scientific reports

Check for updates

OPEN

An information theoretic approach to link prediction in multiplex networks

Seyed Hossein Jafari✉, Amir Mahdi Abdolhosseini-Qomi, Masoud Asadpour, Maseud Rahgozar & Naser Yazdani

The entities of real-world networks are connected via different types of connections (i.e., layers). The task of link prediction in multiplex networks is about finding missing connections based on both intra-layer and inter-layer correlations. Our observations confirm that in a wide range of real-world multiplex networks, from social to biological and technological, a positive correlation exists between connection probability in one layer and similarity in other layers. Accordingly, a similarity-based automatic general-purpose multiplex link prediction method—SimBins—is devised that quantifies the amount of connection uncertainty based on observed inter-layer correlations in a multiplex network. Moreover, SimBins enhances the prediction quality in the target layer by incorporating the effect of link overlap across layers. Applying SimBins to various datasets from diverse domains, our findings indicate that SimBins outperforms the compared methods (both baseline and state-of-the-art methods) in most instances when predicting links. Furthermore, it is discussed that SimBins imposes minor computational overhead to the base similarity measures making it a potentially fast method, suitable for large-scale multiplex networks.

Link prediction has been an area of interest in the research of complex networks for over two decades[1], studying the relationships between entities (nodes) in data represented as graphs. The main goal is to reveal the underlying truth behind emerging or missing connections between node pairs of a network. Link prediction methods have a wide range of applications, from discovery of latent and spurious interactions in biological networks (which is basically quite costly if performed in traditional methods)[2,3] to recommender systems[4,5] and better routing in wireless mobile networks[6]. Numerous perspectives have been adopted to attack the problem of link prediction.

According to similarity-based methods, similarity between nodes determines their likelihood of linkage. This approach is a result of assuming that two nodes are similar if they share many common features[7]. A whole lot of nodes' features stay hidden (or are kept hidden intentionally) in real networks. Further, an interesting question is, despite the fact that a considerable amount of information is hidden in a network, what fraction of the truth can still be extracted by merely including *structural features*? That is one of the main drives to utilize structural similarity indices for link prediction. Several different classifications of similarity measures have been proposed, among all, classifying based on locality of indices is of great importance. To name a few, Common Neighbors (CN)[1], Preferential Attachment (PA)[8], Adamic-Adar (AA)[9] and Resource Allocation (RA)[10] are popular indices focusing mostly on nodes' structural features, each with unique characteristics. Even though these indexes are simple, they are popular because of their low computational cost and reasonable prediction performance. On the other hand, global indices take features of the whole network structure into account, tolerating higher cost of computation, usually in favor of more accurate information. Take length of paths between pairs of nodes for instance, which the well-known Katz[11] index operates on. Average Commute Time (ACT)[1] and PageRank[12] are some other notable global indices. In between lie the quasi-local methods which are able to combine properties from both local and global indices, meaning they include global information, but their computational complexity is similar to that of local methods, such as the Local Path (LP)[13] index and Local Random Walk (LRW)[14]. For more detailed information on these similarity indices (also described as *unsupervised* methods in the literature[15]), readers are advised to refer to[16].

Some researchers have tackled the link prediction problem using the ideas of information theory. These works are based on the fact that similarity of node pairs can be written in term of the uncertainty of their connectivity. At the beginning, the uncertainty of connectivity can be estimated based on priors. Later, all structures around the unconnected node pairs can be considered as evidences to reduce the level of uncertainty in connectedness

School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran. ✉email: jafari.h@ut.ac.ir

of node pairs. In[17] mutual information (MI) of common neighbors is incorporated to estimate the connection likelihood of a node pair. In addition, Path Entropy (PE)[18] similarity index takes quantity and length of paths as well as theirentropy into account. This results in a better assessment of connection likelihood for node pairs. In[19], authors proposed an information theoretic method to benefit from several structural features at the same time. By using information theory, they score each structural feature separately and then combine them by weighted summation. Then they apply the idea on common neighbors and connectivity of neighbor sets as two structural features. Although, most of literature about link prediction is devoted to unweighted networks but a few works have targeted the weighted networks. In[20], authors use a weighted mutual information to predict weighted links which benefits from both structural properties and link weights. The results are promising when compared to both weighted and unweighted methods.

In a coarse-grained sense, learning-based link prediction models reside in a different class than aforementioned similarity-based ones. They learn a group of parameters by processing input graph and use certain models, such as feature-based prediction (HPLP[21]) and latent feature extraction (Matrix Factorization[15]). Representation learning has helped automating the entire process of link prediction, especially feature selection; node2vec[22] and DGI[23], for instance. Recently, an interesting multiplex embedding model has also been proposed called DMGI[24] which is basically an extension of DGI. Learning-based methods often yield better results than their similarity-based counterparts, but that does not mean these models are obsolete. On the one hand, similarity-based models provide a better understanding of the underlying characteristics of networks. Take common neighbors (CN) for example, which indicates the high clustering property of networks[18] or Adamic-Adar index which is based on the size of common nodes' neighborhoods[9]. On the other hand, similarity-based methods often take less computation effort, making them suitable for online prediction without costly training procedures or feature selection stages[25].

Related works

Complex networks research was focused on single-layer networks (simplex or mono-plex) for many years. The study of multi-layer (multiplex or heterogeneous) networks has gained the attention of researchers in the past few years. Refs.[26,27] provide noteworthy reviews on history of multi-layer networks. The attempts to predict multi-layer links are not abundant and some are discussed here.

Hidden geometric correlation in real multiplex networks[28] is an interesting work which depicts how multiplex networks are not just random combinations of single-layer networks. They employ these geometric correlations for trans-layer link prediction i.e., incorporating observations of other layers for predicting connections in a specific layer. This work is followed by a study that argues the requirement of a link persistence factor to explain high edge overlap in real multiplex systems[29]. In heterogeneous networks (i.e., networks with different types of nodes and relations), several similarity-search approaches have been proposed. PathSim[30] is a meta path-based similarity measure that can find similar peers in heterogeneous networks (e.g. authors in similar fields in a bibliographic network). The intuition behind PathSim is that two peer objects are similar if they are not only strongly connected, but also share comparable visibility (number of path instances from a node to itself). HeteSim[31] is another method of the same kind which can measure similarity of objects of different type, inspired by the intuition that two objects are related if they are referenced by related objects. Their drawback, however, is their dependence on connectivity degrees of node-pairs (neglecting further information provided by meta paths themselves) and their necessity of using one and usually symmetric meta-path. In[32], a mutual information model has been employed to tackle these problems. Most meta path-based models suffer from lack of an automated meta-path selection mechanism, in other words, pre-defined meta paths (mostly specific to the dataset under study) are utilized for prediction. In the previously discussed methods, including longer meta paths required much more computation to analyze them and determine their effects.

Link prediction for multiplex networks has been addressed by researchers using features and machine learning. A study of a multiplex online social network, demonstrates the importance of multiplex links (link overlap) in significantly higher interaction of users based on available side information[33]. The authors consider Jaccard similarity of extended neighborhood of nodes in the multiplex network as a feature for training a classifier for link prediction task. A similar work on the same dataset benefits from node-based and meta-path-based features[34]. A specialized type of these meta-paths is tailored to be originated from and ending at communities. The effectiveness of the features has been examined by a binary classification for link predication task. Recently, other interlayer similarity features, based on degree, betweenness, clustering coefficient and similarity of neighbors has been used[35].

Furthermore, the issue of link prediction has been investigated in a scientific collaboration multiplex network[36]. The authors have proposed a supervised rank aggregation paradigm to benefit from the node pairs ranking information which is available in other layers of the network. Another study uses rank aggregation method on a time-varying multiplex network[37].

Yao et al. in[38] discuss the issue of layer relevance and its effect on link prediction task. The authors use global link overlap rate (GOR) and Pearson correlation coefficient (PCC) of node features as measures of layer relevance and later they use it to combine the basic similarity measures of each layer. The results support that the more layers are relevant, the better performance of link prediction is attained. In this work, well-known single-layer similarity measures like CN, RA, and LPI are used. We compare our work with their best performing methods. They show that LPI as a quasi-local metric is the best choice of base similarity measure. For interlayer relevance both GOR and PCC perform well and we refer to them as YaoGL and YaoPL, respectively. Samei et al. have studied the effect of other layers on the target layer using global link overlap rate[39]. Two features based on hyperbolic distance are used, WCN and HP. WCN uses embedded network in geometric space and calculates hyperbolic distance of nodes to weigh the importance of common neighbors. HP considers the hyperbolic distance of nodes

as a dissimilarity measure. Similar to Yao et al., they use GOR to aggregate the score of the two layers. Our results are also compared with this work.

Recently, link prediction problem is studied with the focus of community structure of the layers[40]. This study reveals the importance of similarity of community structure of different layers in link prediction. In[41], it is shown that similarity of eigenvectors of the layers' adjacency matrices is an important source of information for multiplex link prediction. Authors propose reconstruction of one layer with eigenvectors of another layer that proves to be very helpful even if a large portion of links is missing in the target layer.

A systematic approach is extending the basic similarity measures to multiplex networks. However, when it comes to multiplex networks, it's hard to extend the notion of similarity[42]. In a recent work, MAA is presented which extends AA similarity measure to encode diverse forms of interactions[43]. It is suggested that this approach can improve the results of link prediction in certain circumstances compared to the single-layer counterpart.

In this paper, an information-theoretic model is devised that employs other layers' structural information for better link prediction in some arbitrary (target) layer of the network. Through the incorporation of various similarity indices (RA, CN, ACT and LPI) as the base proximity measures, we demonstrate that the proposed method -SimBins- can be used to predict multiplex links without degrading the time complexity significantly. Finally, it is shown that SimBins improves prediction performance on several different real-world social, biological and technological multiplex networks.

Methods

Link prediction in multiplex networks. Consider a multiplex network $G\big(V, E^{[1]}, \ldots, E^{[M]};\ E^{[\alpha]} \subseteq V \times V\ \forall\, \alpha \in \{1, 2, \ldots, M\}\big)$[1] where M, V and E^{α} are the number of layers, the set of all nodes and existing edges in layer α of the multiplex network, respectively. Let $U = V \times V$ be the set of all possible node pairs. Current research aims to study undirected multiplex networks; therefore, it is assumed that $G(V, E^{\alpha})$ for any arbitrary layer α is an undirected simple graph. The link prediction in multiplex networks is concerned with the issue of predicting missing links in an arbitrary target layer $T \in \{1, 2, \ldots, M\}$ with the help of other auxiliary layers. To be able to evaluate the proposed method, E^T i.e. the edges in target layer is divided into a training set E^T_{train} (90% of E^T) and a test set E^T_{test} (10% of E^T) so that $E^T_{\text{train}} \cup E^T_{\text{test}} = E^T$ and $E^T_{\text{train}} \cap E^T_{\text{test}} = \emptyset$. Only the information provided by the training set is used in the prediction task and eventually, E^T_{test} is compared to the output of the proposed algorithm (link-existence likelihood scores for a subset of $U - E^T_{\text{train}}$, including E^T_{test}), determining the performance of the method. To be more specific, link likelihood scores are calculated for node pairs of E^T_{test} and a random subset Z^T_{test} of $U - E^T$ where $|Z^T_{\text{test}}| = 2|E^T_{\text{test}}|$ for which all of them are disconnected in E^T_{train}. To put it in a few words; only a subset of non-observed links in training set are scored for the sake of complexity which will be discussed in detail later. Notice coefficient 2, a ratio incorporated to implement the link imbalance assumption in real networks (that are mostly sparse by nature[44]).

