coefficient of social distance of common neighbors and the influence of the combination of social distance between neighbors on the social distance between node pairs. Therefore, CCND2 is defined as shown in Formula (2).

$$S_{xy}^{CCND2} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{C_b(z)^{\theta}}{s(z)} * \left(\alpha * c(z) + \beta * (1 - c(z))\right)$$
(2)

Among them, c(z) represents the agglomeration coefficient of the node, indicating the social distance between neighbor nodes, namely, $c(z) = \frac{2*e_z}{|\Gamma(z)|*(|\Gamma(z)|-1)}$, where e_z means the actual number of edges between neighbors of node z, $|\Gamma(x)|$ expresses the number of neighbors of node x, α , and β are the constant parameters.

(c) CCND3

For any node *x*, we analyze the influence of the social distance between its neighbors and other nodes, as well as the influence of the social distance between neighbors on the social distance of other nodes, and define the combination influence coefficient of social distance, i.e., $\frac{C_b(x)^{\theta}}{s(x)*c(x)}$.

At the same time, the moderating effect of the focus degree of common neighbors to node pairs (represented by 1-c(z)) on this combination influences the coefficient. Through the combination of the above coefficients, we describe the combination coefficient of social distance influence (namely CCND3). The larger the coefficient, the better the conductivity of social distance influence, as shown in Formula (3).

$$S_{xy}^{CCND3} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \left(\frac{C_b(z)^\theta * (1 - c(z))}{s(z)} + \frac{C_b(x)^\theta}{s(x) * c(x)} + \frac{C_b(y)^\theta}{s(y) * c(y)} \right)$$
(3)

(d) CCND4

We integrate the social distance influence coefficient CND4 of the micro-network with the social distance influence coefficients CCND2 and CCND3 of the macro-network to define a new combined index CCND4, as shown in Formula (4).

$$S_{xy}^{CCND4} = \delta * S_{xy}^{CND4} + \varepsilon * S_{xy}^{CCND2} + \sigma * S_{xy}^{CCND3}$$
(4)

4. FILPA

To overcome the cold start problem of new users' friend recommendation, this section deeply excavates the information contained in the network structure, that is, the different nodes' social distance within the network is different, and then puts forward FILPA. The possibility of new users becoming friends with other nodes in different circles is fully described by adaptively building a high-performance FISDI for a specific circle structure. FILPA considers the relationship between the joint effect of various characteristics of the local network and the prediction performance of the algorithm, selects the NB model, and chooses the appropriate SDIs according to the maturity of the network. Moreover, because a single SDI often makes too high or too low of an estimation, the random combination of SDIs cannot guarantee good results every time. FILPA adopts the AdaBoost meta-learning strategy to identify the fully integrated index composed of suitable SDIs screened by NB, then selects the index with the highest precision. Figure 2 shows the structure of FILPA. In Figure 2, when the accuracy of the combination of scoring index A and the other two indexes is higher than that of index A, the two indexes are considered to be the integrated index.



Figure 2. Structure of FILPA.

4.1. Algorithm Evaluation

The area under the curve (AUC) [31] is the most commonly used standard index for measuring the accuracy of link prediction algorithms. AUC randomly selects the connected and unconnected node pairs in the test set, and compares their score values obtained by SDIs. In *m* independent comparisons, if the number of connected node pairs with a higher score is *m*1 times, and the number of connected node pairs with a higher or equal score is *m*2 times, then the AUC is shown in Formula (5).

$$AUC = \frac{m1 + 0.5(m2 - m1)}{m}$$
(5)

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

4.2. NB for Screening the Best SDI

Since the NB model requires few estimated parameters, is not very sensitive to missing data, and the algorithm is relatively robust, FILPA adopts the NB model to adaptively select SDIs suitable for the link prediction characteristics of new users, then selects appropriate SDIs according to the network maturity. In this paper, the NB model is used to select the appropriate SDI according to the density of common neighbors and the measure of social distance.

4.2.1. Discriminant Factors for the Best Indexes

In order to select SDIs suitable for different networks, this paper uses network maturity to distinguish the best indexes. The evaluation of network maturity includes the quality and quantity characteristics of networks. The quantity aspect refers to the network density and network connection scale. The qualitative aspects include the social distance of the node in the networks, and the stability, diversity, and dispersion of network connections [32]. This paper describes the network maturity characteristics of the local network comprehensively from two dimensions, that is, the density of common neighbors and the social distance between the common neighbor and other nodes, then selects the appropriate SDIs accordingly. In order to overcome the influence of cold start on the accuracy of the algorithm, for each node pair to be predicted, the characteristics of its direct common neighbors and one-step, two-step, and three-step indirect common neighbors are considered simultaneously.

