mathematics

MDPI

Article

LPGRI: A Global Relevance-Based Link Prediction Approach for Multiplex Networks

Chunning Wang [1,*], Fengqin Tang [2] and Xuejing Zhao [1]

[1] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China
[2] School of Mathematics Sciences, Huaibei Normal University, Huaibei 235000, China
* Correspondence: leven_wong@163.com

Abstract: The individuals of real-world networks participate in various types of connections, each forming a layer in multiplex networks. Link prediction is an important problem in multiplex network analysis owing to its wide range of practical applications, such as mining drug targets, recommending friends in social networks, and exploring network evolution mechanisms. A key issue of link prediction within multiplex networks is how to estimate the likelihood of potential links in the predicted layer by leveraging both interlayer and intralayer information. Several studies have shown that incorporating interlayer topological information can improve the performance of link prediction in the predicted layer. Therefore, this paper proposes the Link Prediction based on Global Relevance of Interlayer (LPGRI) method to estimate the likelihood of potential links in the predicted layer of multiplex networks, which comprehensively utilizes both types of information. In the LPGRI method, the contribution of interlayer information is determined using the global relevance (GR) index between layers. Experimental studies on six real multiplex networks demonstrate the competitive performance of our method.

Keywords: complex network; link prediction; multiplex network; interlay relevance

MSC: 05C82

Citation: Wang, C.; Tang, F.; Zhao, X. LPGRI: A Global Relevance-Based Link Prediction Approach for Multiplex Networks. *Mathematics* **2023**, *11*, 3256. https://doi.org/10.3390/math11143256

Academic Editor: Andrea Scozzari

Received: 14 June 2023
Revised: 15 July 2023
Accepted: 23 July 2023
Published: 24 July 2023

1. Introduction

Networks are very useful tools for studying various real-world systems, such as computer communications, social systems, technological systems, and biological systems. In these networks, nodes represent individuals, and edges (or links) reflect the interactions between nodes. Traditionally, complex systems are structured as monolayer networks, where the interactions between nodes are of the same type. For example, in social networks, nodes represent individuals and edges indicate friendships; in biological networks, nodes may represent genes or proteins while edges represent regulatory relationships. However, recent studies [1–4] have shown that there exist different interactions among individuals within the same group. Such systems can be described as multiplex networks, where individuals interact with each other in different types. For example, transportation between cities may involve air and rail networks, while interpersonal communication may occur through mobile phones, email, and WeChat. Each type of connection between cities or individuals mentioned above can be considered a distinct layer within the corresponding multiplex networks. Multiplex networks are often employed to describe multiple relationships within a fixed group of individuals [5].

The study of networks has attracted significant attention in recent decades and the large number of studies in the literature demonstrates the development of this topic [6–10]. Among them, an important problem in network research is link prediction. The aim of link prediction is to estimate the connection probability of missing or forthcoming links by leveraging the observed information of networks. For a detailed description of

link prediction, we refer the interested reader to [10–12]. The problem of link prediction has drawn extensive attention from researchers in various disciplines. On the one hand, link prediction bears important theoretical significance. It can reveal the generation and evolution mechanism of a network structure. On the other hand, it has a wide range of applications, such as recommending friends in online social networks, identifying interacting genetic pairs in genetic networks, and analyzing signal propagation in complex networks [13], etc.

Link prediction methods for monolayer networks can be categorized into the following types. The first is based on various similarity measures. Most of these methods treat link prediction as a ranking problem, assigning similarity scores to pairs of unconnected nodes, where higher scores indicate a higher likelihood of an edge [14]. Similarity scores can be calculated using node features or network structure. Typical similarity measures leveraging network structure include local and global similarity indices. The Common Neighbor index [15], the Jaccard index [10], the Adamic–Adar index [16], the Resource Allocation index [17], and the weighted forms of these indices [18] are all local indices. Global indices include the Katz index [19], Local Path index [11], and SimRank [20]. For more metrics, please refer to reference [11]. The second type of approach considers the organizational principles of the network when predicting links. For example, the hierarchical structure model [21] and the stochastic block model [22]. Another type of link prediction approaches involves probabilistic models for incomplete networks, such as probabilistic relational models [23], the probabilistic entity relationship model [24] and the stochastic relational model [25]. Comprehensive reviews of these methods can be referred to [10,11]. Additionally, ensemble learning, matrix completion, and network embedding techniques are also utilized in link prediction research, interested readers can refer to the review literature [26,27].

Recently, several studies have engaged in the link prediction problem of multiplex networks. Chen et al. [28] proposed the SimBins method, which leverages the effect of overlapping links between layers to enhance the quality of link predictions. Davis et al. [29] introduced a novel measure, an extension of the probabilistically weighted Adamic–Adar index, for link prediction in multiplex networks. Desislava et al. [30] investigated the geo-social properties of multiplex links and proposed a variant of the Jaccard index as an extended neighborhood of nodes in multiplex networks to the problem of link prediction. Li et al. [31] introduced a new similarity measure based on local random walk measures on reliable paths and weighted networks. This method utilizes pure random walks to capture the network structure for computing similarity and discovering unknown links. Based on deep learning, Cao et al. [32] proposed a novel framework called LPSMN for link prediction in multiplex social networks. It leveraged the information from external layers to predict links in a monolayer network. The literature considers a two-layer social network and proposes a meta-path-based algorithm for link prediction. Jalili et al. [33] developed a meta-path-based algorithm incorporating three classical classifiers, i.e., the naive Bayes, the support vector machines and the k-nearest neighbor for predicting the links in multiplex networks. Their experimental results demonstrated that incorporating cross-layer information can significantly improve prediction performance. Bacco et al. [34] designed a probabilistic model and expectation maximization algorithms for link prediction in multiplex networks. Yang et al. [35] proposed a novel probabilistic method called Multi-Relational Influence Propagation for heterogeneous networks with multiple types of links. This method uses temporal-related features to improve link prediction performance. Different from the above method, Sharma et al. [36,37] proposed a method based on the structure similarity of multiplex networks. The likelihood of links in the predicted layer is assigned as a weighted combination of scores from other layers, with the weights determined by checking the link correspondence between the two layers. Yao et al. [38] combined intralayer and interlayer information for link prediction by calculating interlayer similarity. Najari [39] derived a more complex combination of similar index, where the utilization of interlayer information for link prediction in the predicted layer depends

on whether the node pairs have edges in other layers. These methods employed various approaches to measure the correlation between layers. However, they did not incorporate corresponding interlayer relevance measures based on the structural characteristics of directed multiplex networks. Additionally, they did not fully exploit intralayer information when making link predictions. For instance, they relied on similarity-based methods such as CN, AA, etc., to measure the similarity between nodes in the same layer. As is well known, a limitation of similarity-based approaches is that they often trade-off accuracy for computational efficiency.