In the present study, the issue under scrutiny is how employing one layer of the multiplex network such as A, facilitates the task of link prediction in another layer T where $T, A \in \{1, \ldots, M\};\ T \neq A$i.e., a *duplex* subset of the multiplex network. In 'Discussion' section, it is argued that how one can extend the proposed method to utilize the structural information of multiple layers for link prediction.

Evaluation methods. In their ideal form, link prediction algorithms tend to rank non-observed links in a network so that all latent links are situated on top of the ranking and all other non-existent links underneath. This ranking is based on a link-likelihood score that is dedicated to node pairs corresponding to non-observed links in the network. For imperfect rankings a metric is required to assess the quality of the ranking. Here, we describe two evaluation metrics used in this research.

AUC: Using of Area Under Receiver Operating Characteristic Curve (AUC or AUROC)[45] is prominent in the literature for evaluating link prediction methods[16]. AUC indicates the probability that a randomly chosen missing link is scored higher than a randomly chosen non-existent link, denoted as:

$$\text{AUC} = \frac{n' + 0.5n''}{n} \tag{1}$$

where by performing n times of independent comparisons ($n = 10000$ in our experiments), a randomly chosen latent link has a higher score compared to a randomly chosen non-existent link in n' times and are equally scored in n'' times. AUC will be 1 if the node pairs are flawlessly ranked and 0.5 if the scores follow an identical and independent distribution i.e., the higher the AUC, the better the scoring scheme is.

Precision: Given the ranked (by score) list of the non-observed links, the precision is defined as the ratio of the missing links to the number of selected items from the top of the list. That is to say, if we take the top-L links as the predicted ones, among which L_r links are known missing links; Precision is defined as:

$$\text{Precision} = \frac{L_r}{L} \tag{2}$$

Here, we consider $L = \left|E^T_{test}\right|$. Clearly, higher precision indicates higher prediction accuracy.

Data. Various real-world multiplex network datasets from different domains are selected for investigation; from social (Physicians, NTN and CS-Aarhus) to technological (Air/Train and London Transport) and biological systems (C. Elegans, Drosophila and Human Brain). They also have diverse characteristics that are briefly introduced in Table 1.

Multiplex name	No. of layers	No. of nodes	Node multiplexity	Layer name	No. of active nodes	No. of links
Air/train	2	69	1	Air	69	180
				Train	69	322
C. Elegans	3	280	0.98	Electric	253	515
				Chem-mono	260	888
				Chem-poly	278	1703
Drosophila	2	839	0.89	Suppress	838	1858
				Additive	755	1424
Brain	2	90	0.85	Structure	85	230
				Function	80	219
Physicians	3	246	0.93	Advice	215	449
				Discuss	231	498
				Friend	228	423
NTN	4	78	0.94	Communication	74	200
				Financial	13	15
				Operational	68	437
				Trust	70	259
London	3	368	0.13	Tube	271	312
				Overground	83	83
				DLR	45	46
CS-Aarhus	5	61	0.96	Lunch	60	193
				Facebook	32	124
				Co-author	25	21
				Leisure	47	88
				Work	60	194
SacchPomb	5	4092	0.28	Direct	936	1332
				Colocalization	346	370
				Physical	2400	6973
				Synthetic	897	2540
				Association	181	218

Table 1. Basic characteristics of multiplex networks used in experiments.

Air/Train (AT). This dataset consists of Indian airports network and train stations network and their geographical distances[46]. To relate the **train** stations to the geographically nearby **air**ports, in[28] they have aggregated all train stations within 50 km from an airport into a super-node. Then, the super-nodes are considered as connected if they share a common train station, or if one train station of one super-node is directly connected to a station of the other super-node. Air is the network of airports and Train is the network of aggregated train station super-nodes.

C. Elegans. The network of neurons of the nematode Caenorhabditis Elegans that are connected through miscellaneous synaptic connection types: **Electric**, **Chemical Monadic** and **Chemical Polyadic**[47].

Drosophila Melanogaster (DM). Layers of this network represent different types of protein–protein interactions belonged to the fly Drosophila Melanogaster, namely **suppressive** genetic interaction and **additive** genetic interaction. More details can be found in[48,49].

Human Brain (HB). The human brain multiplex network is taken from[28,50]. It consists of a **structural** or anatomical layer and a **functional** layer that connect 90 different regions of the human brain (nodes) to each other. The structural network is gathered by dMRI and the functional network by BOLD fMRI[50]. In this multiplex network, the structural connections are obtained by setting a threshold on connection probability of brain regions (which is proportional to density of axonal fibers in between)[28]. The functional interactions are derived in a similar manner, by putting a threshold on the connection probability of regions which is proportional to a correlation coefficient measured for activity of brain region pairs[28].

Physicians. Taken from[51], the Physicians multiplex dataset contains 3 layers which relate physicians in four US towns by different types of relationships; to be specific, **advice**, **discuss** and **friendship** connections.

Noordin Top Terrorist Network (NTN). Taken from[52], this multiplex dataset is made of information among 78 individuals i.e. Indonesian terrorists that depicts their relationships with respect to exchanged **communications**, **financial** businesses, common **operations** and mutual **trust**.

London Transport. For the purpose of studying navigability performance under network failures, De Domenico et al.[53] gathered a dataset for public transport of London consisting of 3 different layers; the **tube**, the **overground**, and the docklands light railway (**DLR**). Nodes are stations which are linked to each other if a real connection exists between them in the corresponding layer.

CS-Aarhus. This dataset is collected from[54] which is conducted at the Department of Computer Science at Aarhus University in Denmark among the employees. The network consists of 5 different interactions

corresponding to current **work** relationships, repeated **leisure** activities, regularly eating **lunch** together, **co-author**ship of publications and friendship on **Facebook**.

SacchPomb. The SacchPomb dataset is taken from[28,48] and represents the multiplex genetic and protein interaction network of the Saccharomyces Pombe (fission yeast). The multiplex consists of 5 layers corresponding to 5 different types of interactions. Layer 1 corresponds to **direct interaction**, Layer 2 to **colocalization**, Layer 3 to **physical association**, Layer 4 to **synthetic genetic interaction**, and Layer 5 to **association**. More details on the data can be found in[48].

Node multiplexity in Table 1 shows the fraction of nodes in a multiplex network that are active (have at least one link attached) in more than one layer.

Information theory background. This sub-section is concerned with the issue of introducing necessary concepts of information theory, as it lays out the main mathematical background of the proposed method. What follows is the definition of self-information and mutual information.

Given a random variable X, the *self-information* or surprisal of occurrence of event $x \in X$ with probability $p(x)$ is defined as[55]:

$$I(X = x) = -\log p(x) \tag{3}$$

The self-information implies how much uncertainty or surprise there is in the occurrence of an event; the less probable the outcome is, the more the surprise it conveys. The base of the logarithmic functions is assumed to be 2 throughout the paper, as they measure uncertainty in *bits* of information.

Let's proceed with the definition of mutual information between two random variables X and Y with joint probability mass function $p(x, y)$ and marginal probability mass functions $p(x)$ and $p(y)$, respectively. The *mutual information* $I(X; Y)$ is[56]:

$$\begin{aligned} I(X;Y) &= \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= \sum_{x,y} p(x,y)\log\frac{p(x|y)}{p(x)} \end{aligned} \tag{4}$$

Consequently, the mutual information of two events $x \in X$ and $y \in Y$ can be denoted as[17]:

$$\begin{aligned} &I(X = x; Y = y) = \log\frac{p(x|y)}{p(x)} = -\log p(x|y) - (-\log p(x)) \\ &= I(x) - I(x|y) \end{aligned} \tag{5}$$

In fact, the mutual information indicates how much two variables are dependent to each other i.e., for a variable X, how much uncertainty is reduced due to observation of another variable Y. The mutual information would be zero if and only if two variables are independent. In the following section, we will describe how these two measures play their roles in designation of our method.

Base similarity measures. There is extensive literature on similarity measures that determine how similar two nodes are in a single-layer network; as it was partially presented on introduction of this paper. In our proposed method, a subset of these similarity indices (both local and global) is used as base measures that the multiplex link prediction model is built on top of them.

CN[1]: Maybe, the most well-known and typical way to measure similarity of two nodes x and y is to count the number of their common neighbors:

$$S_{xy}^{CN} = |\Gamma(x) \cap \Gamma(y)| \tag{6}$$

where $\Gamma(x)$ and $\Gamma(y)$ are the set of neighbors of x and y, respectively.

RA[10]: In Resource Allocation, degree of a node is considered as a resource that is allocated to the neighbors of that node negatively proportional to its degree:

$$S_{xy}^{RA} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} |\Gamma(z)|^{-1} \tag{7}$$

ACT[1]: Random-walk based methods account for the steps required for reaching one node starting from some arbitrary node. Average Commute Time measures the average number of steps required for a random walker to reach node y starting from node x. For the sake of computational complexity, pseudo-inverse of Laplacian matrix is utilized to calculate the commute time:

$$S_{xy}^{ACT} = \frac{1}{l_{xx}^{+} + l_{yy}^{+} - 2l_{xy}^{+}} \tag{8}$$

where l_{xy}^{+} is the $[x, y]$ entry in pseudo-inverse Laplacian matrix i.e., $l_{xy}^{+} = [L^{+}]_{xy}$. The pseudo-inverse of Laplacian is calculated as[57]:

$$L^{+} = \left(L - \frac{ee'}{n}\right)^{-1} + \frac{ee'}{n} \tag{9}$$

where e is a column vector of 1's (e' is its transpose) and n is the total number of the nodes.

LPI[10,13]: To provide a good tradeoff of accuracy and computational complexity, the Local Path Index (LPI) is introduced as an index that takes consideration of local paths, with wider horizon than CN. It is defined as:

$$S^{LPI} = A^2 + \varepsilon A^3 \tag{10}$$

where ε is a free parameter. Clearly, this measure degenerates to CN when $\varepsilon = 0$. And if x and y are not directly connected, $(A^3)_{xy}$ is equal to the number of different paths with length 3 connecting x and y. This index can be extended for higher order paths and considering paths of infinite length this similarity measure converges to Katz index. The LP index performs remarkably better than the neighborhood-based indices, such as RA and CN. Throughout the current work, ε is set to 10^{-4} wherever LPI is used. This is the same for the compared methods. In[16], it is stated that the value of can be directly set as a very small number instead of finding its optimum, which may take a long time. In particular, the essential advantage of using a second-order neighborhood is to improve the distinguishability of similarity scores.