(a) Density of common neighbors

The more common neighbors a node pair has, the more intimate the social distance between them, and the more likely the pair is to become friends. The network features related to the density of common neighbors include density [33] and centrality of mean eigenvector [34], as shown in Formulas (6) and (7).

$$d = \frac{2 * |E|}{|V| * (|V| - 1)} \tag{6}$$

$$C_E = \frac{1}{|V|} * \sum_{i=1}^{|V|} C_E(i)$$
(7)

where $C_E(i) \propto \sum_{j \in \Gamma(i)} C_E(j)$ can be obtained by recursively solving the centrality of eigenvector of node *i*, |E| represents the total number of edges of the network. The specific algorithm is described as follows.

set
$$C_E(i) = 1$$
 for all i;
do
 $\{compute \ C_E^*(i) = \sum_{j \in \Gamma(i)} C_E(j) \text{ for all } i;$
 $set \ \lambda = \sqrt{\sum_{i=1}^{|V|} [C_E^*(i)]^2};$
 $set \ C_E(i) = \frac{C_E^*(i)}{\lambda} \text{ for all } i;$
 $\} Until (\lambda stops changing)$

(b) Social distance between the common neighbor and other nodes

Since a greater number of the shortest paths pass through the nodes of common neighbors, the social distance between the common neighbor and other nodes is more intimate. Therefore, we considered the indexes related to the shortest paths, such as average proximity centrality, average betweenness centrality, and the size of average connected groups [35], as shown in Formulas (8)–(10).

$$C_C = \frac{1}{|V|} \sum_{i=1}^{|V|} \left(\frac{1}{|V| - 1} \sum_{j=1, j \neq i}^{|V|} d_{ij} \right)^{-1}$$
(8)

$$C_B = \frac{1}{|V|} \sum_{i=1}^{|V|} \frac{2}{(|V|-1)(|V|-2)} \sum_{j < k} \frac{n_{jk}(i)}{n_{jk}}$$
(9)

$$s = \frac{1}{t} \sum_{i=1}^{t} N(i)$$
 (10)

where *t* represents the number of connected groups, N(o) represents the size of the o-th connected group, and d_{ij} represents the shortest path between node *i* and node *j*.

4.2.2. NB Model

The independent variable (X) of the learning sample is defined as the above five local network characteristic indexes, and the dependent variable (Y) is the SDI label with the largest AUC value. Suppose that the whole training network set is $W = (X_1, X_2, X_3, X_4, X_5, Y)$, for each network, according to its AUC value on each network characteristic index mentioned in Section 4.2.1, it is assigned to the class y_i corresponding to the SDIs with the largest AUC value. Assume that there are L kinds of labels, that is, $y_i \in F = \{y_1, y_2, \dots, y_L\}$, a partition $\{T_1, T_2, \dots, T_L\}$ of W is obtained. When the input network feature vector is $(x_1, x_2, x_3, x_4, x_5)$, the probability that the sample belongs to class y_i is shown in Formula (11).

$$P(Y = y_i | X_1 = x_1, X_2 = x_2, \dots, X_5 = x_5) = \frac{P(X_1 = x_1, X_2 = x_2, \dots, X_5 = x_5 | Y = y_i) * P(Y = y_i)}{P(X_1 = x_1, X_2 = x_2, \dots, X_5 = x_5)}$$
(11)

Based on the definition of each feature, assuming that they are independent of each other, according to the principle of NB, the molecules of Formula (11) can be further expressed as $P(X_1 = x_1, X_2 = x_2, ..., X_5 = x_5 | Y = y_i) * P(Y = y_i) = \prod_{j=1}^5 P(X_j = x_j | Y = y_i) * P(Y = y_i)$. Since the denominator $P(X_1 = x_1, X_2 = x_2, ..., X_5 = x_5)$ of Formula (11) is the same for all classes, the class with the largest molecule is the class to which the sample belongs. Therefore, the NB classifier can be expressed as Formula (12).

$$h_{nb} = argmax_{y_i \in F} \prod_{j=1}^{5} P(X_j = x_j | Y = y_i) * P(Y = y_i)$$
(12)

In Formula (12), $P(Y = y_i)(i = 1, 2, ..., L)$ represents the prior probability of the class y_i , which is obtained by the maximum likelihood estimation criterion, namely, $P(Y = y_i) = \frac{|T_i|}{|W|}$. $P(X_j = x_j | Y = y_i)$ represents the conditional probability density of the occurrence of the j-th local network feature of the training set in class y_i . Although normal distribution is the most popular approach to deal with continuous variables, it has the least information and the most uncertainty, which will result in a highly robust algorithm. Therefore, this paper adopts a non-parametric density estimation method, namely, kernel density estimation, to estimate the probability density function directly from the training samples, that is, $P(X_j = x_j | Y = y_i) = \frac{1}{|W|*h} \sum_{j=1}^{|W|} K(\frac{x_j - x_{ji}}{h})$. Among them, x_j represents the j-th local network feature value, x_{ji} represents the j-th local network feature value value value value value value value value

cates the kernel function, being set to the most common Gaussian kernel function, namely, $K\left(\frac{x_j-x_{ji}}{h}\right) = \frac{1}{h*\sqrt{2\pi}} * exp\left[-\frac{(x_j-x_{ji})^2}{2*h^2}\right].$

During the training of the NB model, the AUC value of each SDI is calculated for each sample in the training set, and the normalized standard deviation of it is represented by w, which will be used for the weight corresponding to the SDI in the integration algorithm in Section 4.3.