Traditional link prediction methods typically consider only the topological information derived from single-layer networks, referred to as intralayer information, while disregarding the additional information originating from other layers in multiplex networks, known as interlayer information. Therefore, a key issue in link prediction within multiplex networks is how to effectively leverage both intralayer and interlayer information. To this end, we aim to discover an effective method that integrates intralayer and interlayer information to enhance the accuracy of link prediction. In this paper, we introduce the Link Prediction based on Global Relevance of Interlayer (LPGRI) method to estimate the likelihood of potential links in the predicted layer of multiplex networks. The LPGRI method effectively leverages interlayer information by utilizing the global relevance (GR) index between layers. The steps to achieve the main contribution of this paper are as follows. First, we propose a novel index called the GR index to measure the correlation between layers in undirected and directed multiplex networks. The GR index is obtained by measuring the average similarity of the connection modes between corresponding nodes in different layers. It reflects the correlation between the two layers of the multiplex network from the perspective of connection patterns. Specifically, for directed multiplex networks, the GR index can measure the correlation between layers according to their corresponding structural characteristics. Second, we estimate the likelihood of potential links in predicted layers by effectively utilizing both intralayer and interlayer information. The LPGRI method combines the two types of information using the GR index for target layer prediction. The experimental results show that incorporating information from other layers improves the prediction performance to some extent. Third, the experimental results also demonstrate that the LPGRI method can handle the cold-start problem and solve the problem of sparse networks. Due to the comprehensive utilization of information from other layers in the multiplex network, the LPGRI method achieves enhanced prediction accuracy, even in cases where the predicted layer contains limited valuable information.

The rest of this paper is organized as follows: In Section 2, we detail the framework of the LPGRI approach. The datasets and performance evaluation metric are introduced in Section 3. Experiment analysis is presented in Section 4. Lastly, concluding remarks are provided in Section 5.

2. Methodology

We consider a multiplex network with N nodes and L layers, denoted by $\boldsymbol{G} = (G^{(1)}, G^{(2)}, \cdots, G^{(L)})$. The notations used in this paper and their explanations are provided in Table 1. Here, the set of nodes is the same across all layers in a multiplex network, i.e., $|V^{(1)}| = |V^{(2)}| = \cdots = |V^{(L)}| = N$. Then, $G^{(\alpha)}$ can be written as $G^{(\alpha)} = (V, E^{(\alpha)})$, where $|V| = N$. The adjacency matrix $A^{(\alpha)} = (A_{ij}^{(\alpha)})_{N\times N}$ is defined as

$$A_{ij}^{(\alpha)} = \begin{cases} 1 & \text{if there is an edge from } i \text{ to } j \text{ in the layer } \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

Typically, the predicted layer in the multiplex network is referred to as the target layer and each of the remaining layers in the multiplex network is regarded as an auxiliary layer.

In a multiplex network, the task of link prediction is to estimate the likelihood of potential links between nodes that are currently unconnected in the target layer, utilizing the information from both the target layer itself and the other auxiliary layers. Previous research [40] showed that similarity in interlayer structural features can enhance link pre-

diction performance in multiplex networks using the layer reconstruction method and experimental analysis on real-world multiplex networks from various domains. Tang et al. [5] introduced a semi-supervised learning method that considers interlayer structural information to predict links in the target layer of a multiplex network. Experimental results on both simulated and real-world multiplex networks demonstrate that the proposed method outperforms prediction methods that solely rely on single-layer network information. These studies indicate that incorporating interlayer topological information can improve the performance of link prediction in the target layer. This implies that the connections between pairs of nodes in the target layer may be correlated with the connections between corresponding node pairs in other layers. Consequently, if the correlation between the target layer and the auxiliary layers can be accurately measured, the information from the auxiliary layers can be effectively utilized to improve the performance of link prediction in the target layer. Motivated by this, we propose a metric called the global relevance (GR) index to quantify the correlation between layers. A higher GR index between the target layer and auxiliary layer implies a stronger correlation in the network structure between the two layers. This suggests that the auxiliary layer may provide more useful information for link prediction in the target layer. Specifically, for a given node pair i and j in the target layer $G^{(T)}$, the likelihood of an edge existing between them is influenced by both the intralayer structure information of the target layer $G^{(T)}$ and the interlayer information from the auxiliary layers $G^{(l)}(l = 1, 2, \cdots, L, l \neq T)$. By utilizing the GR index, we can effectively incorporate the information from auxiliary layers into the prediction task. In the following sections, we will introduce the GR index first, followed by the presentation of the prediction method LPGRI for multiplex networks.

Table 1. Notation definitions.

Notations	Explanations
$V^{(\alpha)}$	the set of nodes in the α-th layer ($\alpha = 1, 2, \cdots, L$)
$E^{(\alpha)}$	the set of edges in the α-th layer
$G^{(\alpha)} = (V^{(\alpha)}, E^{(\alpha)})$	the layer α
$A^{(\alpha)} \in \mathbb{R}^{N \times N}$	the adjacency matrix of $G^{(\alpha)}$
$G^{(T)}$	the target layer
$A_{i\cdot}^{(\alpha)}$	the i-th row of the adjacency matrix of layer α
$A_{\cdot i}^{(\alpha)}$	the i-th column of the adjacency matrix of layer α
$\bar{A}_{i\cdot}^{(\alpha)}$	the mean value of $A_{i\cdot}^{(\alpha)}$
$\sigma(A_{i\cdot}^{(\alpha)})$	the standard deviation of $A_{i\cdot}^{(\alpha)}$
$P_{ij}^{(T)}$	probability of potential links between node pair (i, j) in the target layer

2.1. Measurement of Interlayer Correlation in Multiplex Networks

In monolayer network link prediction, Liao et al. [41] utilized the Pearson correlation coefficient as the similarity of a node pair. Specifically, the vectors $\boldsymbol{v}_x = (a_{x1}, a_{x2}, \cdots, a_{xn})$ and $\boldsymbol{v}_y = (a_{y1}, a_{y2}, \cdots, a_{yn})$ represent the feature vectors of node x and y, respectively. $\boldsymbol{v}_x$ can be set as $A_{x\cdot}$, which denotes the x-th row of the monolayer network adjacency matrix. The similarity between nodes x and y is then defined as the Pearson correlation coefficient of vectors $\boldsymbol{v}_x$ and $\boldsymbol{v}_y$. Inspired by this, we calculate the average correlation between corresponding nodes across interlayers to measure the relevance between layers in multiplex networks. Initially, the correlation between corresponding nodes in different layers is computed using the Pearson correlation coefficient, which reflects the degree of similarity in connection patterns between corresponding nodes across different layers. Subsequently, the average correlation of all nodes is utilized as the global relevance index between different layers, which reflects the overall similarity in connection patterns between the two layers. The GR index between undirected layers α and β is defined as

$$\mathrm{GR}_{\alpha\beta}^{un} = \frac{1}{N}\sum_{i=1}^{N}\frac{(A_{i\cdot}^{(\alpha)} - \bar{A}_{i\cdot}^{(\alpha)})(A_{i\cdot}^{(\beta)} - \bar{A}_{i\cdot}^{(\beta)})'}{\sigma(A_{i\cdot}^{(\alpha)})\sigma(A_{i\cdot}^{(\beta)})}, \tag{1}$$

and the GR index between directed layers α and β is

$$\mathrm{GR}_{\alpha\beta}^{dir} = \frac{1}{2N}\sum_{i=1}^{N}\left\{\frac{(A_{i\cdot}^{(\alpha)} - \bar{A}_{i\cdot}^{(\alpha)})(A_{i\cdot}^{(\beta)} - \bar{A}_{i\cdot}^{(\beta)})'}{\sigma(A_{i\cdot}^{(\alpha)})\sigma(A_{i\cdot}^{(\beta)})} + \frac{(A_{\cdot i}^{(\alpha)} - \bar{A}_{\cdot i}^{(\alpha)})'(A_{\cdot i}^{(\beta)} - \bar{A}_{\cdot i}^{(\beta)})}{\sigma(A_{\cdot i}^{(\alpha)})\sigma(A_{\cdot i}^{(\beta)})}\right\}, \tag{2}$$

where $A_{i\cdot}^{(\alpha)}$ and $A_{\cdot i}^{(\alpha)}$ represent the i-th row and column of the adjacency matrix of layer α. They reflect the connection pattern of node i and the connection pattern to node i in layer α, respectively.