For more details on base similarity measures, readers are encouraged to see surveys on link prediction algorithms[16,58].

Results

Does the structure of one layer of a multiplex, provide any information on the formation of links in some other layer of the same network? Take a social multiplex network, for example, in which one layer states people's work relationships and the other layer represents their friendship. Intuitively it can be conjectured that in a real multiplex like our sample social network, structural changes in one layer can affect the other; if two people become colleagues, the conditions of them being friends will probably not be the same as it was before. More specifically, is there any correlation among the structure of layers of a multiplex network? This question has been positively answered in previous studies with different approaches. In[28] a null model is created for a multiplex network, by randomly reshuffling inter-layer node-to-node mappings. Subsequently, it is shown that geometric inter-layer correlations are destroyed in the null model compared to the original network.

Various structural features can be analyzed to uncover correlations between layers. Direct links, common neighbors, paths[1] and eigenvectors[59] are such examples. In the following sections we will develop a set of tools that assist in collection of evidences about inter-layer correlations in multiplex networks, as basic intuitions supporting the proposed link prediction framework.

Partitioning Node Pairs (Binning). Consider two layers $T, A \in \{1, 2, \ldots, M\}$; $T \neq A$ of a multiplex network with Mlayers and V nodes. T is the target layer, so it is intended to predict likelihood of presence of links in that layer, and A is the auxiliary layer assisting the prediction task. A subset U' of $U = V \times V$ is constituted so that $U' = E^{T}_{\text{train}} \cup Z^{T}_{\text{train}}$ where Z^{T}_{train} is a random sample of non-observed links from $U - E^{T}$ and $|Z^{T}_{\text{train}}| = 2|E^{T}_{\text{train}}|$. The size of Z^{T}_{train} is twice as large as E^{T}_{train}, so that U' would be a suitable representative of the target layer due to the link imbalance phenomenon in real complex systems. Two different partitions of U' is formed (using equal-depth binning, described in the following paragraph):

(i) w.r.t the target layer T:

$$\{S^{T}_{1}, S^{T}_{2}, \ldots, S^{T}_{b_T}\} \text{ where } \bigcup_{i=1}^{b_T} S^{T}_{i} = U' \text{ and } \forall i, j \in \{1, 2, \ldots, b_T\}, i \neq j \Rightarrow S^{T}_{i} \cap S^{T}_{j} = \emptyset.$$

(ii) With respect to the auxiliary layer A:

$$\{S^{A}_{1}, S^{A}_{2}, \ldots, S^{A}_{b_A}\} \text{ where } \bigcup_{j=1}^{b_A} S^{A}_{j} = U' \text{ and } \forall i, j \in \{1, 2, \ldots, b_A\}, i \neq j \Rightarrow S^{A}_{i} \cap S^{A}_{j} = \emptyset.$$

These partitions are introduced as ***bins*** of node pairs in current study. The number of bins w.r.t target and auxiliary layer are b_T and b_A, respectively. An equal-depth (frequency) binning strategy is applied to the target layer similarity scores of the node pairs in U', in order that each partition $S^{T}_{i}; i \in \{1, 2, \ldots, b_T\}$ contains approximately the same number of members (node pairs). The same strategy goes for similarity scores in auxiliary layer A, establishing $S^{A}_{j}; j \in \{1, 2, \ldots, b_A\}$ partitions. It should be noted that S^{T}_{i} andS^{A}_{j} are two different partitions of the same set, namely U'. To make distinction between these two partitions, readers should pay attention to the superscript in the notation. Therefore, for $i = j$, S^{T}_{i} is not necessarily equal to S^{A}_{j} because the former partitioning is based on similarity in the target layer while the latter is based on similarity in the auxiliary layer.

Aforementioned partitions (bins) form the building blocks of how the multiplex networks are scrutinized in this paper, as they put forward a coarse-grained view of the data; tolerating the insignificant fluctuations observed in particular regions of the networks. The setting denoted above will be used from now onwards, to avoid any further repetitions.

Intra-layer and trans-layer connection probabilities. The foregoing discussion introduces two key measures for target and auxiliary layer bins, namely S^{T}_{i} and S^{A}_{j}: (1) intra-layer connection probability $p_{\text{intra}}(S^{T}_{i})$,

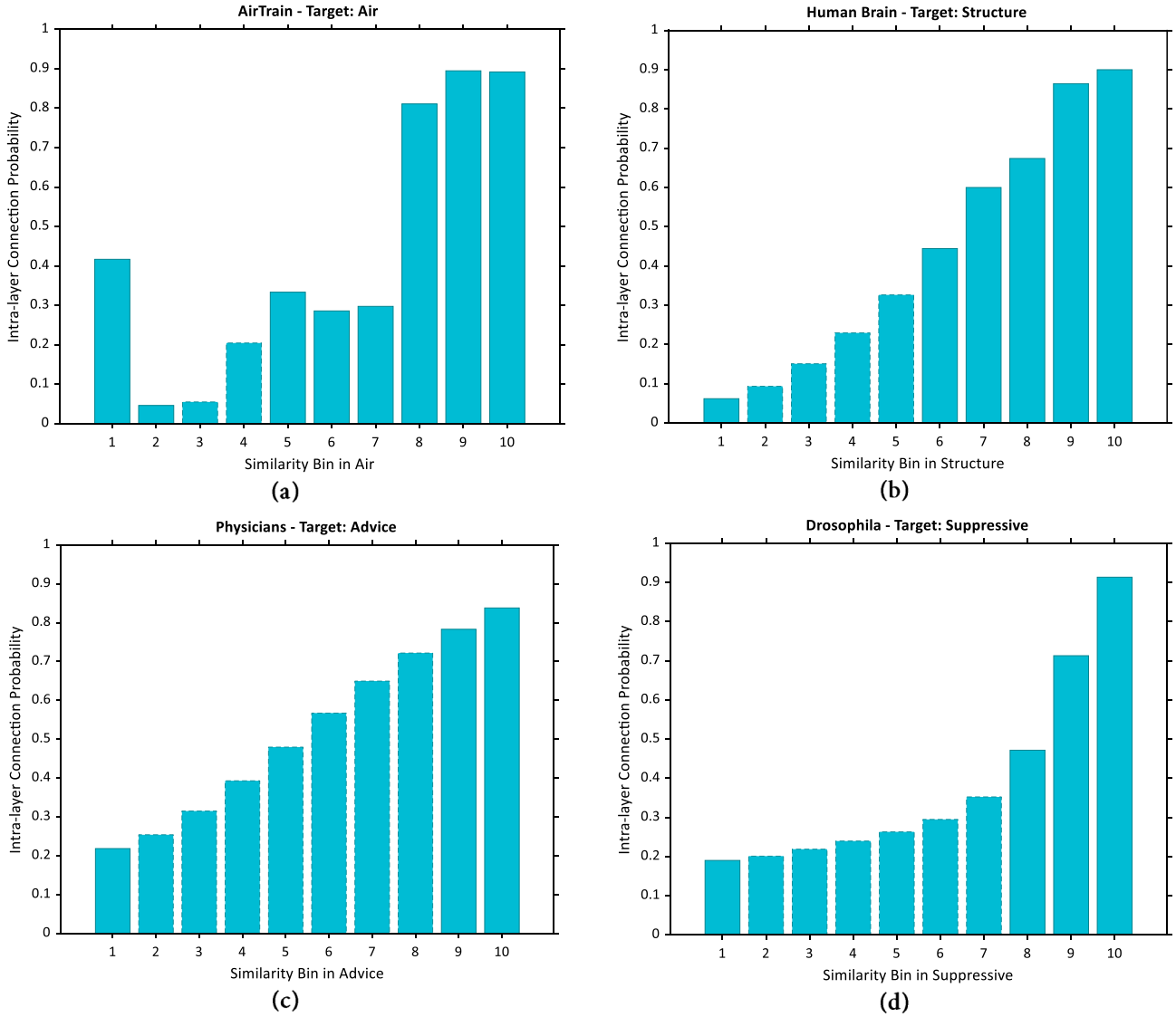

Figure 1. Intra-layer connection probability in target layer bins. Intra-layer connection probability or fraction of node pairs in a bin that are linked in layer (**a**) 'Air' of the network Air/Train, (**b**) 'Structure' of Human Brain, (**c**) 'Advice' of Physicians, (**d**) 'Suppressive' of Drosophila. Bars with dashed lines represent imputed probabilities.

and (2) trans-layer connection probability $p^T_{\text{trans}}(S^A_j)$. Intra-layer connection probability in S^T_i is the connection likelihood of pairs existing in that bin. This measure can also be expressed as conditional probability of connection of an arbitrary node pair x, y in layer T, given their similarity (bin) in the same layer:

$$p_{\text{intra}}(S^T_i) = p(L^T = 1 | S^T_i);\ i \in \{1, 2, \ldots, b_T\} \tag{11}$$

Notice $L^T = 1$, which is the event that any randomly selected pair (x, y) are linked in layer T. Empirically, $p_{\text{intra}}(S^T_i)$ is computed as proportion of linked node pairs in S^T_i to all of node pairs in the set:

$$\tilde{p}_{\text{intra}}(S^T_i) = \frac{|S^T_i \cap E^T_{\text{train}}|}{|S^T_i|};\ i \in \{1, 2, \ldots, b_T\} \tag{12}$$

Intra-layer connection probability for four different multiplex (duplex) networks is provided for each bin in (Fig. 1). In data-driven observations of this paper, wherever a similarity measure is involved, Resource Allocation (RA) index is used; otherwise specified. Additionally, it is assumed that the number of bins in both the target and auxiliary layers i.e., b_T and b_A are set to 10. Our experiments show that too small number of bins leads to significant decrement in prediction results.

In most of the cases, increasing the number of bins either has no effect on prediction results or degrades them (although not quite significantly). Additionally, large number of bins brings unnecessary computational complexity to our algorithm. We have also tried a more adaptive approach for choosing the number of bins by maximizing the entropy of node-pairs distribution in bins which lead to no substantial improvement in prediction. A value

between 10 and 50 is recommended as SimBins shows no significant sensitivity in terms of accuracy within the mentioned range and the computational overhead is miniscule.