4.3. Identifying FISDI Based on AdaBoost Meta-Learning Strategy

In order to further improve the accuracy of the model and avoid the problem of algorithm performance degradation caused by the random combination of SDIs, FILPA introduces the AdaBoost meta-learning strategy to identify FISDI composed of the best SDIs that are selected by NB. Because the process of new user link prediction must tackle the problems of an extremely sparse local network structure, the accuracy of a single model to identify FISDI is not high. Through the AdaBoost meta-learning strategy, the weak recognition models with low accuracy can be enhanced to become strong recognition models with high accuracy. In the AdaBoost meta-learning strategy, Discriminant Analysis (DA) is used as the base classifier and linear regression is used as the meta classifier, multiple base classifiers are combined to improve the classification accuracy. On the basis of the learning results of the base classifier, the meta classifier is used for relearning how to obtain the final results, so that the low-level learning can be fully used in the high-level induction process. Figure 3 shows an example of the AdaBoost meta-learning strategy.



Figure 3. An example of AdaBoost meta-learning strategy.

4.3.1. Discriminates Factors of Fully Integrated Index

On the type of all indexes of similarity criterion, Canberra, Sum of Absolute Difference (SAD), Reciprocal of Absolute Value (RAV), and Max–min are the most representative. Therefore, this paper uses them to measure the similarity between SDIs; their Formulas are shown as (13)–(16), respectively. Where $sim_k(i, j)$ is the similarity scores of the node pair composed of node *i* and *j* and calculated according to index *k* in Section 3.2, symbol \land represents the smaller one selected from the two alternatives, and symbol \lor represents the selection of the larger one.

$$L_{AB}^{Canberra} = \sum_{i < i} \frac{|sim_A(i,j) - sim_B(i,j)|}{|sim_A(i,j) + sim_B(i,j)|}$$
(13)

$$L_{AB}^{SAD} = \sum_{i < j} |sim_A(i, j) - sim_B(i, j)|$$
(14)

$$L_{AB}^{RAV} = \frac{0.001}{\sum_{i < j} |sim_A(i,j) - sim_B(i,j)|}$$
(15)

$$L_{AB}^{Max-min} = \frac{\sum_{i < j} (sim_A(i,j) \land sim_B(i,j))}{\sum_{i < j} (sim_A(i,j) \lor sim_B(i,j))}$$
(16)

4.3.2. DA Base Classifier

In order to improve the efficiency of the algorithm, we limited the composite index to be composed of *Q* SDIs. The basic idea of the DA base classifier is to judge whether the *Q* SDIs can constitute a fully integrated index according to the similarity between two scores of the *Q* SDIs, which is calculated by Canberra, SAD, RAV, and Max–min.

Meta-learning is used in the process of training the DA base classifier, assuming that the total training sample set is $S = \{(a_i, b_i) | i = 1, 2, ..., n\}$, where a_i vector represents the similarity of any two of the *Q* SDIs calculated by the above four algorithms, $b_i \in Y = \{0, 1, 2, 3, 4\}$, where 0 indicates *Q* SDIs cannot form a fully integrated index, 1, 2, 3, and 4 mean that the combination way to form a fully integrated index, where 1 indicates that all indexes are additive, 2 represents the index is made up of a number of blocks, which is composed of the combination of addition and subtraction of three indexes, 3 means that the index is comprised of a number of blocks, which is composed of the combination of three indexes, and 4 expresses that one index is subtractive from other indexes. Based on meta-learning, the weight *w* calculated in Section 4.2 is used to combine the *Q* SDIs with above different linear combinations for comprehensively mining characteristics of the implied social distance, y_i is the label corresponding to the fully integrated index with the largest AUC value.

The base DA base classifier is denoted as h(u, v), where u is the similarity between algorithm scores, v is the category label, (namely b_i), and its output value is the probability that u belongs to class v. Suppose that the i-th training input sample is (u_i, v_i) , p represents other classes except v_i , and we also define operator $\langle r \rangle$, when r is true, $\langle r \rangle = 1$, when r is false, $\langle r \rangle = 0$. When $p \neq v_i$, DA makes three judgments on the sample: $u_i \in v_i$ or $u_i \in p$. There are three situations when u_i is judged and classified: (1) When $\langle u_i, v_i \rangle = 0$ and $\langle u_i, p \rangle = 1$, then $u_i \notin v_i$; (2) When $\langle u_i, v_i \rangle = 1$ and $\langle u_i, p \rangle = 0$, then $u_i \in v_i$; (3) When $\langle u_i, v_i \rangle = \langle u_i, p \rangle$, the possibility of $u_i \in v_i$ is the same as $u_i \in p$, then choose one of them at random. Therefore, the probability that u_i is wrongly classified as p is shown in Formula (17) [36].

$$\frac{1}{2}(1 - h(u_i, v_i) + h(u_i, p))$$
(17)