Remark 1. *In the study presented in [41], the authors considered link prediction on the monolayer network. The Pearson correlation coefficient was used to calculate the similarity score of the features of two nodes. This similarity score was then utilized as a measure of similarity between the two nodes. Based on the ranking of these scores, link predictions can be made. However, the GR index is obtained by averaging the Pearson correlation coefficient of the connection patterns between corresponding nodes across two layers (i.e., calculating the Pearson correlation coefficient between $A_{i\cdot}^{(\alpha)}$ and $A_{i\cdot}^{(\beta)}$). The GR index defines interlayer correlations in both undirected and directed multiplex networks by considering the connection patterns between corresponding nodes across layers.*

2.2. Probability Estimation of Potential Links in Multiplex Networks

To estimate the probability of potential connections in the target layer by leveraging the information both from the target layer and auxiliary layers, we propose a method called Link Prediction based on Global Relevance of Interlayer (LPGRI) for link prediction in multiplex networks.

Similar to the study presented in [39], for the target layer, the ultimate probability of a potential link between the pair of nodes (i, j) is defined as

$$P_{ij} = (1-\lambda)P_{ij}^{(T)} + \lambda\sum_{\substack{l=1\\ l\neq T}}^{L}\mathrm{GR}_{lT}P_{ij}^{(l)}, \tag{3}$$

where $\lambda \in [0, 1]$ is the tunable parameter. It controls the amount of information provided by all auxiliary layers for the link prediction of the target layer. $P_{ij}^{(l)}$ represents the probability of potential links between node pair (i, j) in the auxiliary layer $G^{(l)}$. These probabilities are calculated using the NBS method [42], which has better predictive performance in monolayer networks. The ultimate probability of a potential link between the nodes (i, j) in the target layer is represented by P_{ij}. The index GR_{lT}, which represents the correlation between layers $G^{(l)}$ and $G^{(T)}$, can be viewed as the weight of information provided by each auxiliary layer $G^{(l)}$ for link prediction in the target layer. It is defined by Equation (1) for undirected layers and Equation (2) for directed layers.

Equation (3) reveals that the prediction outcome of the potential link probability in the target layer depends on the information derived from both the target layer and auxiliary layers. The first term represents the contribution of intralayer information obtained from the target layer, while the second term represents the contribution of interlayer information derived from all auxiliary layers. The extent of the utilization of these two types of information is controlled by the parameter λ. A higher λ value indicates that more information is provided by all auxiliary layers for the prediction, and vice versa. Specifically, when λ equals 0, only the information from the target layer is utilized in the LPGRI method. In this case, the LPGRI method degenerates to the prediction problem of a monolayer network. When λ equals 1, only the information from all auxiliary layers is employed in the proposed method. When λ takes a value between 0 and 1, information from both intralayer and

interlayer sources is combined. The tunable parameter λ needs to be optimized for each multiplex network.

Equation (3) is used to calculate the final connection probability in our paper. Similar formulas have also appeared in [38,39]. Compared to these studies, the LPGRI method has two advantages. First, the GR index in the LPGRI method measures interlayer relevance from the perspective of connection patterns and provides different measurement methods for undirected and directed networks. Second, the calculation of within-layer connection probabilities utilizes NBS instead of local metrics such as CN or AA. It is well known that although local metrics offer higher computational efficiency, they are limited by their lower calculation accuracy.

3. Data Description and Evaluation Metric

3.1. Datasets

To verify the validation of the proposed method, we conduct experiments on six real datasets from various fields, including biological networks, social networks, etc. The descriptions of these datasets are provided below.

- Lazega [43]: The dataset consists of a multiplex network representing corporate law partnerships among employees. It contains three kinds of relationships, namely advice, co-work, and friendship. There are a total of 71 nodes, with the number of active nodes in each layer being 71, 70, and 69, respectively.
- C. elegans [44]: The Caenorhabditis elegans dataset comprises three layers that correspond to different synaptic junctions, namely electric links, chemical monadic links, and chemical polyadic links. The multiplex network consists of 279 nodes in total. Each layer has a different number of active nodes, with 253, 260, and 278 nodes, respectively.
- Kapferer [45]: The Kapferer tailor shop dataset describes the interactions in a tailor shop in Zambia over a period of ten months. The four layers of the network are generated by two different types of interaction. The first two layers, TS1 and TS2, represent "sociational" interactions, specifically friendship and socioemotional connections. The last two layers, TI1 and TI2, record "instrumental" interactions related to work and assistance at two different time points. The multiplex network consists of 39 nodes, with the number of active nodes in each layer being 39, 39, 35, and 37, respectively.
- Vicker [46]: This dataset is a multiplex social network depicting the relationships between 29 Grade 7 students in a school in Victoria, Australia. It consists of three layers, with each layer corresponding to different types of relationships, namely getting on, best friends, and preferring working together. There are 29 nodes in total. The number of active nodes in each layer is also 29.
- CKM [47]: This dataset constitutes a multiplex network that captures the interaction among physicians during the adoption of a new drug. It consists of three layers that represent different types of relationships: friendship, discussion, and asking for advice. There are 245 nodes in total. The number of active nodes in each layer is 215, 231, and 227, respectively.
- Rattus [48]: This dataset provides a multiplex network of genetic and protein interactions in Rattus Norvegicus. The raw data comprise 2640 nodes and 6 layers. In order to remove uninformative layers, we exclude those with only a few dozen edges. Within the paper, the multiplex network is analyzed using three layers: physical association, direct interaction, and colocalization. There are a total of 2538 nodes, with 1948, 979, and 149 active nodes in each layer, respectively.

Among the datasets, the first three datasets are undirected multiplex networks and the others are directed multiplex networks. Table 2 provides a summary of the basic topological features of these multiplex networks. In each multiple network, L represents the number of layers, N denotes the number of nodes, $N^{(\alpha)}$ indicates the number of active nodes (A node is considered active in a layer if it has at least one link in this layer.) in each layer, $k^{(\alpha)}$ represents the layer sequence number of the layer α and $E^{(\alpha)}$ represents the number

of edges in the layer. Additionally, D represents the layer density, which is calculated as $\frac{2E^{(\alpha)}}{N(N-1)}$ for undirected networks and $\frac{E^{(\alpha)}}{N(N-1)}$ for directed networks.