The bars with dashed lines in (Fig. 1) represent imputed values. Because of high frequency of some certain similarity values (such as 0 scores in RA for node pairs with no common neighbors), a perfect equal-depth binning may not be feasible; as a result, a number of bins will contain no sample node pairs. The value of intra-layer connection probability for these bins has been imputed using a penalized least squares method which allows fast smoothing of gridded (missing) data[60]. In addition to more clear observations, this imputation will let us fix the number of bins and handle missing data in a systematic way. The results indicate that by the increment of similarity (higher bin numbers) intra-layer connection probability increases respectively, depicting a positive correlation between similarity (bin number) and intra-layer connection probability; as stated in seminal work of Liben-nowell and Kleinberg[1].

Trans-layer connection probability is defined analogously except that although connection in target layer T is concerned, the similarity scores of node pairs are given in auxiliary layer A. Similar to formula (11), $p^T_{\text{trans}}(S^A_j)$ can be defined as follows:

$$p^T_{\text{trans}}(S^A_j) = p(L^T = 1|S^A_j);\ j \in \{1, 2, \ldots, b_A\} \tag{13}$$

Empirical value of trans-layer connection probability is calculated likewise:

$$\tilde{p}^T_{\text{trans}}(S^A_j) = \frac{|S^A_j \cap E^T_{\text{train}}|}{|S^A_j|};\ j \in \{1, 2, \ldots, b_A\} \tag{14}$$

In other words, p^T_{trans} w.r.t A relates the similarity of node pairs in layer A to their probability of connection in layer T. Trans-layer connection probability of four duplexes is depicted in the left column of (Fig. 2). Moreover, the node pairs in S^A_j can be divided into two disjoint sets based on their connectivity in the auxiliary layer. Then the trans-layer connection probability for connected node pairs in auxiliary layer $S^A_j \cap E^A$ and unconnected ones $S^A_j \cap (U - E^A)$ will be:

$$\tilde{p}^T_{\text{trans}}(S^A_j \cap E^A) \tag{15}$$

and:

$$\tilde{p}^T_{\text{trans}}\left(S^A_j \cap (U - E^A)\right) \tag{16}$$

as shown in the middle and right columns of (Fig. 2), respectively.

The bars with dotted lines represent imputed trans-layer connection probabilities, similar to intra-layer connection probabilities in (Fig. 1). By inspecting the values of trans-layer connection probabilities for the datasets under study, a rising pattern is prominent by moving to bins corresponding to higher similarity ranges. Drosophila in (Fig. 2d1-3) brings up an exceptional case, where similarity in the auxiliary (Additive) layer shows no correlation with connection in the target (Suppressive) layer. Except these kind of irregularities in data, the available evidence appears to suggest that in most of the real multiplex networks, probability of connection in one (target) layer of the network does have positive correlation with similarity in some other (auxiliary) layer i.e., as similarity grows higher in the auxiliary layer, it can be a signal of higher connection probability in target layer. This observation develops the claim that for link prediction in target layer, not only the similarity of nodes in that same layer, but also their similarity in some other auxiliary layer can be utilized. Notice that this rising pattern in p_{trans} is observed in almost all datasets under scrutiny, independent from the choice of similarity measure.

The previously described property of trans-layer connection probability lies at the heart of the current study, shaping the main idea of the proposed multiplex link prediction method. In addition, the connectedness of the node pairs in the auxiliary layer leads to significant increase in the trans-layer connection probabilities. In Human Brain and Physicians networks the presence of link in the auxiliary is a strong evidence of connectivity in the target layer. The case is similar for AirTrain network but with lower certainty. The Drosophila network is an exception as before. These findings are in consistence with the link persistence phenomenon as reported in[29]. Here, we propose a consolidated method which considers the similarity of node pairs in the target and auxiliary layers, and also their connectedness in the auxiliary layer as the underlying evidences for calculating the uncertainty of linkage in the target layer.

Furthermore, by simultaneously partitioning U' based on their similarity in both target and auxiliary layers, we obtain $b_T \times b_A$ partitions or *2d-bins*. Within each 2d-bin, the fraction of target layer links to total node pairs is included i.e., the empirical connection probability in target layer is computed. In (Fig. 3), empirical probability of connection in 2d-bins is presented for the same duplexes as in (Fig. 2).

Several results can be inferred by scrutinizing (Fig. 3). Increment of the empirical probability of connection in the horizontal axis expresses the effectiveness of the similarity measure in target layer; the higher the bin number, the larger the fraction of node pairs that have formed links. Another aspect of the above figure is the ascension of the empirical probability of connection by moving to higher bin number in the auxiliary layer i.e., the vertical axis (except Drosophila in Fig. 3. d1-3), which is a sign of positive correlation between the probability of connection in target layer and similarity in the auxiliary layer; so far totally consistent with Figs. 1 and Fig. 2**.** This cross-layer connection and similarity correlation are observed in the majority of datasets under study, in which a subset of them is presented above. It is interesting that when similarity of a node-pair is very low in the target layer, high similarity in the auxiliary layer leads to stronger connection probability between them.

Figure 2. Empirical trans-layer connection probability in auxiliary layer bins. (**a1**–**d1**) Trans-layer connection probability of all node pairs, (**a2**–**d2**) Trans-layer connection probability of node-pairs connected in auxiliary layer, (**a3**–**d3**) Trans-layer connection probability of node-pairs unconnected in auxiliary layer, for sample duplexes of 4 datasets.

The following sub-sections are concerned with the issue of how to estimate probability of connection in the target layer of a multiplex network by incorporating other layers' structural information with a systematic approach that generalizes beyond specific data.

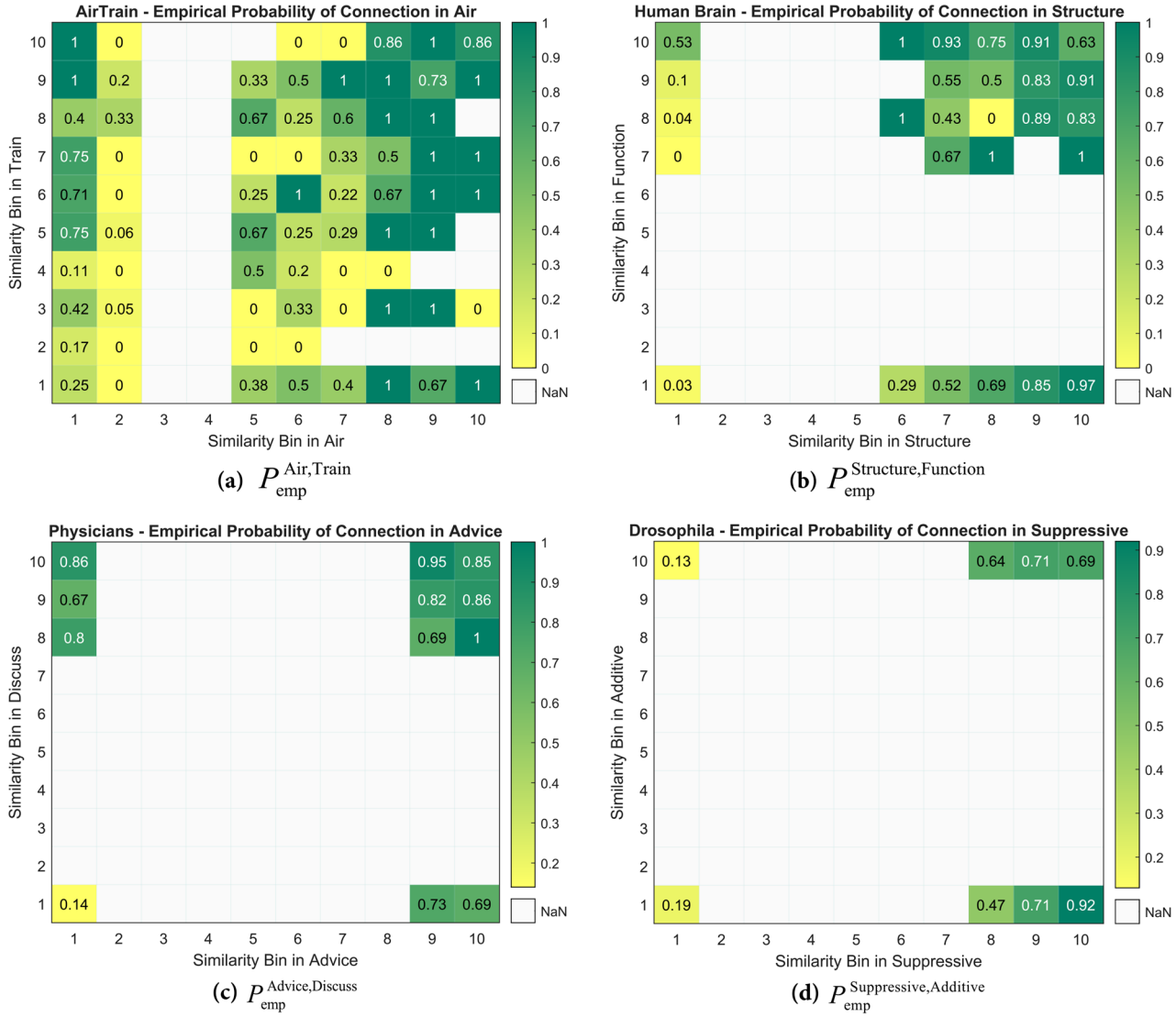

Figure 3. Empirical probability of connection in 2d-bins. The fraction of node pairs in the 2d-bins that are connected in the target layer(**a**) 'Train' of the network Air/Train w.r.t 'Air', (**b**) 'Function' of Human Brain w.r.t 'Structure', (**c**) 'Discuss' of Physicians w.r.t 'Advice', (**d**) 'Additive' of Drosophila **w.r.t** 'Suppressive' layer. NaN (Not a Number) values represent 2d-bins that contain no sample pairs.

Fusion of decisions. Consider two independent decision makers that determine the probability of occurrence of a certain event corresponding to a binary random variable. Each of them declares a probability p and q (where $0 \leq p, q \leq 1$) for the same event, respectively. One would want to reach to a consensus based on these two different opinions. This goal can be achieved by incorporating various functions that operate on input probabilities. The AND operator is one such function:

$$\mathrm{AND}(p, q) = pq \tag{17}$$

Another option could be the OR operator, defined as:

$$\mathrm{OR}(p, q) = p + q - pq \tag{18}$$

The more interesting function in the context of current research is the OR operator because it fits much better in the problem of link prediction as it is less prone to variations of only one of the input probabilities. We will return to the issue of fusion of decisions in the following sub-section when characterizing the link prediction model.