For the above five kinds of problems, there are four different kinds of p, and because each different p may have different importance in different situations, each p is given a specified weight q(i, p), $(\sum_{p \neq v_i} q_t(i, p) = 1)$. Therefore, Formula (17) is modified to Formula (18).

$$\frac{1}{2} \left(1 - h(u_i, v_i) + \sum_{p \neq v_i} q_t(i, p) * h(u_i, p) \right)$$
(18)

4.3.3. AdaBoost Framework

The AdaBoost meta-learning strategy shares outstanding performance in a multiclassification problem, so it is selected to identify FISDI composed of the best SDIs. According to Formula (18), its pseudo-error can be expressed as Formula (19).

$$\varepsilon_t = \frac{1}{2} \sum_{i=1}^n D_t(i) * \left(1 - h_t(u_i, v_i) + \sum_{p=0(p \neq v_i)}^4 q_t(i, p) * h_t(u_i, p) \right)$$
(19)

where $D_t(i)$ represents the weight of the i-th sample, and the larger its value, the more likely the i-th sample is to be misjudged. The label weighting function $q_t(i, p)$ indicates the

probability of classifying u_i into class p wrongly. The larger its value, the more easily the sample can be misclassified, which needs to be examined in the next iteration of learning. $q_t(i, p)$ changes with multiple iterations, so as to get the final global classification model and achieve a better classification effect.

The main steps of the AdaBoost meta-learning strategy proposed in this paper are as follows:

Step 1: Generate the raw data S. For each sample in the training network set, the optimal indexes identified by the NB model are firstly eliminated from all L indexes, and then the remaining L-1 indexes are combined in pairs to form a composite index with the optimal indexes, respectively, and the w calculated in Section 4.2 is taken as the weight of the corresponding SDI. For each group of indicators, the similarity between two scores is calculated according to Canberra, SAD, RAV and Max-min, and the label is judged based on AUC value;

Step 2: Input. The total training sample set $S = \{(a_i, b_i) | i = 1, 2, ..., n\}$, and the number of iterations is T = 100. In each iteration, samples with the size of m * n are selected according to the sample distribution weight D obtained from the previous iteration, where $m \in (0,1)$ represents the proportion of selected samples. This algorithm ranks the weight vectors of sample distribution in descending order and selects the first m * n samples in total;

Step 3: Initialize variables. Let $D_1(i) = 1/n$, the weight of an error label *p* in the i-th

sample is $\omega_{i,p}^1 = \frac{D_1(i)}{4}$, where i = 1, 2, ..., n and $p \in \{Y - y_i\}$; Step 4: At iteration T, generate T DA based classifiers. Cycle the following steps at the t-th iteration ($t = 1, 2, \ldots, T$):

a. Calculate the label weight according to Formula (20) and compute the sample distribution weight of the i-th sample based on Formula (21);

$$q_t(i,p) = \frac{\omega_{i,p}^t}{\sum_{p=0(p \neq y_i)}^4 \omega_{i,p}^t} (p \neq y_i)$$
(20)

$$D_t(i) = \frac{\sum_{p=0(p \neq y_i)}^4 \omega_{i,p}^t}{\sum_{i=1}^n \left(\sum_{p=0(p \neq y_i)}^4 \omega_{i,p}^t \right)}$$
(21)

b. According to the new sample set S_i obtained from the sample distribution $D_t(i)$, DA is trained and to obtain the classifier $h_t(u, v)$;

c. The pseudo-error ε_t of h_t is calculated according to Formula (19), if $\varepsilon_t \ge 0.5$, then jump to Step 5;

d. Calculate the proportion $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$ of the current base classifier and update the weight vector, as shown in Formula (22).

$$\omega_{i,p}^{t+1} = \omega_{i,p}^{t} * \beta_{t}^{\frac{1}{2}(1-h(u_{i},v_{i})+h(u_{i},p))}$$
(22)

Step 5: At the end of iteration T, the base classifiers $h_t(u, v)$ are linearly combined with different weights to get the final meta classifier H(u, v), as shown in Formula (23). H(u, v) is used to test the test samples. According to the similarity between the scores of two indexes calculated by Canberra, SAD, RAV, and Max-min, the fully integrated index composed of the most suitable SDIs selected by NB is distinguished. The final classification results are obtained by weighted voting rules, as shown in Formula (24).

$$H(u,v) = \sum_{t=1}^{T} \left(\log \frac{1}{\beta_t} \right) * h_t(u,v)$$
(23)

$$g = \arg\max_{v \in Y} \{H(u, v)\}$$
(24)

4.4. High-Performance FISDI

FILPA further filters out high-precision FISDI to overcome the deficiency of available information faced by new users link prediction, and comprehensively describes the possibility of establishing links between users. Supposing that the AdaBoost meta-learning strategy selects *k* fully integrated indexes composed of the most suitable SDIs (expressed as *B*) which are identified by the NB model (i.e., $E_1(B), E_2(B), \ldots, E_k(B)$), FILPA selects the algorithm with the highest accuracy from the *k* fully integrated indexes, as shown in Formula (25).