Table 2. Summarized statistics of the six multiplex networks.

Network Type	Dataset	L	N	$N^{(\alpha)}$	$k^{(\alpha)}$	Layer Name	$E^{(\alpha)}$	D
undirected	Lazega	3	71	71	1	advice	717	0.298
				70	2	work	378	0.152
				69	3	friend	399	0.161
	C. elegans	3	279	253	1	electric	514	0.013
				260	2	mono	888	0.023
				278	3	poly	1703	0.044
	Kapferer	4	39	39	1	TS1	158	0.213
				39	2	TS2	223	0.301
				35	3	TI1	76	0.103
				37	4	TI2	95	0.128
directed	Vicker	3	29	29	1	get on	361	0.445
				29	2	friend	181	0.223
				29	3	co-work	198	0.244
	CKM	3	245	215	1	advice	480	0.008
				231	2	discussion	565	0.009
				227	3	friend	504	0.008
	Rattus	3	2538	1948	1	physical association	2894	0.00052
				979	2	direct interaction	1024	0.00018
				149	3	colocalization	119	0.00002

3.2. Evaluation Metric

Considering a target layer $G^{(T)} = (V, E^{(T)})$, we define $U^{(T)}$ as the universal set, which contains all $N(N-1)/2$ edges in the target layer. To test the accuracy of the proposed method, the edge set $E^{(T)}$ is randomly divided into two sets: the training set $E_t^{(T)}$, which is treated as known information, and the test set $E_p^{(T)}$, which is used for testing and considered to contain no information for prediction. Obviously, $E_t^{(T)} \cup E_p^{(T)} = E^{(T)}$ and $E_t^{(T)} \cap E_p^{(T)} = \emptyset$. In our experiments, the proportion of edges contained in the probe set and the training set is set to d and $1-d$, respectively. This means we randomly select a certain proportion of the observed edges, denoted as $1-d$, as the training set, and the remaining proportion of the observed edges, d, as the test set in a layer.

Considering that the area under the receiver operating characteristic curve (AUC) [49] provides a comprehensive assessment of algorithm performance, is not sensitive to the balance of the dataset, and has a relatively simple and interpretable nature, we adopt it as a metric to evaluate the accuracy of link prediction approaches in this paper. It evaluates the performance of a method based on the ranking of predicted links. Given the rank of all non-observed links, the AUC can be interpreted as the probability that a randomly chosen missing link from $E_p^{(T)}$ is assigned a higher score (In this paper, the score refers to the probability of a link.) than a randomly chosen nonexistent link from $U^{(T)} - E^{(T)}$. If among n independent comparisons, n_1 times the missing link has a high score and n_2 times is the same, the calculation of AUC can be written as follows:

$$\text{AUC} = \frac{n_1 + 0.5n_2}{n}. \tag{4}$$

AUC is one of the most widely used indices for measuring link prediction accuracy. If all scores are generated from an independent and identical distribution, the AUC value is approximately 0.5. In practice, AUC values range from 0.5 to 1, with larger values indicating higher prediction accuracy.

4. Experimental Analysis

In this section, we begin by employing the GR index to calculate the interlayer correlation within each multiplex network. Secondly, we discuss the influence of the tunable parameter λ on the predictive performance of the LPGRI method. Next, we study the different contribution of each auxiliary layer to the prediction performance of the target layer in the multiplex networks. Finally, we conduct experiments to evaluate the performance of the LPGRI method in comparison to other competitive methods. All experiments are conducted using R 4.2.2.

4.1. Correlation Between Layers in Real Datasets

In this section, we analyze the correlation between layers in each multiplex network. Figure 1 shows the correlation between layers in the multiplex networks, as calculated by the GR index. Compared with other multiplex networks, the Vicker and Rattus networks exhibit a higher correlation between their layers. The values of GR between layers range from 0.513 to 0.705 for the Vicker network and 0.429 to 0.756 for the Rattus network, excluding autocorrelation among layers. This indicates that each pair of layers within these two multiplex networks share considerable similarities. According to Equation (3), the auxiliary layer that is more similar to the target layer will provide more information for link prediction. As for the remaining multiplex networks, the maximum GR value is around 0.5.

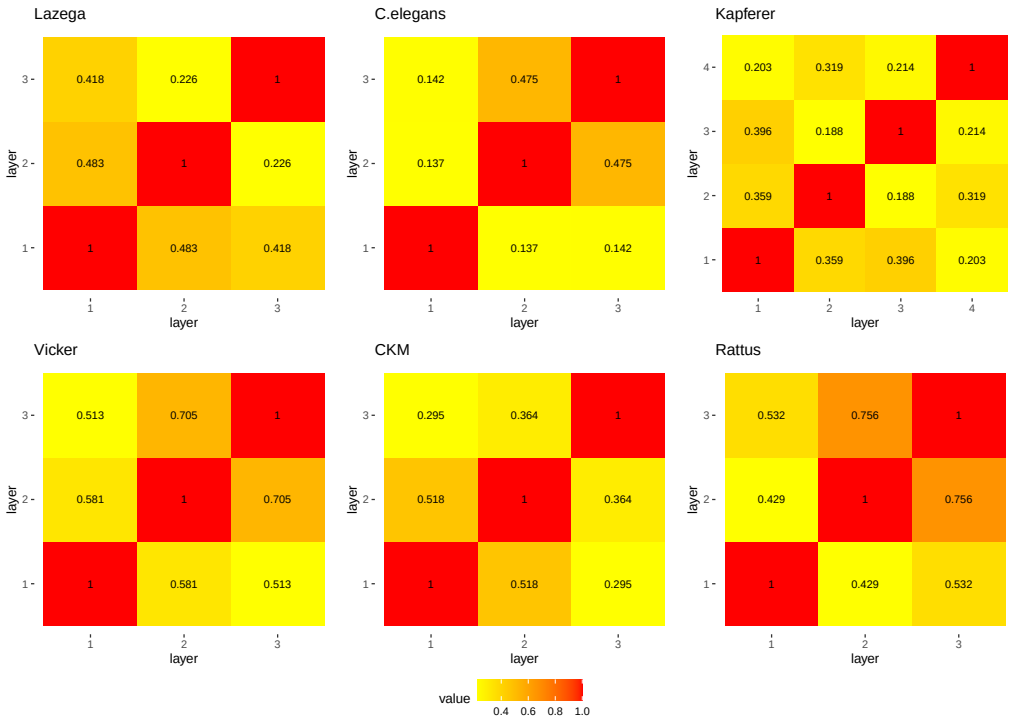

Figure 1. Layer relevance of six multiplex networks. The heat map illustrates the interlayer relevance within various multiplex networks. Each subfigure represents the outcome of a specific multiplex network, quantified by the GR index.