The multiplex link prediction model. On these grounds, a model is suggested to predict probability of connection between node pairs in a layer of the multiplex network such as T which incorporates information both from the layer itself and from some other auxiliary layer A. The similarity between two distinct nodes x and y is defined as:

$$SB_{xy}^{T,A} = -I(L_{xy}^{T} = 1|S_i^T, S_j^A);\ \ (x,y) \in S_i^T \cap S_j^A \tag{19}$$

where $I(L_{xy}^{T} = 1|S_i^T, S_j^A)$ is the uncertainty of existence of a link between (x, y) in the target layer when their target and auxiliary bin numbers are known. According to Eq. (5), we can write:

$$-I(L_{xy}^{T} = 1|S_i^T, S_j^A) = -I(L_{xy}^{T} = 1) + I(L_{xy}^{T} = 1; S_i^T, S_j^A) \tag{20}$$

The first term in Eq. (20) can be derived by incorporating Eq. (3):

$$-I(L_{xy}^{T} = 1) = \log p(L_{xy}^{T} = 1) \approx \log(\tilde{S}_{xy}^{T}) \tag{21}$$

where $\tilde{S}_{xy}^{T}$ is the min–max normalized similarity score of the pair (x, y) in target layer T i.e., the probability of connection in target layer (without any knowledge on bins partitioning) is estimated with similarity in that same layer, intuitively. The second term in Eq. (20) is the mutual information of (x, y) being connected in the target layer and belonging to S_i^T and S_j^A bins; which is estimated as follows:

$$I(L_{xy}^{T} = 1; S_i^T, S_j^A) \approx I(L^T = 1; S_i^T, S_j^A) \tag{22}$$

Equation (22) propounds the view that a group of node pairs dwelling in known target and auxiliary bins can be looked at similarly. To be more specific, if the goal is to obtain the mutual information between the event that (x, y) are connected and the event that it resides in both S_i^T and S_j^A, a possible workaround is to estimate it with the reduction in uncertainty of connection of *any* node pair due to which bins (target and auxiliary) it belongs to. Thus, according to Eq. (5), we proceed by expanding the right-hand side of Eq. (22):

$$I(L^T = 1; S_i^T, S_j^A) = I(L^T = 1) - I(L^T = 1|S_i^T, S_j^A) \tag{23}$$

The term $I(L^T = 1)$ in Eq. (23) is the self-information of that a randomly chosen node pair is linked in target layer T. Clearly, $I(L^T = 1)$ is the same for every node pair in the multiplex network; therefore, it does not affect the scoring (node pairs ranking), and it can be safely neglected. Thus, to carry out the model specification, $I(L^T = 1|S_i^T, S_j^A)$ needs to be calculated; which is the conditional self-information of that a randomly chosen node pair is linked in layer T when the pair's state of binning in target and auxiliary layer is known. Using Eq. (3) we have $I(L^T = 1|S_i^T, S_j^A) = \log p(L^T = 1|S_i^T, S_j^A)$. On the basis of our discussion on fusion of decisions, the probability $p(L^T = 1|S_i^T, S_j^A)$ for any randomly selected node pair (x, y) which is a member of $S_i^T \cap S_j^A$ is estimated by incorporating $p_{\text{intra}}(S_i^T)$ i.e. intra-layer connection probability in target layer T and $p_{\text{trans}}^T(S_j^A)$ i.e. trans-layer connection probability in T w.r.t auxiliary layer A. Therefore, similar to Eq. (18), the OR operation on intra and trans-layer connection probabilities concludes in:

$$\begin{aligned} p(L^T = 1|S_i^T, S_j^A) &= p_{\text{intra}}(S_i^T) + p_{\text{trans}}^T(S_j^A) - p_{\text{intra}}(S_i^T)p_{\text{trans}}^T(S_j^A) \\ &= \left[P_{\text{est}}^{T,A}\right]_{ij} \end{aligned} \tag{24}$$

It should be noticed that the trans-layer connection probability can be divided for connected and unconnected node pairs in the auxiliary layer according to Eqs. (15) and (16), respectively. To put it altogether, we incorporate Eqs. (15) and (16) into (24). Then, plugging Eq. (24) into Eq. (19) results in the final scoring scheme. Thus, SimBins similarity score of a node pair (x, y) in target layer T with the aid of auxiliary layer A where $(x, y) \in S_i^T \cap S_j^A;\ i \in \{1, \ldots, b_T\}, j \in \{1, \ldots, b_A\}$ and $T, A \in \{1, \ldots, M\};\ T \neq A$ is (empirical values of intra and trans-layer connection probabilities are used):

$$SB_{xy}^{T,A} = \begin{cases} \log(\tilde{S}_{xy}^{T}) + \log\left(\tilde{p}_{\text{intra}}(S_i^T) + \tilde{p}_{\text{trans}}^T(S_j^A \cap E^A) - \tilde{p}_{\text{intra}}(S_i^T)\tilde{p}_{\text{trans}}^T(S_j^A \cap E^A)\right) & ;\ \multimap(x,y) \in E^A \\ \log(\tilde{S}_{xy}^{T}) + \log\left(\tilde{p}_{\text{intra}}(S_i^T) + \tilde{p}_{\text{trans}}^T\left(S_j^A \cap (U - E^A)\right) - \tilde{p}_{\text{intra}}(S_i^T)\tilde{p}_{\text{trans}}^T\left(S_j^A \cap (U - E^A)\right)\right) & ;\ \multimap(x,y) \in U - E^A \end{cases} \tag{25}$$

Algorithm 1 outlines the entire scheme. Now that our multiplex scoring model is complete, we will proceed by evaluating the method on the datasets section introduced earlier.

Algorithm 1 Link Prediction and Evaluation using SimBins

input : Multiplex network $G(V, E^T, E^A; E^T, E^A \in U = V \times V)$
outputs: Similarity scores for test set based on the proposed method
MeanAUC

T *is the target layer and* ***A*** *is the auxiliary layer*;
Number of bins in the target and the auxiliary layer are set to $\boldsymbol{b_T}, \boldsymbol{b_A}$, *respectively*;
Number of evaluation iterations is initialized to ***iters***;
AUCVec *is initialized to an empty vector*;

for *iteration* $\leftarrow 1$ **to** *iters* **do**
  $E^T_{train}, E^T_{test} \leftarrow$ `DivideNetToTrainTest`$(E^T,\ TrainingRatio)$;
  $Z^T_{train} \leftarrow$ `Sample`$(U - E^T_{train}, size = 2|E^T_{train}|)$;
  $Z^T_{test} \leftarrow$ `Sample`$(U - E^T_{test}, size = 2|E^T_{test}|)$;
  $U' \leftarrow E^T_{train} \cup Z^T_{train}$;
  foreach *node-pair* (x, y) *in* U' **do**
    $sim^T_{xy} \leftarrow$ `BaseSimilarity`(x,y,T);
    $sim^A_{xy} \leftarrow$ `BaseSimilarity`(x,y,A);
  end
  $S^T_1, S^T_2, ..., S^T_{b_T} \leftarrow$ `EqualDepthBinning`$(U', [sim^T])$;
  $S^A_1, S^A_2, ..., S^A_{b_A} \leftarrow$ `EqualDepthBinning`$(U', [sim^A])$;
  foreach *partition* S^T_i **do**
    Compute $p_{intra}(S^T_i)$ according to equation (12);
  end
  foreach *partition* S^A_j **do**
    Compute $p^T_{trans}(S^A_j \cap E^A)$ and $p^T_{trans}(S^A_j \cap (U - E^A))$ according to equations (15) and (16), respectively;
  end
  foreach *2d-bin* (S^T_i, S^A_j) **do**
    Compute $[P^{T,A}_{est}]_{ij}$ according to equation (24);
  end
  foreach *node-pair* (x, y) *in* $E^T_{test} \cup Z^T_{test}$ *residing in 2d-bin* (S^T_i, S^A_j) **do**
    $SB^{T,A}_{x,y} \leftarrow$ according to equation (25);
  end
  The output of `CalcAUC`$(E^T_{test}, Z^T_{test}, SB^{T,A}_{x,y})$ is appended to **AUCVec**;
end
return $[SB^{T,A}], Mean(AUCVec)$;

The diagram in Fig. 4 illustrates the process of node-pairs similarity calculation in SimBins. The main source of information are the structure of the target and auxiliary layers. The train and test sets are derived from the target layer including both links and non-existent link (the test set is later used for evaluation). The rest of the process includes partitioning of the train set (U') according to the base similarity scores in T, A and connectedness in A. Accordingly, intra-layer and trans-layer connection probabilities of each partition (bin) is calculated and fed to the final SimBins scoring Eq. (25).

Experimental results. The link prediction performance on9 different datasets, a total of 29 network layers forming 52 layer-pairs has been reported based on both AUC (Table 2) and Precision (3) evaluation metrics. The evaluation metrics are the mean over 100 iterations with train ratio set to 90% as described in 'Evaluation Method' section. Four base measures comprising local, global and quasi-local indices have been incorporated i.e., RA, CN, ACT and LPI that were introduced in 'Base Similarity Measures' section. SimBins ($SB^A_T \equiv SB^{T,A}$) is

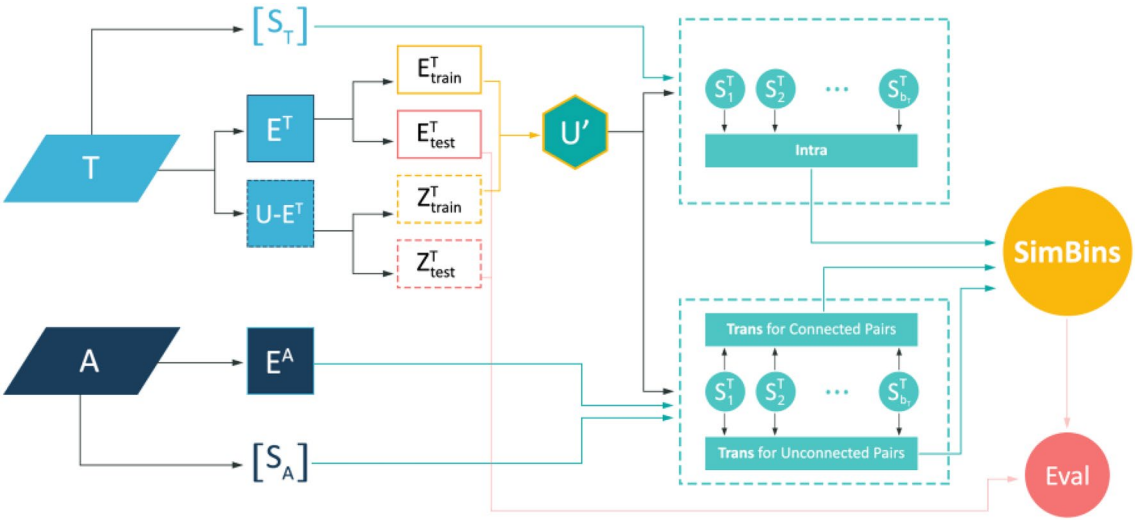

Figure 4. An overview of simbins method.

compared with baseline methods including scoring based on similarity in the target layer (S_T) and simple addition of similarity scores of the target and auxiliary layers ($S_T + S_A$).