r

$$= \max_{1 \le i \le k} \{AUC(E_i(B))\}$$
(25)

5. Experiments and Results Analysis

5.1. Experimental Design

The validity of FILPA was verified by 971 ego-net data sets of Twitter provided by Stanford University. Ego-net is a social network composed of users and their friends. Each ego-net represents a brand community, where users who like the same kind of products will form a circle. The data sets divide community members into different circles according to products they favorite. 224 networks with product circles were selected from Twitter. In each experiment, 180 networks were randomly selected from 224 networks as training samples and the remaining 44 networks were used as test samples. All experiments in this paper were implemented by MATLAB software. If there is no special statement, the default values of the parameters in the software are set for each algorithm.

Table 2 shows the average, minimum, and maximum statistical characteristics of 224 network samples in Twitter. In the simulation, the data were preprocessed. To be specific, the nodes with few links were screened out from all the nodes, and they were approximately defined as new users.

		Twitter	
Statistical Indicators –	Minimum	Mean	Maximum
Number of nodes	41	116.13	205
Number of edges	59	293.86	637
Average degree	2.48	4.89	8.48
Average shortest path	2.47	3.28	4.15
Efficiency of network	0.28	0.34	0.42
Average node betweenness	78.34	264.16	471.53
Average edge betweenness	31.67	78.68	138.67
Average clustering coefficient	0	0.11	0.38
Assortativity coefficient	-0.77	-0.27	0.11
Degree heterogeneity	5.04	15.51	36.28
Network density	0.03	0.04	0.08
Centrality of mean eigenvector	0.28	0.53	0.86
Centrality of mean proximity	0.25	1.21	8.26
Scale of mean connected group	20.50	91.98	205

Table 2. Statistical information of network samples in Twitter.

Table 3 shows the indexes with specific parameter values. A large number of experimental results show that the algorithm can achieve high precision by setting the parameter value in this way.

_						
	Algorithm Name	Parameters	Algorithm Name	Parameters	Algorithm Name	Parameters
	CCND2a	$\alpha = 3 \beta = 8$	CCND4d	$\delta = 3 \varepsilon = 3 \theta = 3 \alpha = 8$ $\beta = 8$	CCND4k	$\delta = 3 \varepsilon = 8 \theta = 3 \alpha = 8$ $\beta = 3$
	CCND2b	$\alpha = 3 \beta = 3$	CCND4e	$\delta = 8 \varepsilon = 3 \theta = 3 \alpha = 3$ $\beta = 8$	CCND41	$\delta = 3 \varepsilon = 8 \theta = 3 \alpha = 8$ $\beta = 8$
	CCND2c	$\alpha = 8 \beta = 3$	CCND4f	$\delta = 8 \varepsilon = 3 \theta = 3 \alpha = 3$ $\beta = 3$	CCND4m	$\delta = 3 \varepsilon = 3 \theta = 8 \alpha = 3$ $\beta = 8$
	CCND2d	$\alpha = 8 \beta = 8$	CCND4g	$\delta = 8 \varepsilon = 3 \theta = 3 \alpha = 8$ $\beta = 3$	CCND4n	$\delta = 3 \varepsilon = 3 \theta = 8 \alpha = 3$ $\beta = 3$
	CCND4a	$\delta=3\;\theta=3\;\alpha=3\;\beta=\!8$	CCND4h	$\delta = 8 \varepsilon = 3 \theta = 3 \alpha = 8$ $\beta = 8$	CCND40	$\delta = 3 \varepsilon = 3 \theta = 8 \alpha = 8$ $\beta = 3$
	CCND4b	$\delta = 3 \varepsilon = 3 \theta = 3 \alpha = 3$ $\beta = 3$	CCND4i	$\delta = 3 \varepsilon = 8 \theta = 3 \alpha = 3$ $\beta = 8$	CCND4p	$\delta = 3 \varepsilon = 3 \theta = 8 \alpha = 8$ $\beta = 8$
	CCND4c	$\delta = 3 \varepsilon = 3 \theta = 3 \alpha = 8$ $\beta = 3$	CCND4j	$\delta = 3 \varepsilon = 8 \theta = 3 \alpha = 3$ $\beta = 3$		·

Table 3. Algorithm abbreviation with parameters.

In NB experiments, based on the training set, 27 kinds of SDIs were screened and 14 kinds of SDIs with high performance were selected, which were CND3, CND4, CCND2a, CCND2c, CCND2d, CCND4a, CCND4b, CCND4c, CCND4e, CCND4g, CCND4j, CCND4l, CCND4m, and CCND4n, respectively. For each of the selected SDI, the AdaBoost metalearning strategy was trained by the training set, and then the fully integrated indexes were found. These 14 kinds of algorithms correspond to 14 AdaBoost meta-learning strategies.

In addition, in order to verify the performance of the proposed FILPA model, in the experiment, 10 kinds of classical SLPAs were adopted to compare with CN, Salton, Jaccard, Sorenson, HPI, HDI, LHN, AA, RA, and RAA.