4.2. Tunable Parameter Analysis

As mentioned in Section 2.2, the tunable parameter λ needs to be optimized in the multiplex networks. Therefore, we now investigate how the parameter λ influences the prediction performance of the LPGRI method by varying the ratio of training sets from 90% to 50%, and select the optimal value of λ, denoted by λ^*, for each multiplex network shown in Table 2. In this experiment, we consider three different values for d, which represents the proportion of edges in the test set. Specifically, for each target layer, we set d to be 0.1, 0.3, and 0.5, which corresponds to training sets containing 90%, 70%, and 50% of the

edges, respectively. In each training set, the AUC values are calculated by the varying of parameter λ from 0 to 1. It is important to note that when calculating the GR_{lT} value, we incorporate all available information from the auxiliary layer $G^{(l)}$ and the training set information from the target layer $G^{(T)}$. Figure 2 illustrates the prediction performance variation of the LPGRI method with changes in the λ parameter.

From Figure 2, we can observe that the information from the auxiliary layers is helpful for link prediction in the target layer. The performance of the LPGRI method exhibits similar variation characteristics across different multiplex networks as parameter λ changes. With the increase in λ, the AUC values of the LPGRI method first increase and then tend to plateau in each target layer. This indicates that a moderate utilization of auxiliary layer information can enhance the prediction performance. However, when λ is close to 1, the AUC values begin to decrease in most target layers. This may be due to the noise from the auxiliary layers affecting the prediction performance of the target layer. On the other hand, when $\lambda = 0$, i.e., only the target layer information is used, the AUC values are minimal in most cases. Notably, when $\lambda = 0.1$, indicating that a small amount of information from the auxiliary layers is added, the AUC values significantly improve the most in target layers. This result further illustrates that, in comparison to solely relying on target layer information, the incorporation of auxiliary layer information can enhance the prediction performance to a certain extent.

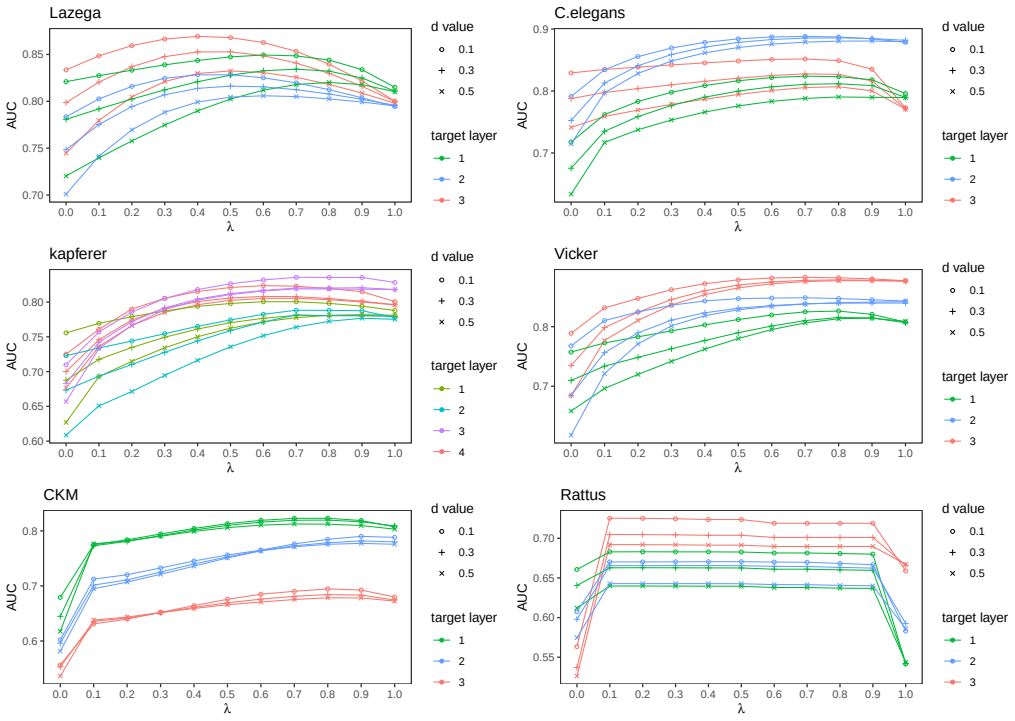

Figure 2. Variation in AUC values as a function of parameter λ in six multiplex networks. Each subfigure shows how the prediction performance of the LPGRI method changes with the parameter λ. Different colors of the curves in each subfigure represent the variation in AUC for different target layers. For the same target layer, three different values of d are considered, namely 0.1, 0.3, and 0.5. Each result is obtained by averaging 30 experiments.

Second, from Figure 2, we also observe that the AUC values corresponding to $\lambda = 1$ are consistently higher than the AUC values corresponding to $\lambda = 0$ in multiplex networks such as Kapferer, Vicker, and CKM. This finding is helpful for addressing the cold-start problem. For example, in a recommender system, it is difficult to obtain sufficient topology information to predict the potential connections for new members or commodities. In other words, the cold-start problem arises when valuable information regarding new or low-degree nodes is unavailable in a monolayer network. In comparison, multiplex networks

capture diverse relationships among individuals. Thus, the auxiliary layers based on the behavior or background of new individuals can assist in predicting the topology of these individuals in the target layer.

Third, the experimental results under different values of d show that the LPGRI method effectively addresses the network sparsity problem. Taking the CKM network as an example, when λ changes from 0 to 0.1, the AUC values of the first target layer increase by 14.08%, 19.88% and 25.71%, corresponding to a d value of 0.1, 0.3 and 0.5, respectively. Similar results are observed in other target layers of CKM. The reasons for this phenomenon are as follows. A larger value of d indicates that the training set contains less information for prediction. As mentioned before, $\lambda = 0$ implies that only the information from target layer is used for prediction, while when $\lambda = 0.1$, this indicates that in addition to the information from the target layer, the information from the auxiliary layers is also used for prediction. When the target layer contains limited useful information (i.e., d = 0.5), the likelihood of potential links is mostly assigned to 0 when only the target layer information is considered (i.e., $\lambda = 0$). It becomes challenging to distinguish between nonexistent links in $U^{(T)} - E^{(T)}$ and observed links in $E_p^{(T)}$. As a result, the prediction performance is poor. However, when the information from the auxiliary layers is simultaneously taken into account (i.e., $\lambda = 0.1$), the distinction between them becomes evident, leading to a significant improvement in prediction accuracy. In summary, the experimental results demonstrate that even in sparse networks where the target layer contains limited useful information (i.e., d = 0.5), the LPGRI method can still enhance prediction accuracy with the assistance of auxiliary layer information. This indicates that the LPGRI method can achieve improved prediction accuracy even in sparse networks. A similar phenomenon can be observed in Figure 2 for other multiplex networks.

Finally, we will discuss the selection of optimal values for λ in different multiplex networks. According to the results shown in Figure 2, it can be observed that the optimal values of λ for different layers within the same multiplex network are generally consistent. Specifically, the optimal values of λ are approximately 0.5 for the Lazega network, and 0.7 for the C.elegans, Kapferer, and Vicker networks. Furthermore, the CKM network exhibits optimal values of λ close to 0.8. For the Rattus network, when λ falls within the range of [0.1, 0.5], the AUC values show minimal variation. Based on the preceding discussion, the optimal value of λ, denoted as λ^*, varies across the multiplex networks analyzed in this paper. Specifically, in the Lazega network, λ^* is determined to be 0.5. For the C.elegans, Kapferer, and Vicker networks, λ^* is set to 0.7. In the CKM network, λ^* is determined to be 0.8, while for the Rattus network, λ^* is found to be 0.2.