In Table 2, for each base measure, the highest mean AUC is shown in bold and, for each duplex (all 52 rows), the highest AUC among all of the methods (independent from the base measure) is highlighted with an underscore. SimBins dominates other baseline methods and proves to be an effective multiplex link prediction method due to several reasons: (i) Most of the time, SimBins is superior to the other baseline methods (i.e., bold entries). This can be further verified with the fact that SimBins achieves higher average of all mean AUCs (the last row of the table) (ii) In a large fraction of duplexes (37 of 52), the overall best mean AUC belongs exclusively to SimBins (in 6 other duplexes, SimBins achieves the best performance alongside another method, non-exclusively) (iii) SimBins performs better than the single-layer method (or S_T) in most of the cases whereas for similarities addition method ($S_T + S_A$) this is less frequently observed; meaning our method is capable of using other layer's information effectively. And, $SB^{T,A}$ is more robust against deceptive signals compared to $S_T + S_A$. Consider Drosophila for example. The slightly negative correlation between similarity in the auxiliary layer (Suppressive) and connection probability in the target layer (Additive), as previously discussed on (Fig. 2-d), has caused performance reduction for $S_T + S_A$ whereas SimBins still performs as good as—if not better than—S_T. A similar outcome can be observed for NTN and London Transport, more clearly when ACT is used as the base similarity measure. In CS-Aarhus, where Facebook is the target layer, both S_T and $S_T + S_A$ perform even worse than random scoring (expected 50% AUC) while SimBins keeps the performance up about $70 - 80\%$. As the last row indicates, the average mean AUC of SimBins is higher than both other baseline methods, no matter the choice of base measure.

There exist occasions in which SimBins cannot improve the link prediction performance compared to the base similarity measure. Specifically, Drosophila which the absence of inter-layer correlation as discussed earlier is the underlying reason. And, in London Transport, node multiplexity is far too low as shown in Table 1. Consequently, very few nodes are shared among different layers that makes utilization of structural similarities between layers a hard task.

The above discussion holds true for Adamic-Adar[9], Preferential Attachment[8], and LRW[15] similarity measures, as we have performed similar experiments which led to resembling results, but we have avoided bringing the corresponding details for the sake of brevity.

Interestingly, the results appear to suggest that choosing LPI as the base similarity measure, leads to the best overall performance in most of the multiplex networks. Using LPI as the base similarity measure for SimBins gives the best performance with average mean AUC of 85.0% for all 52 duplexes under study.

The evaluation of methods based on Precision metric as reported in 3, confirms our earlier discussions. This metric measure quantifies the quality of top entries of the sorted list of unobserved links while AUC considers the quality of the ranking in the whole list. Here, also SimBins is superior compared to other two baseline methods. Specifically, in 38 duplexes out of 52 the best performance based on Precision metric is for SimBins while in 2 duplexes it shares the best performance with another baseline method. So, the results of Tables 2 and 3 confirm the superiority of SimBins over baseline methods regardless of the choice of base similarity measure and evaluation metric and also suggest that using SimBins along with LPI as the base similarity measure leads to the best performance.

Finally, we compare SimBins with three state-of-the-art methods, namely, YaoPL, YaoGL[38], and SameiHP[39]. An introduction to these methods is given in 'Related Works' section. The scoring schema of these methods can be summarized as Eq. (26). The base similarity measure used in these methods (S^T and S^A for the target and auxiliary layers T and A respectively) is LPI for the two former methods and HP for the latter. Moreover, the layer relevance measure ($\mu^{T,A}$)is PCC for YaoPL and GOR for YaoGL and SameiHP. Based on the recommendation of the authors, the parameter $\varphi = 0.5$ is considered. The results of the experiments are shown in Table 4.

	Target layer	Auxiliary layer	RA			CN			ACT			LPI		
			S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A
AT	Air	Train	83.9	89.9	**90.6**	79.8	**85.0**	84.9	87.7	85.9	**89.2**	80.1	**86.1**	82.8
	Train	Air	83.3	**84.0**	83.8	83.1	83.3	**84.2**	79.6	80.3	**80.9**	84.1	84.0	**84.8**
C. ELEGANS	Electric	Chem-Mono	70.6	79.0	**80.3**	70.6	78.5	**80.3**	64.7	65.8	**69.6**	76.6	82.4	**83.0**
		Chem-Poly	70.6	84.0	**85.6**	71.0	83.3	**85.9**	65.5	68.7	**72.5**	76.4	84.2	**85.9**
	Chem-Mono	Electric	76.2	77.0	**78.2**	75.8	76.3	**77.8**	67.3	67.8	**70.5**	84.3	83.9	**84.8**
		Chem-Poly	76.2	87.3	**90.8**	75.7	85.4	**91.7**	68.4	73.4	**89.0**	84.1	88.3	**89.9**
	Chem-Poly	Electric	85.8	85.9	**86.5**	84.0	83.9	**84.5**	**72.3**	72.0	**73.9**	**86.3**	86.1	**86.3**
		Chem-Mono	85.6	86.9	**88.7**	84.1	85.3	**87.6**	72.3	73.1	**81.9**	86.3	87.5	**87.6**
DM	Suppressive	Additive	76.5	75.9	**76.7**	**76.6**	75.7	**76.6**	**80.9**	74.3	77.2	**82.3**	81.2	**82.3**
	Additive	Suppressive	**74.2**	73.8	**74.2**	**73.9**	73.1	73.7	**73.6**	70.2	69.4	**79.5**	77.7	79.2
HB	Structure	Function	91.2	91.3	**92.9**	89.9	88.9	**91.9**	75.4	69.2	**78.6**	92.1	90.8	**94.2**
	Function	Structure	86.0	88.8	**89.9**	85.6	88.5	**89.9**	68.9	72.5	**79.9**	89.0	90.0	**91.0**
PHYSICIANS	Advice	Discuss	71.4	81.9	**87.3**	71.9	**82.6**	**88.7**	50.9	**66.3**	**77.0**	84.7	**93.7**	93.4
		Friendship	71.6	78.0	**81.3**	72.1	78.4	**81.8**	50.0	58.0	**62.2**	84.6	89.5	**89.6**
	Discuss	Advice	75.2	81.3	**87.2**	74.6	80.7	**87.3**	52.7	61.8	**74.1**	83.4	91.6	**91.7**
		Friendship	74.6	81.2	**84.6**	74.0	80.1	**84.8**	51.9	62.1	**67.9**	83.9	**90.5**	90.3
	Friendship	Advice	69.9	77.6	**80.9**	69.8	77.5	**81.2**	56.3	57.3	**66.9**	77.9	86.6	**87.1**
		Discuss	69.8	82.1	**86.0**	69.7	81.6	**86.7**	56.2	65.6	**72.8**	78.1	**89.9**	**89.9**
NTN	Communi	Financial	**84.2**	83.8	83.0	**82.7**	82.6	**82.7**	**74.8**	63.6	71.8	82.0	81.7	**82.8**
		Operation	84.3	84.3	**87.2**	82.6	82.9	**87.9**	75.0	68.0	**84.8**	82.4	82.1	**87.4**
		Trust	84.0	84.1	**89.4**	83.3	81.2	**88.9**	73.6	71.3	**82.6**	82.0	81.3	**86.8**
	Financial	Communi	91.5	**92.1**	90.7	**90.5**	78.6	90.0	52.7	40.6	**68.7**	**89.7**	77.8	87.9
		Operation	89.5	83.8	**90.2**	90.0	67.4	**92.1**	54.1	54.1	**67.5**	92.0	66.1	**92.6**
		Trust	91.7	92.7	**96.9**	90.2	79.3	**93.3**	50.6	41.0	**77.6**	93.0	83.4	**96.2**
	Operation	Communi	98.0	98.0	**98.8**	97.3	97.5	**98.2**	66.9	68.3	**81.4**	96.7	97.3	**97.8**
		Financial	**98.2**	97.9	**98.2**	**97.3**	**97.3**	97.2	67.1	58.8	**73.9**	96.7	96.7	**96.8**
		Trust	98.3	95.6	**98.7**	97.2	94.6	**97.7**	67.6	65.5	**78.7**	97.0	94.2	**97.7**
	Trust	Communi	88.5	92.4	**94.8**	87.7	91.6	**94.7**	78.2	80.3	**90.5**	88.6	**92.9**	92.6
		Financial	**88.5**	88.3	**88.5**	**87.5**	87.4	**87.5**	77.9	67.4	**80.6**	88.5	88.4	**88.7**
		Operation	88.6	88.3	**91.6**	88.1	86.9	**92.1**	78.3	71.3	**84.1**	88.3	85.5	**91.0**
LONDON TRANS	Tube	Over-ground	53.2	53.2	**55.0**	53.4	53.4	**55.0**	53.3	47.1	**61.0**	58.0	59.6	**59.9**
		DLR	**53.5**	53.4	**53.5**	**53.7**	**53.7**	**53.7**	**54.7**	50.4	50.1	57.7	57.6	**57.8**
	Over-ground	Tube	49.9	50.3	**55.6**	49.9	50.4	**56.0**	49.1	51.7	**81.5**	49.9	**55.3**	55.0
		DLR	**49.9**	**49.9**	**49.9**	50.0	49.9	**50.1**	49.7	48.9	**56.0**	**49.9**	49.7	**49.9**
	DLR	Tube	52.8	**53.2**	50.4	53.0	**53.6**	49.8	56.5	58.5	**64.3**	52.3	**53.2**	**53.2**
		Over-ground	**52.2**	**52.2**	49.8	**52.8**	52.7	50.4	**57.9**	57.4	50.6	**53.0**	52.8	**53.0**
CS-AARHUS	Lunch	Facebook	94.7	93.3	**94.8**	**94.7**	91.0	**94.7**	83.5	61.5	**84.0**	93.9	89.5	**94.5**
		Co-author	**95.4**	95.3	95.3	**93.5**	93.4	93.4	83.4	56.0	**83.6**	94.2	94.1	**94.4**
		Leisure	94.5	94.2	**94.9**	94.0	93.9	**94.4**	82.8	68.7	**85.8**	93.5	93.2	**93.9**
		Work	94.7	94.7	**95.5**	93.8	93.3	**95.3**	84.1	82.3	**88.1**	94.6	92.9	**95.9**
	Facebook	Lunch	93.5	90.5	**93.8**	**92.8**	90.3	92.5	43.6	51.6	**78.8**	95.0	91.3	**95.3**
		Co-author	92.5	92.1	**92.9**	93.2	93.1	**93.6**	42.7	47.3	**74.7**	94.7	94.6	**94.8**
	Co-author	Lunch	73.0	**92.2**	89.7	69.1	**91.5**	91.2	45.6	58.9	**72.0**	73.3	92.0	**94.8**
		Facebook	72.9	70.5	**79.6**	69.8	66.2	**73.6**	43.1	62.2	**68.6**	73.3	71.9	**81.2**
	Leisure	Lunch	82.8	**90.5**	90.2	81.4	89.2	**89.7**	58.9	75.1	**81.8**	81.7	89.1	**89.5**
	Work	Lunch	88.1	91.0	**91.3**	86.2	**89.9**	**89.9**	71.6	**83.2**	82.0	85.4	**89.4**	**89.4**