In order to evaluate the predictive performance of the algorithm, 100 experiments were carried out, and the average AUC value was taken as the index to evaluate the performance of the algorithm. The results are shown in Table 4, where FILPAa represents FILPA with different weight combinations, FILPAb represents FILPA without weight combinations, that is, all the weights are 1, and single NB means the direct use of the FISDI selected by the NB for prediction.

Table 4. The performance of all algorithms ($\theta = 1$).

Algorithm name	Average AUC	Algorithm Name	Average AUC	Algorithm Name	Average AUC
CND0	0.6490383	CCND3	0.6504578	CCND4j	0.6507415
CND1	0.6497224	CCND4a	0.6503807	CCND4k	0.6513481
CND2	0.6499278	CCND4b	0.6507177	CCND41	0.6507386
CND3	0.6497900	CCND4c	0.6512201	CCND4m	0.6504118
CND4	0.65056450	CCND4d	0.6507350	CCND4n	0.6506732
CCND1	0.65004691	CCND4e	0.6504217	CCND40	0.6510824
CCND2a	0.65005695	CCND4f	0.6507004	CCND4p	0.6507198
CCND2b	0.65052006	CCND4g	0.6511380	Single NB	0.6513114
CCND2c	0.65127149	CCND4h	0.6507184	FILPAa	0.6647313
CCND2d	0.65054185	CCND4i	0.6502706	FILPAb	0.6647088

Figure 4 shows the performance comparison between single NB and non-combined SDIs. Figure 5 and Table 5 present the performance comparison between FILPA and all reference methods, and Figure 6 and Table 6 demonstrate the performance comparison between original SLPA and new SDIs



Figure 4. Performance comparison of single NB and non-combined SDIs.



Figure 5. Performance comparison of FILPA and all reference methods.

Table 5. Performance comparison of FILPA and all reference methods.

Algorithm Name	Average AUC	Algorithm Name	Average AUC
FILPAa	0.6647313	HDI	0.6452804
CN	0.6429730	LHN	0.6437872
Salton	0.6452023	AA	0.6500761
Jaccard	0.6451619	RA	0.6501544
Sorenson HPI	0.6451657 0.6439117	RAA	0.6521583

5.2. Performance Analysis of Algorithms

Table 5 and Figure 5 show that FILPA demonstrates a significant improvement over the existing excellent algorithms. In other words, FILPA can accurately predict the likelihood that new users will establish friendships with other nodes in the community when marketing the brand products to new users.

It can be seen from Table 4 that the performance of FILPAa is better than that of unweighted FILPAb, which indicates that the link prediction effect can be improved by

constructing a high-performance fully integrated index with different weights. It can also be seen from Table 4 that the performance of FILPAa and FILPAb is better than that of NB and other algorithms, which shows that the mechanism of selecting FISDI through the AdaBoost meta-learning strategy and further identifying high-performance FISDI can provide better prediction results than the blind combination of SDIs.



Figure 6. Performance comparison of newly defined SDIs and original SLPA.

Algorithm Name	Average AUC	Algorithm Name	Average AUC
CND0	0.6490383	CN	0.6429730
CND1	0.6497224	HPI	0.6439117
CND2	0.6499278	HDI	0.6452804
CND3	0.6497900	LHN	0.6437872
CND4	0.65056450	RA	0.6501544

Table 6. Performance comparison of newly defined SDIs and original SLPA.

It can also be seen from Figure 6 and Table 6 that the performance of the new indicator SDI developed by integrating the characteristic information of social distance into the original SLPA is significantly higher than that of the original SLPA itself, which suggests that using rich information such as the social distance of related nodes within a large range to predict the likelihood of social relationship generation can overcome the problem that algorithms based on a traditional local triadic closure structure cannot solve the problem of a sparse network due to the scarce information of local common neighbors of new users, and cold start. In addition, it can be concluded from Figure 4 that the accuracy of NB is higher than other non-combined SDIs, which means that the mechanism of selecting the best SDIs based on the density and social distance characteristics of common neighbors in a local network is effective.

Finally, paired samples of AUC of any two algorithms in 100 experiments were collected, and Z-test was used to determine whether the mean values of the two samples were statistically equal. Subsequently, it is found that all P values are less than 5% of the significance level, which indicates that there is a significant difference between the algorithms. These results show that FILPA can effectively overcome the cold start problem faced by new users in the prediction of their friendships, accurately recommend other nodes in the community to new users, and effectively realize brand product marketing to new users.

6. Conclusions

By leveraging the influence of a friend relationship between new users and existing users in a brand community, new users will be attracted to join that community to purchase the brand's products, which will be of benefit to realizing the "products rapidly attracting users" by expanding the scale of the brand community. Therefore, this paper proposes FILPA for predicting possible friendships between new users and existing users in a brand community.