4.3. Analysis of the Influence of Auxiliary Layers on Prediction Performance

In this section, we discuss the different contributions of each auxiliary layer to the prediction performance of the target layer in multiplex networks. Specifically, we study the GR index in the LPGRI method, which allows us to understand the distinct impacts of interlayer relevance on prediction performance. To uncover these effects, we analyze the changes in prediction performance by gradually eliminating the contributions of auxiliary layers. This elimination process is performed in descending order based on the GR values assigned to each auxiliary layer. To be specific, for a given target layer $G^{(T)}$, we remove the contribution of each auxiliary layer $G^{(l)}(l = 1, 2, \cdots, l \neq T)$ to the target layer step by step through setting the GR_{IT} to 0 in descending order. This process is stopped when the prediction results only depend on the target layer. In this case, $GR_{IT} = 0$ for all auxiliary layers, indicating that the second term in Equation (3) has no impact on the prediction results. Consequently, Equation (3) solely comprises the first term. The results are shown in Figure 3.

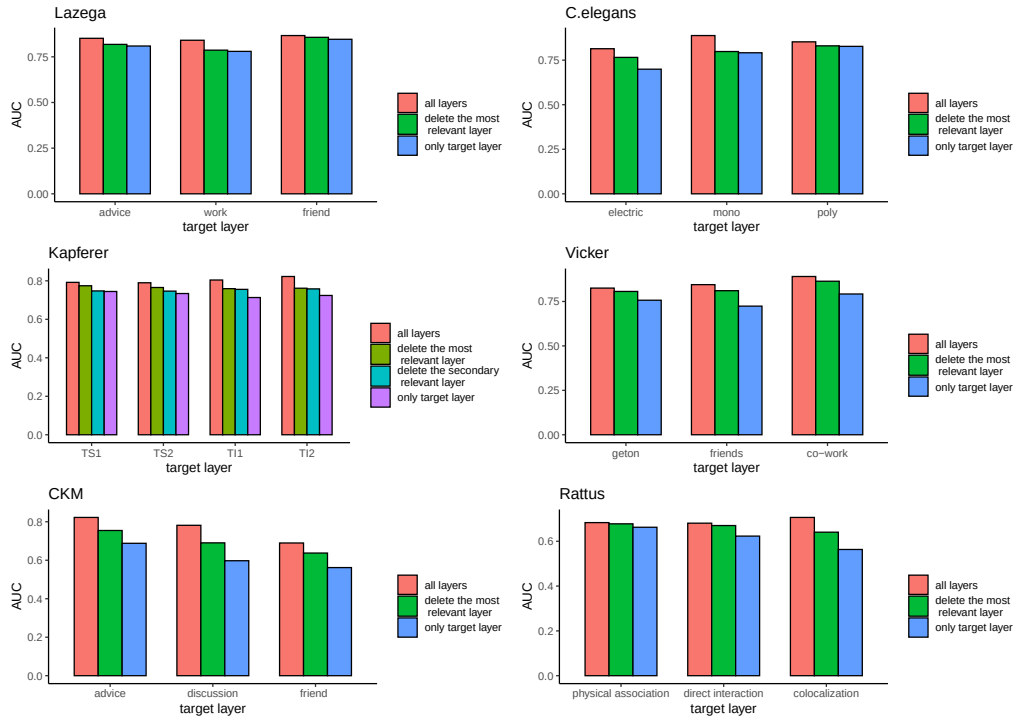

Figure 3. Variation in AUC values with changes in auxiliary layers in six multiplex networks. Each subfigure shows how the LPGRI method's prediction performance for the target layer is affected by the gradual removal of auxiliary layers in descending order of GR value. For each multiplex network, λ is equal to λ^*. Each result is obtained by averaging 30 realizations with a training set containing 90% edges.

As depicted in Figure 3, the prediction performance of the target layer consistently declines as auxiliary layers are gradually removed. In particular, in most multiplex networks, removing the auxiliary layer that exhibits a high relevance to the target layer leads to a more significant decrease in predictive performance. For example, take the first layer of the Lazega multiplex network as the target layer. According to Figure 1, the GR values for the first layer and the other two layers are $GR_{12} = 0.483$ and $GR_{13} = 0.418$, respectively. We first set $GR_{12} = 0$ in Equation (3), which means the second layer is removed from the LPGRI method. The performance of the LPGRI method drops from 0.8513 to 0.8180, corresponding to a rate of decline of 3.91%. However, when the third layer is removed from the LPGRI method (i.e., setting GR_{13} to 0), the performance of the LPGRI method decreases from 0.8180 to 0.8091, with a decline rate of 1.09%. Similar patterns emerge in other multiplex networks. Experimental results demonstrate that the higher the correlation between the auxiliary layer and the target layer, the more useful the auxiliary layer is in predicting the target layer.

4.4. Comparison of LPGRI with Other Methods

In this section, we perform experiments to evaluate the performance of the LPGRI method and compare it with other competitive methods: NSILR [38], LPIS [39], NBS [42] and LPPON [50]. The first two of these methods are designed for predicting links in multiplex networks and the second two are designed for predicting links in a monolayer network.

Table 3 presents a comparison of prediction accuracy quantified by AUC. Note that, the parameters in LPIS and LPPON are tuned to their optimal values subject to maximal AUC. From Table 3, it is evident that the LPGRI method outperforms the other methods in three multiplex networks: Kapferer, Vicker, and Rattus. On the other hand, the LPIS method works the best in the third layer of the C.elegans network, while the NSILR method achieves the best results for the third layer of the Lazega network, the first layer of the

C.elegans network, the fourth layer of the Kapferer network, and the third layer of the CKM network. Furthermore, the LPGRI method demonstrates a strong capability in handling directed multiplex networks, as it performs the best in two out of the three multiplex networks: Vicker and Rattus. For the directed multiplex network CKM, the LPGRI method performs better in the first two layers. In most cases, the predictive performance of the methods utilizing the auxiliary layers information is better than that of those methods designed solely for predicting links in monolayer networks.

Table 3. Mean of AUC for six multiplex networks. For each target layer, λ is equal to λ^* and the training set contains 90% of the edges. Each result is obtained by averaging 30 realizations. The maximum value of AUC in each layer is emphasized in bold.