Continued

	Target layer	Auxiliary layer	RA			CN			ACT			LPI		
			S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A
SACCH-POMB	Direct	Colocalization	62.9	64.3	**65.8**	62.9	64.3	**65.7**	51.7	50.1	**68.2**	73.1	**75.9**	**75.9**
		Physical	62.9	71.8	**76.8**	63.0	70.1	**76.5**	51.8	52.6	**75.9**	74.0	82.2	**85.5**
		Synthetic	62.8	69.8	**70.9**	62.8	69.6	**70.9**	50.9	50.2	**69.3**	73.3	**80.4**	**80.4**
		Association	63.1	63.7	**64.3**	63.0	63.7	**64.2**	52.0	51.6	**72.8**	73.4	**74.6**	**74.6**
	Physical	Direct	77.8	78.4	**79.4**	77.4	78.0	**79.0**	69.4	57.8	**76.3**	88.4	**89.5**	88.6
	Synthetic	Direct	80.3	81.5	**82.2**	80.3	81.5	**82.3**	65.2	54.2	**82.8**	90.9	**92.1**	**92.1**
	AVERAGE AUC		*78.5*	*81.0*	***82.8***	*77.8*	*79.4*	***82.4***	*63.7*	*62.9*	***75.1***	*81.6*	*82.9*	***85.0***

Table 2. Average AUC over 100 iterations for the networks under study. Each row shows the performance of link prediction methods on a duplex subset of a multiplex network grouped by the corresponding base similarity measure in use. Columns show the average AUC over 100 iterations for the prediction methods S_T (similarity score of only the target layer), $S_T + S_A$ (addition of similarity scores of the target and auxiliary layer), $\mathrm{SB}_T^A \equiv \mathrm{SB}^{T,A}$ (SimBins).

$$S_{x,y}^{T,A} = (1-\varphi)S_{x,y}^{T} + \varphi\mu^{T,A}S_{x,y}^{A}m \tag{26}$$

Clearly, SimBins achieves the best performance (85.0%) in term of average mean AUC over all 52 duplexes. Also, in 25 duplexes SimBins is the best performing method (the best in 18 cases and sharing the best performance in 7 cases with another method) while the second best is SameiHP with the best performance in 13 duplexes. It should be also noted that SameiHP method has large fluctuation across different networks and the lowest average mean AUC. So, using SimBins based on LPI is our choice that performs well across diverse set of multiplex networks.

Complexity analysis. Consider a duplex network $G(V, E^{[1]}, E^{[2]}; E^{[i]} \subseteq V \times V), m_i = |E^{[i]}| \forall i \in \{1,2\}$ where layer 1 is the target, and layer 2 is the auxiliary layer. Let $O(\theta)$ be a representative of computational complexity for the base similarity measures. The similarity of node pairs in both layers is needed for subset U' of $U = V \times V$ as formulated in 'Partitioning Node Pairs (Binning)' section. Therefore, the computing complexity of measuring similarities is $O(\sum_{i=1,2} \theta m_i)$. Partitioning U' into equal-depth bins requires sorting of similarities, consequently it would have complexity of $O(\sum_{i=1,2} m_i \log m_i)$. Total estimation complexity of intra-layer and trans-layer connection probabilities is $O\left(\sum_{i=1,2} m_i b_i\right)$ where b_i is the number of bins in corresponding layer. And, estimation of probability of connection in all 2d-bins according to Eq. (24) would be of order$O(b_1 b_2)$ which is negligible w.r.t bounded number of bins. Accordingly, the total computational complexity of scoring a node pair in SimBins would be $O(m \log m)$ where m is in the same order as m_1, m_2 if the sparsity of multiplex layers is comparable. This tolerable computing complexity indicates that SimBins can be scaled for usage in large networks.

Notice that for obtaining a full ranking of propensity of links, SimBins, like the majority of link prediction algorithms would need at least $O(n^2)$; $n = |V|$ computations which is not easily scalable to very large networks without pruning the n^2 space. To be specific, for a full ranking, SimBins would have a computing complexity of $O(\theta n^2 + m \log m)$ in which $O(\theta n^2)$ is the dominating term in real-networks; meaning that SimBins imposes minor overhead to the base similarity measures. This makes SimBins appropriate for using with large networks like SaccPomb that we studied in this paper.

Discussion

In this manuscript, we explored the intra-layer and trans-layer connection probabilities in multiplex networks and verified that in many real multiplex networks, connection probability within an arbitrary layer is correlated with similarity in other layers of the same multiplex. We also observe that connectedness in one layer of the multiplex, increases the probability of linkage in other layers. Subsequently, we developed a consolidated link prediction model by incorporating information theory concepts for characterizing intuitions gathered from the observed evidences.

The proposed method works on a pair of multiplex's layers i.e., a duplex. Different ideas can be conducted to extend it to use multiple layers' topology for link prediction. Considering a target layer T and auxiliary layers $A_1, \ldots, A_M$, the simplest idea is to add up the SimBins scores for each possible layer pairs, symbolically $\mathrm{SB}^{T,\{A_1,\ldots,A_M\}} = \sum_{i=1}^{M} \mathrm{SB}^{T,A_i}$ where SB^{T,A_i} is computed according to Eq. (25). The other—not as straightforward as previous– idea is to compose and study bins of more than two dimensions. This extension, although more systematic, might suffer from heavy sparsity of samples (imagine node pairs residing in 3d-bins).

Eventually, SimBins is compared with two baseline methods (base similarity measure in the target layer and simple addition of similarities in target and auxiliary layers) and three state-of-the-art methods (YaoPL, YaoGL and SameiHP) on 9 multiplexes. It is shown that SimBins outperforms the other two baseline methods in most cases. Besides, it rarely performs worse than target similarity and is more robust to deceptive signals compared to the simple addition of similarities. It is mentioned that in some networks, such as London Transport and Drosophila, SimBins seems to be unprofitable as a result of massively condensed node pairs similarity distribution

	Target layer	Auxiliary layer	RA			CN			ACT			LPI		
			S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A
AT	Air	Train	71.8	**74.3**	73.8	61.4	65.3	**68.8**	72.1	71.7	**73.5**	68.8	**69.9**	69.1
	Train	Air	65.8	66.3	**66.4**	58.2	60.9	**64.4**	60.2	61.9	**62.0**	65.9	66.6	**67.0**
C. ELEGANS	Electric	Chem-Mono	42.7	**67.2**	66.9	17.7	47.5	**63.6**	48.9	47.9	**53.3**	55.3	64.9	**67.1**
		Chem-Poly	43.2	67.6	**69.4**	18.4	58.2	**68.4**	49.3	49.9	**55.5**	55.4	65.7	**67.4**
	Chem-Mono	Electric	60.7	64.4	**66.1**	31.4	38.7	**63.4**	48.8	49.7	**51.1**	68.3	65.0	**68.7**
		Chem-Poly	60.5	72.8	**75.6**	31.7	64.0	**80.7**	50.0	56.5	**77.5**	67.9	**72.0**	69.6
	Chem-Poly	Electric	72.9	72.6	**73.3**	61.2	55.3	**64.1**	54.1	53.4	**55.6**	**69.9**	68.8	**69.9**
		Chem-Mono	72.4	73.3	**74.0**	60.7	62.7	**68.1**	54.5	55.0	**64.7**	70.1	71.2	**72.1**
DM	Suppressive	Additive	54.9	56.1	**56.3**	30.5	31.5	**56.7**	**64.2**	57.0	60.3	**70.3**	68.6	**70.5**
	Additive	Suppressive	48.6	**53.0**	50.8	24.6	26.8	**50.5**	**57.3**	53.4	53.5	60.6	63.5	**66.2**
HB	Structure	Function	77.6	76.0	**79.7**	58.7	60.6	**76.1**	55.1	48.9	**62.3**	74.7	75.1	**81.3**
	Function	Structure	69.9	72.8	**74.8**	55.1	65.7	**73.6**	51.5	53.4	**63.3**	73.2	73.8	**75.6**
PHYSICIANS	Advice	Discuss	44.2	67.0	**78.2**	14.6	43.3	**78.7**	32.3	45.9	**63.4**	66.0	80.4	**81.4**
		Friendship	43.9	59.2	**68.3**	14.2	32.6	**68.8**	31.1	39.5	**43.6**	66.0	75.9	**76.7**
	Discuss	Advice	51.8	66.2	**78.3**	15.3	42.0	**77.1**	34.3	43.4	**58.4**	64.5	79.3	**79.8**
		Friendship	50.6	65.5	**74.2**	14.6	33.0	**73.7**	33.8	45.5	**50.7**	65.1	77.2	**77.8**
	Friendship	Advice	39.0	58.3	**66.9**	13.6	35.4	**67.2**	38.1	39.8	**49.7**	51.6	73.2	**73.8**
		Discuss	38.7	66.7	**76.0**	12.6	37.0	**75.5**	38.2	46.0	**56.8**	51.9	**77.8**	77.6
NTN	Communi	Financial	**68.5**	68.0	**68.5**	**56.9**	**56.9**	**56.9**	**56.3**	45.0	53.3	64.1	63.8	**65.8**
		Operation	69.5	68.2	**72.0**	59.0	64.1	**69.8**	56.7	50.0	**67.7**	64.8	67.2	**71.5**
		Trust	67.9	67.8	**70.9**	60.1	58.5	**71.2**	54.5	48.8	**65.7**	65.4	62.6	**70.0**
	Financial	Communi	9.5	**36.0**	26.0	0.0	18.0	**23.0**	8.0	6.5	**12.0**	0.0	**25.5**	24.0
		Operation	11.5	31.0	**35.5**	0.0	16.5	**21.5**	8.5	12.0	**16.0**	0.0	**17.0**	15.0
		Trust	13.5	**30.0**	29.5	0.0	**18.5**	**18.5**	7.0	10.0	**21.5**	0.0	**29.5**	20.5
	Operation	Communi	90.7	90.0	**91.6**	84.8	85.1	**88.8**	48.4	51.0	**64.2**	87.4	87.5	**89.4**
		Financial	90.6	90.3	**90.7**	84.8	84.8	**84.9**	48.3	37.3	**53.9**	**87.6**	**87.6**	**87.6**
		Trust	91.0	84.4	**91.5**	84.8	77.1	**88.2**	48.9	42.3	**61.0**	88.2	80.2	**89.4**
	Trust	Communi	76.8	**81.8**	79.6	70.0	75.0	**78.0**	62.3	61.8	**76.4**	74.8	**79.8**	78.6
		Financial	**77.4**	77.1	**77.4**	71.0	70.6	**71.1**	**61.8**	44.7	61.6	**75.2**	75.0	75.0
		Operation	76.7	73.0	**78.1**	71.2	67.9	**75.7**	62.5	49.4	**66.9**	73.8	71.4	**76.0**
LONDON TRANS	Tube	Overground	4.3	4.3	**10.8**	0.3	0.3	**10.7**	**33.4**	31.5	16.1	8.1	11.6	**21.0**
		DLR	4.3	4.4	**7.7**	0.1	0.1	**8.0**	**34.3**	31.3	12.8	7.3	7.3	**16.7**
	Overground	Tube	0.0	0.0	**13.3**	0.0	0.0	**14.0**	17.0	27.1	**56.3**	0.0	4.3	**12.1**
		DLR	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	16.6	**18.3**	12.0	**0.0**	**0.0**	**0.0**
	DLR	Tube	0.0	0.4	**3.6**	0.0	0.0	**3.6**	14.4	20.2	**30.8**	0.0	2.0	**8.4**
		Overground	0.0	0.0	**2.2**	0.0	0.0	**3.2**	15.6	**20.0**	10.6	0.0	0.0	**6.0**
CS-AARHUS	Lunch	Facebook	83.8	78.2	**84.4**	75.5	64.7	**80.0**	61.8	37.4	**63.7**	80.9	70.8	**82.6**
		Co-author	**85.1**	84.6	85.0	74.9	74.5	**75.7**	61.6	38.5	**62.5**	81.2	81.0	**81.5**
		Leisure	83.1	82.8	**84.3**	75.8	75.9	**81.6**	61.4	47.1	**66.1**	**81.7**	81.6	**83.0**
		Work	83.7	82.6	**85.0**	75.6	72.3	**84.0**	62.7	62.4	**71.7**	82.0	78.5	**84.7**
	Facebook	Lunch	79.2	71.1	**80.3**	71.0	66.6	**77.4**	12.7	23.6	**58.1**	77.5	71.6	**79.9**
		Co-author	77.2	75.6	**77.7**	71.5	71.6	**72.1**	10.0	20.5	**50.8**	77.1	76.7	**77.7**
	Co-author	Lunch	14.7	**56.3**	53.3	2.0	46.0	**51.3**	17.3	27.3	**35.3**	13.7	54.7	**58.3**
		Facebook	14.7	38.3	**49.0**	1.3	24.0	**43.7**	18.7	28.0	**30.3**	10.3	33.7	**49.7**
	Leisure	Lunch	60.4	**73.9**	72.2	36.2	67.9	**70.0**	33.9	54.8	**66.7**	61.7	**71.6**	70.0
	Work	Lunch	71.0	**73.5**	73.2	58.3	67.1	**68.5**	57.0	**64.7**	63.6	65.2	**72.1**	70.7