The traditional link prediction algorithm derived from a local triadic closure structure cannot overcome the problems of sparse network and cold start caused by the scarcity of information related to new users' local common neighbors. Consequently, this paper proposes FILPA as a method for solving these problems by measuring the social distance of relevant nodes in an expanse region. Compared with the existing algorithms, the distinctive characteristics of this algorithm are as follows:

Firstly, the traditional link prediction method based on triadic closure structure only focuses on local information, while the SDI proposed in this paper takes the global information into account, which enriches the available information when predicting the connection between new users and existing users in the brand community.

Secondly, the algorithm adopts a naive Bayes (NB) method to adaptively select the appropriate SDI according to network density and social distance. In contrast with the traditional fixed link prediction algorithm, it can not only choose the appropriate SDI in different network structures more efficiently, but can also improve the prediction accuracy.

Thirdly, the existing single index prediction method can only describe the characteristics of social distance from one perspective, and blind combination of single index can not guarantee the comprehensive characterization of network information. In order to adequately describe the characteristics of implicit social distance, FILPA applies an AdaBoost meta-learning strategy to identify the FISDI composed of the best SDI screened by NB, and then explores the possibility of users becoming friends.

Twitter data was used to verify the performance of FILPA. The experiment results on 224 communities in the Twitter network show that the five kinds of SDIs in micro network structure are better than the method of using the original SLPA, and macro-based SDI has significantly high accuracy in contrast to an SLPA based on triadic closure structure. In addition, compared with the existing 10 classical algorithms, FILPA has obviously superior results. Therefore, FILPA can overcome the cold start problem faced by predicting the friendships of new users. By predicting the possibility of users becoming friends through FILPA, marketers can accurately recommend friends to new users in social networks and realize brand product marketing to new users, which will assist brand communities in quickly growing their customer bases, promote word-of-mouth, increase user stickiness, and realize successful product marketing in the embryonic period when the number of product users is still small. Additionally, the cost of marketing can be reduced by the way of attracting customers recommended by friends.

This paper has some limitations. First, it only takes link prediction in a static network into consideration, and we will consider link prediction in a dynamic network in the future. Second, this paper only considers the social relations of nodes in homogeneous networks, and does not consider the real-time interactive information in the network. Both social contact and real-time interactive information of the network will be considered in the future when predicting friendships in the networks.