Type	Dataset	Layer Name	Method					RI
			LPGRI	NSILR	LPIS	NBS	LPPON	
undirected	Lazega	advice	**0.8513**	0.8498	0.8417	0.8245	0.8011	0.18%
		work	**0.8407**	0.8101	0.7625	0.7983	0.7788	0.38%
		friend	0.8664	**0.8769**	0.8639	0.8245	0.8011	−0.12%
	C. elegans	electric	0.8141	**0.8522**	0.7901	0.7011	0.7441	−4.47%
		mono	**0.8880**	0.8648	0.8047	0.7585	0.8139	2.68%
		poly	0.8527	0.8493	**0.8571**	0.8292	0.7952	−0.51%
	Kapferer	TS1	**0.7922**	0.7875	0.7289	0.7540	0.7212	0.60%
		TS2	**0.7899**	0.7786	0.7554	0.7391	0.7148	1.45%
		TI1	**0.8042**	0.7914	0.7561	0.7096	0.7072	1.62%
		TI2	**0.8225**	**0.8225**	0.7685	0.7330	0.7517	0.00%
directed	Vicker	get on	**0.8253**	0.7755	0.8046	0.7576	0.7256	2.57%
		friend	**0.8448**	0.8026	0.8208	0.7585	0.7889	2.92%
		co-work	**0.8906**	0.7924	0.8443	0.7741	0.7926	5.48%
	CKM	advice	**0.8225**	0.7117	0.5913	0.6792	0.7006	15.57%
		discussion	**0.7816**	0.7667	0.5791	0.6022	0.6848	1.94%
		friend	0.6897	**0.7397**	0.5875	0.5442	0.5971	−6.76%
	Rattus	physical association	**0.6830**	0.5458	0.5464	0.6647	0.5872	2.75%
		direct interaction	**0.6702**	0.5175	0.5601	0.6229	0.5923	7.59%
		colocalization	**0.7252**	0.5742	0.5938	0.5588	0.5757	22.13%

In order to draw some further comparisons, we also calculate the ratio of improvement, defined as

$$\text{RI(LPGRI,B)} = \frac{\text{AUC of LPGRI} - \text{AUC of B}}{\text{AUC of B}} \times 100\%, \tag{5}$$

where B represents the method with the best prediction performance among all methods except the LPGRI method in each target layer. The RI value quantifies the degree of improvement of the LPGRI method compared to method B. A positive RI value indicates that the LPGRI method performs better than B, and vice versa. The RI values can be found in the last column of Table 3.

From Table 3, it is clear that the LPGRI method exhibits the highest ratio of improvement compared to method B, especially in the case of directed multiplex network. For example, when compared to other methods, the LPGRI method achieves over 15% improvement in the first layer of CKM. In contrast, the largest improvement made by other methods on the LPGRI method is only up to 6.76% in the third layer of CKM. Additionally, we observe that the LPGRI method has a higher improvement rate in sparse multiplex networks such as CKM and Rattus (the layer density of multiplex networks is shown in Table 2), indicating its ability to effectively handle sparse problems in networks. These results are consistent with the conclusions in Section 4.2.

In summary, according to the results based on the metric of AUC, the LPGRI approach exhibits superior overall performance for link prediction, particularly in the context of

directed multiplex networks. This can be attributed to the effectiveness of the GR index employed by the LPGRI method, which accurately measures the relevance between layers not only in undirected multiplex networks but also in directed multiplex networks.

5. Conclusions

Link prediction represents an important problem in network analysis, with its purpose being to estimate the missing edges or forthcoming edges in the network by using the observed topology structure or node attributes of the network. In the real world, individuals participate in various types of connections within networks, with each type of connection forming a layer. Multiplex networks are employed to represent multiple relationships within a defined set of objects. In this paper, we studied the link prediction problem of multiplex networks. A key issue of link prediction in multiplex networks is estimating the likelihood of potential edges in the target layer by leveraging information from both the target layer and the auxiliary layers. To address this, we introduce the LPGRI method. This approach combines information from the target layer and all auxiliary layers using a tuning parameter λ, where the contribution of each auxiliary layer is determined by the GR index. We evaluate the performance of the LPGRI method on six multiplex networks, including both undirected and directed networks. The experimental results show that the LPGRI method can be applied for friend recommendations in social networks as well as the discovery of interacting gene pairs in biological networks.

In the LPGRI method, the parameter λ is used to regulate the degree of utilization of information between the target layer and auxiliary layers in link prediction. Moveover, it needs to be optimized in different multiplex networks. When the network contains a large number of layers, parameter optimization becomes a challenging task. For this reason, finding a more flexible approach to select the extent of interlayer information's influence on predicting links in the target layer is an area that can be further explored in this work. Additionally, interlayer relevance can be measured from multiple perspectives, and ways to measure the reliability of these indices, including the GR index, is a direction for future research.

Author Contributions: Conceptualization, methodology, software, investigation, writing—original draft preparation, C.W.; writing—review and editing, F.T. and X.Z.; resources, funding acquisition, data curation, F.T. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Natural Science Foundation of China (11971214, 12201235), Natural Science Foundation of Anhui Province (2108085QA14).

Data Availability Statement: The datasets used in this study can be obtained from https://manliodedomenico.com/data.php, accessed on 30 June 2023.

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

Abbreviations

The following abbreviations are used in this manuscript:

LPGRI	Link Prediction based on Global Relevance of Interlayer
GR	Global Relevance
AUC	Area Under the Curve
RI	Ratio of Improvement

References

1. Cardillo, A.; Gómez-Gardeñes, J.; Zanin, M.; Romance, M.; Papo, D.; del Pozo, F.; Boccaletti, S. Emergence of network features from multiplexity. *Sci. Rep.* **2013**, *3*, 1344. [CrossRef] [PubMed]
2. Nicosia, V.; Bianconi, G.; Latora, V.; Barthelemy, M. Growing multiplex networks. *Phys. Rev. Lett.* **2013**, *111*, 058701. [CrossRef] [PubMed]

3. Szell, M.; Lambiotte, R.; Thurner, S. Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13636–13641. [CrossRef]
4. Kivelä, M.; Arenas, A.; Barthelemy, M.; Gleeson, J.P.; Moreno, Y.; Porter, M.A. Multilayer networks. *J. Complex Netw.* **2014**, *2*, 203–271. [CrossRef]
5. Tang, F.; Wang, C.; Wang, Y.; Su, J. Link prediction for multilayer networks using interlayer structural information. *Int. J. Mod. Phys. C* **2022**, *33*, 2250003. [CrossRef]
6. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [CrossRef]
7. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.U. Complex networks: Structure and dynamics. *Phys. Rep.* **2006**, *424*, 175–308. [CrossRef]
8. Newman, M.E.J. The structure and function of complex networks. *SIAM Rev.* **2003**, *45*, 167–256. [CrossRef]
9. Albert, R.; Barabasi, A.L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47. [CrossRef]
10. Liben-Nowell, D.; Kleinberg, J. The link prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 1019–1031. [CrossRef]
11. Lü, L.; Zhou, T. Link prediction in complex networks: A survey. *Physica A* **2011**, *390*, 1150–1170. [CrossRef]
12. Wang, P.; Xu, B.; Wu, Y.; Zhou, X. Link prediction in social networks:the state-of-the-art. *Sci. China Inf. Sci.* **2015**, *58*, 1–38.
13. Ji, P.; Ye, J.; Mu, Y.; Lin, W.; Tian, Y.; Hens, C.; Perc, M.; Tang, Y.; Sun, J.; Kurths, J. Signal propagation in complex networks. *Phys. Rep.* **2023**, *1017*, 1–96. [CrossRef]
14. Martínez, V.; Berzal, F.; Cubero, J.C. A survey of link prediction in complex networks. *ACM. Comput. Surv.* **2016**, *49*, 69.1–69.33. [CrossRef]
15. Lorrain, F.; White, H.C. Structural equivalence of individuals in social networks. *J. Math. Sociol.* **1971**, *1*, 49–80. [CrossRef]
16. Adamic, L.A.; Adar, E. Friends and neighbors on the Web. *Soc. Netw.* **2003**, *25*, 211–230. [CrossRef]
17. Zhou, T.; Lü, L.; Zhang, Y. Predicting missing links via local information. *Eur. Phys. J. B* **2009**, *71*, 623–630. [CrossRef]
18. Martinčić-Ipšić, S.; Močibob, E.; Perc, M. Link prediction on Twitter. *PLoS ONE* **2017**, *12*, e0181079. [CrossRef]
19. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43. [CrossRef]
20. Jeh, G.; Widom, J. Simrank: A measure of structural-context similarity. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23–26 July 2002; ACM Press: New York, NY, USA, 2002; pp. 271–279.
21. Clauset, A.; Moore, C.; Newman, M.E.J. Hierarchical structure and the prediction of missing links in networks. *Nature* **2008**, *453*, 98–101. [CrossRef]
22. Holland, P.W.; Laskey, K.B.; Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **1983**, *5*, 109–137. [CrossRef]
23. Friedman, N.; Getoor, L.; Koller, D.; Pfeffer, A. *Learning Probabilistic Relational Models*; Springer: Berlin, Germany, 2000.
24. Heckerman, D.; Meek, C.; Koller, D. Probabilistic entity-relationship models, PRMs, and plate models. In Proceedings of the 21st International Conference on Machine Learning, Banff, AB, Canada, 4–8 July 2004; pp. 55–60.
25. Yu, K.; Chu, W.; Yu, S.; Tresp, V.; Xu, Z. Stochastic relational models for discriminative link prediction. *Adv. Neural Inf. Process. Syst.* **2006**, *19*, 1553–1560.
26. Zhou, T. Progresses and challenges in link prediction. *iScience* **2021**, *24*, 103217. [CrossRef] [PubMed]
27. Kumar, A.; Singh, S.S.; Singh, K.; Biswas, B. Link prediction techniques, applications, and performance: A survey. *Physica A* **2020**, *553*, 124289. [CrossRef]
28. Jafari, S.H.; Abdolhosseini-Qomi, A.M.; Asadpour, M.; Rahgozar, M.; Yazdani, N. An information theoretic approach to link prediction in multiplex networks. *Sci. Rep.* **2021**, *11*, 13242. [CrossRef]
29. Davis, D.; Lichtenwalter, R.; Chawla, N.V. Multi-relational link prediction in heterogeneous information networks. In Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining, ASONAM, Kaohsiung, Taiwan, 25–27 July 2011; pp. 281–288.
30. Hristova, D.; Noulas, A.; Bown, C.; Musolesi, M.; Mascolo, C. A multilayer approach to multiplexity and link prediction in online geo-social networks. *EPJ Data Sci.* **2016**, *5*, 1–17. [CrossRef]
31. Li, W.; Li, T.; Berahmand, K. An effective link prediction method in multiplex social networks using local random walk towards dependable pathways. *J. Comb. Optim.* **2023**, *45*, 31. [CrossRef]
32. Cao, J.; Lei, T.; Li, J.; Jiang, J. A Novel Link Prediction Method for Social Multiplex Networks Based on Deep Learning. *Mathematics* **2023**, *11*, 1705. [CrossRef]
33. Jalili, M.; Orouskhani, Y.; Asgari, M.; Alipourfard, N.; Perc, M. Link prediction in multiplex online social networks. *R. Soc. Open Sci.* **2017**, *4*, 160863. [CrossRef]
34. Bacco, C.D.; Power, E.A.; Larremore, D.B.; Moore, C. Community detection, link prediction, and layer interdependence in multilayer networks. *Phys. Rev. E* **2017**, *95*, 042317. [CrossRef]
35. Yang, Y.; Chawla, N.; Sun, Y.; Hani, J. Predicting links in multi-relational and heterogeneous networks. In Proceedings of the 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; pp. 755–764.
36. Sharma, S.; Singh, A. An efficient method for link prediction in complex multiplex Networks. In Proceedings of the 11th Signal-Image Technology and Internet-Based Systems (SITIS), Bangkok, Thailand, 23–27 November 2015; pp. 453–459.

37. Sharma, S.; Singh, A. An efficient method for link prediction in weighted multiplex networks. *Comput. Soc. Netw.* **2016**, *3*, 1–17. [CrossRef] [PubMed]
38. Yao, Y.; Zhang, R.; Yang, F.; Yuan, Y.; Sun, Q.; Qiu, Y.; Hu, R. Link prediction via layer relevance of multiplex networks. *Int. J. Mod. Phys. C* **2017**, *28*, 1750101. [CrossRef]
39. Najari, S.; Salehi, M.; Ranjbar, V.; Jalili, M. Link prediction in multiplex networks based on interlayer similarity. *Physica A* **2019**, *536*, 120978. [CrossRef]
40. Abdolhosseini-Qomi, A.M.; Jafari, S.H.; Taghizadeh, A.; Yazdani, N.; Asadpour, M.; Rahgozar, M. Link prediction in real-world multiplex networks via layer reconstruction method. *R. Soc. Open Sci.* **2020**, *7*, 191928. [CrossRef]
41. Liao, H.; Zeng, A.; Zhang, Y.C. Predicting missing links via correlation between nodes. *Physica A* **2015**, *436*, 216–223. [CrossRef]
42. Zhang, Y.; Elizaveta, L.; Zhu, J. Estimating network edge probabilities by neighborhood smoothing. *Biometrika* **2017**, *104*, 771–783. [CrossRef]
43. Lazega, E. *The Collegial Phenomenon: The Social Mechanisms of Cooperation among Peers in a Corporate Law Partnership*; Oxford University Press: Oxford, UK, 2001.
44. Chen, B.L.; Hall, D.H.; Chklovskii, D.B. Wiring optimization can relate neuronal structure and function. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4723–4728. [CrossRef]
45. Kapferer, B. *Strategy and Transaction in an African Factory*; Manchester University Press: Manchester, UK, 1972.
46. Vickers, M.; Chan, S. *Representing Classroom Social Structure*; Victoria Institute of Secondary Education: Melbourne, Australia, 1981.
47. Coleman, J.; Katz, E.; Menzel., H. The diffusion of an innovation among physicians. *Sociometry* **1957**, *20*, 253–270. [CrossRef]
48. Domenico, M.D.; Nicosia, V.; Arenas, A.; Latora, V. Structural reducibility of multilayer networks. *Nat. Commun.* **2015**, *6*, 6864. [CrossRef]
49. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [CrossRef]
50. Zhao, Y.; Wu, Y.J.; Levina, E.; Zhu, J. Link Prediction for Partially Observed Networks. *J. Comput. Graph. Stat.* **2017**, *26*, 725–733. [CrossRef]