Author Contributions: Conceptualization, S.L.; methodology, S.L., Z.W. (Ziming Wang) and B.Z. (Boyi Zhu); software, Z.W. (Ziming Wang) and Z.Y.; writing—original draft preparation, B.Z. (Boyi Zhu) and Z.W. (Zhifang Wen); writing—review and editing, B.Z. (Beiyan Zhang), Z.W. (Zhifang Wen); supervision, S.L. and Z.Y.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Chinese National Natural Science Foundation (No. 71871135).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to thank Xuewei Song of Shanghai University who provided technical support and important insights for this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Muniz, A.M.; O'guinn, T.C. Brand community. J. Consum. Res. 2001, 27, 412–432. [CrossRef]
- Kim, Y.A.; Ahmad, M.A. Trust, distrust and lack of confidence of users in online social media-sharing communities. *Knowl.-Based Syst.* 2013, 37, 438–450. [CrossRef]
- 3. Liang, T.P.; Ho, Y.T.; Li, Y.W.; Turban, E. What drives social commerce: The role of social support and relationship quality. *Int. J. Electron. Commer.* **2011**, *16*, 69–90. [CrossRef]
- 4. Shah, P.P.; Dirks, K.T.; Chervany, N. The multiple pathways of high performing groups: The interaction of social networks and group processes. *J. Organ. Behav.* 2006, 27, 299–317. [CrossRef]
- Rapp, A.; Beitelspacher, L.S.; Grewal, D.; Hughes, D.E. Understanding social media effects across seller, retailer, and consumer interactions. J. Acad. Mark. Sci. 2013, 41, 547–566. [CrossRef]
- Chang, H.H.; Chuang, S.S. Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. *Inf. Manag.* 2011, 48, 9–18. [CrossRef]
- Wang, X.; Yu, C.; Wei, Y. Social media peer communication and impacts on purchase intentions: A consumer socialization framework. *J. Interact. Mark.* 2012, 26, 198–208. [CrossRef]
- Xie, J.; Szymanski, B.K.; Liu, X. Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In Proceedings of the IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, Canada, 11 December 2011; pp. 344–349. [CrossRef]
- 9. Chen, D.; Lü, L.; Shang, M.S.; Zhang, Y.C.; Zhou, T. Identifying influential nodes in complex networks. *Phys. A: Stat. Mech. Its Appl.* **2012**, 391, 1777–1787. [CrossRef]
- 10. Gomes, H.M.; Barddal, J.P.; Enembreck, F.; Bifet, A. A survey on ensemble learning for data stream classification. *ACM Comput. Surv.* (*CSUR*) 2017, 50, 1–36. [CrossRef]
- 11. Aziz, F.; Gul, H.; Uddin, I.; Gkoutos, G.V. Path-based extensions of local link prediction methods for complex networks. *Sci. Rep.* **2020**, *10*, 1–11. [CrossRef]
- 12. Walker, S.K. Connected: The surprising power of our social networks and how they shape our lives. *J. Fam. Theory Rev.* 2011, *3*, 220–224. [CrossRef]
- 13. Martínez, V.; Berzal, F.; Cubero, J.C. A survey of link prediction in complex networks. *ACM Comput. Surv.* (*CSUR*) **2016**, 49, 1–33. [CrossRef]
- 14. Fan, X.; Chen, Z.; Cai, F.; Wu, J.; Liu, S.; Liao, Z.; Liao, Z. Local core members aided community structure detection. *Mob. Netw. Appl.* **2019**, *24*, 1373–1381. [CrossRef]
- Han, Q.; Wen, H.; Wu, J.; Ren, M. Rumor spreading and security monitoring in complex networks. *Int. Conf. Comput. Soc. Netw.* 2015, 9197, 48–59. [CrossRef]
- 16. Lü, L.; Zhou, T. Link prediction in complex networks: A survey. Phys. A Stat. Mech. Its Appl. 2011, 390, 1150–1170. [CrossRef]
- 17. Li, S.; Song, X.; Lu, H.; Zeng, L.; Shi, M.; Liu, F. Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm. *Expert Syst. Appl.* **2020**, *139*, 112839. [CrossRef]
- 18. Katz, L. A new status index derived from sociometric analysis. Psychometrika 1953, 18, 39–43. [CrossRef]
- 19. Rafiee, S.; Salavati, C.; Abdollahpouri, A. CNDP: Link prediction based on common neighbors degree penalization. *Phys. A: Stat. Mech. Its Appl.* **2020**, *539*, 122950. [CrossRef]
- 20. Zareie, A.; Sakellariou, R. Similarity-based link prediction in social networks using latent relationships between the users. *Sci. Rep.* **2020**, *10*, 1–11. [CrossRef]
- Li, L.; Liu, X.; Chen, N.; Tian, H. Link prediction based on node centrality. Proc. Int. Conf. Inf. Technol. Electr. Eng. 2018, 17, 1–6. [CrossRef]
- Zhou, M.; Jin, H.; Wu, Q.; Xie, H.; Han, Q. Betweenness centrality-based community adaptive network representation for link prediction. *Appl. Intell.* 2022, 52, 3545–3558. [CrossRef]
- 23. Iqbal, M.M.; Latha, K. An effective community-based link prediction model for improving accuracy in social networks. *J. Intell. Fuzzy Syst.* **2022**, *42*, 2695–2711. [CrossRef]
- 24. Lu, M.; Wei, X.; Ye, D.; Dai, Y. A unified link prediction framework for predicting arbitrary relations in heterogeneous academic networks. *IEEE Access* 2019, *7*, 124967–124987. [CrossRef]
- Ozcan, A.; Oguducu, S.G. Multivariate time series link prediction for evolving heterogeneous network. *Int. J. Inf. Technol. Decis. Mak.* 2019, 18, 241–286. [CrossRef]
- 26. Hristova, D.; Noulas, A.; Brown, C.; Musolesi, M.; Mascolo, C. A multilayer approach to multiplexity and link prediction in online geo-social networks. *EPJ Data Sci.* 2016, *5*, 24. [CrossRef]

- 27. Ozcan, A.; Oguducu, S.G. Link prediction in evolving heterogeneous networks using the NARX neural networks. *Knowl. Inf. Syst.* **2018**, *55*, 333–360. [CrossRef]
- 28. Lü, L.; Zhou, T. Link prediction in weighted networks: The role of weak ties. EPL (Europhys. Lett.) 2010, 89, 18001. [CrossRef]
- 29. Liben-Nowell, D.; Kleinberg, J. The link-prediction problem for social networks. J. Am. Soc. Inf. Sci. Technol. 2007, 58, 1019–1031. [CrossRef]
- 30. Wu, B.X.; Xiao, J. Potential friend recommendation based on user tagging. J. Comput. Appl. 2015, 35, 1663–1667. [CrossRef]
- 31. Chen, B.; Hua, Y.; Yuan, Y.; Jin, Y. Link Prediction on Directed Networks Based on AUC Optimization. *IEEE Access* 2018, 6, 28122–28136. [CrossRef]
- 32. Tichy, N.M.; Tushman, M.L.; Fombrun, C. Social network analysis for organizations. *Acad. Manag. Rev.* **1979**, *4*, 507–519. [CrossRef]
- 33. Marsden, P.V. The reliability of network density and composition measures. Soc. Netw. 1993, 15, 399-421. [CrossRef]
- 34. Bonacich, P. Some unique properties of eigenvector centrality. Soc. Netw. 2007, 29, 555–564. [CrossRef]
- 35. Albert, R.; Barabási, A.L. Statistical mechanics of complex networks. Rev. Mod. Phys. 2002, 74, 47. [CrossRef]
- Palazon, M.; Sicilia, M.; Lopez, M. The influence of "Facebook friends" on the intention to join brand pages. J. Prod. Brand Manag. 2015, 24, 580–595. [CrossRef]