Continued

	Target layer	Auxiliary layer	RA			CN			ACT			LPI		
			S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A	S_T	$S_T + S_A$	SB_T^A
SACCH-POMB	Direct	Colocalization	24.6	27.6	**31.6**	8.3	10.5	**31.5**	22.8	19.2	**55.0**	39.9	45.2	**51.7**
		Physical	24.5	47.2	**58.1**	8.5	25.2	**58.5**	23.6	21.9	**58.7**	40.8	70.1	**71.5**
		Synthetic	24.5	38.9	**42.1**	8.2	20.2	**42.2**	21.1	18.4	**55.2**	39.6	56.4	**61.6**
		Association	25.1	26.4	**28.5**	9.0	9.8	**28.4**	23.1	23.7	**51.5**	40.1	42.0	**49.5**
	Physical	Direct	53.2	54.5	**64.6**	28.7	30.1	**65.5**	50.7	32.6	**58.1**	81.3	**81.8**	81.4
	Synthetic	Direct	60.2	62.5	**64.6**	38.3	41.1	**64.7**	40.7	22.0	**63.7**	78.3	80.8	**84.5**
	AVERAGE PRECISION		*50.5*	*57.3*	***60.6***	*36.2*	*44.6*	***58.1***	*40.5*	*39.8*	***52.4***	*52.1*	*58.8*	***62.0***

Table 3. Average Precision over 100 iterations for the networks under study. Each row shows the performance of link prediction methods on a duplex of a multiplex network grouped by the corresponding base similarity measure in use. Columns show the average Precision over 100 iterations for the prediction methods S_T (similarity score of only the target layer), $S_T + S_A$ (addition of similarity scores of the target and auxiliary layer), $\mathrm{SB}_T^A \equiv \mathrm{SB}^{T,A}$ (SimBins).

and negative inter-layer correlations. On the other hand, when comparing with the state-of-the-art methods, it is observed that the overall best average AUC belongs to SimBins and it performs consistently well across various multiplex networks. This can be attributed to the design of the proposed method in which incorporates information both from connectedness and similarity of nodes in different layers.

It is shown that SimBins imposes negligible computation overhead to the base similarity measures (as we applied the method on a large network with a few thousand nodes and edges like SacchPomb, with minor computational burden). The idea of using an equal-width strategy for partitioning node pairs leads to even more efficiency due to its $O(m)$ complexity (instead of $O(m \log m)$ in equal-depth binning), although the accuracy of prediction might be affected.

Because our method falls under the structural similarity category, it may not beat learning-based approaches that are of higher computational complexity. As discussed earlier in this section, extending SimBins to use similarities in multiple layers simultaneously can be further explored as a future direction. The proposed method integrates intra-layer structural similarities and connectedness in the auxiliary layers in a systematic way; it is proved to boost the performance of link prediction in multiplex networks while maintaining a low computational complexity.

			SIMBINS-LPI	STATE-OF-THE-ART ($\varphi = 0.5$)		
	Target layer	Auxiliary layer	SB_T^A	YaoPL	YaoGL	SameiHP
AT	Air	Train	82.8	86.7	86.3	**88.2**
	Train	Air	**84.8**	82.6	84.1	79.2
C. ELEGANS	Electric	Chem-Mono	83.0	**84.2**	83.7	75.3
		Chem-Poly	**85.9**	85.6	84.4	78.2
	Chem-Mono	Electric	**84.8**	83.6	84.1	73.7
		Chem-Poly	**89.9**	87.9	88.7	74.5
	Chem-Poly	Electric	86.3	**86.8**	86.4	70.4
		Chem-Mono	87.6	**87.9**	87.4	70.5
DM	Suppressive	Additive	82.3	**82.7**	82.2	77.8
	Additive	Suppressive	**79.2**	78.6	79.0	77.9
HB	Structure	Function	**94.2**	93.2	93.5	68.8
	Function	Structure	91.0	90.8	**91.6**	78.2
PHYSICIANS	Advice	Discuss	**93.4**	92.3	92.8	82.8
		Friendship	**89.6**	89.1	88.5	81.9
	Discuss	Advice	91.7	90.6	**92.0**	79.7
		Friendship	**90.3**	90.1	**90.3**	80.7
	Friendship	Advice	**87.1**	86.5	**87.1**	78.5
		Discuss	89.9	**90.1**	89.9	81.1
NTN	Communi	Financial	**82.8**	79.5	82.4	72.4
		Operation	**87.4**	85.9	84.1	71.4
		Trust	**86.8**	84.4	82.4	71.5
	Financial	Communi	87.9	**93.3**	87.5	80.8
		Operation	92.6	87.6	**95.0**	85.4
		Trust	**96.2**	93.3	94.3	81.2
	Operation	Communi	**97.8**	97.3	97.5	64.3
		Financial	**96.8**	96.4	**96.8**	63.1
		Trust	**97.7**	97.0	97.0	62.4
	Trust	Communi	**92.6**	92.1	**92.6**	64.2
		Financial	**88.7**	87.8	**88.7**	65.1
		Operation	91.0	**91.6**	89.1	66.6
LONDON TRANS	Tube	Overground	59.9	59.6	60.5	**68.2**
		DLR	57.8	57.6	57.7	**69.0**
	Overground	Tube	55.0	56.1	55.2	**77.1**
		DLR	49.9	50.0	49.8	**81.6**
	DLR	Tube	53.2	50.9	50.5	**82.6**
		Overground	53.0	53.2	52.7	**82.1**
CS-AARHUS	Lunch	Facebook	94.5	**95.0**	93.4	80.4
		Co-author	**94.4**	**94.4**	94.4	74.6
		Leisure	93.9	94.0	**94.2**	77.2
		Work	**95.9**	94.6	95.6	78.1
	Facebook	Lunch	**95.3**	94.9	94.6	78.9
		Co-author	**94.8**	**94.8**	94.1	75.7
	Co-author	Lunch	**94.8**	92.0	91.7	76.4
		Facebook	**81.2**	79.0	76.1	78.0
	Leisure	Lunch	89.5	**89.9**	86.8	81.0
	Work	Lunch	89.4	89.7	**90.4**	80.8

Continued

	Target layer	Auxiliary layer	SIMBINS-LPI	STATE-OF-THE-ART ($\varphi = 0.5$)		
			SB_T^A	YaoPL	YaoGL	SameiHP
SACCHPOMB	Direct	Colocalization	75.9	75.6	76.6	**93.0**
		Physical	85.5	82.2	81.7	**93.0**
		Synthetic	80.4	80.9	81.5	**93.0**
		Association	74.6	74.6	75.4	**93.1**
	Physical	Direct	88.6	89.3	89.2	**96.5**
	Synthetic	Direct	92.1	92.0	92.3	**95.2**
	AVERAGE AUC		***85.0***	*84.5*	*84.5*	*77.9*

Table 4. Comparison of average AUC over 100 iterations for the networks under study with state-of-the-art methods. Performance evaluation of the link prediction methods on 52 real-world duplex networks based on AUC measure. Three left columns determine the name of the multiplex networks, the target layer of link prediction, and the auxiliary layer which comes to help the prediction task. From left to right, the evaluated methods are SimBins using LPI as base similarity measure: Our proposed method, YaoPL: a state-of-the-art method that utilizes LPI with PCC as the layer relevance measure, YaoGL: LPI with PCC as the layer relevance measure, SameiHP: a state-of-the-art method that utilizes hyperbolic distance as dissimilarity measure within each layer with GOR as the layer relevance measure. **Bold and underlined** are the best results in each row.

Received: 4 December 2020; Accepted: 10 June 2021
Published online: 24 June 2021

References

1. Liben-Nowell, D., Kleinberg, J. The link prediction problem for social networks. Proceedings of the twelfth international conference on Information and knowledge management. ACM: New Orleans. p. 556–559 (2003).
2. Clauset, A., Moore, C. & Newman, M. E. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**(7191), 98–101. https://doi.org/10.1038/nature06830 (2008).
3. Guimera, R. & Sales-Pardo, M. Missing and spurious interactions and the reconstruction of complex networks. *Proc. Natl. Acad. Sci. U S A.* **106**(52), 22073–22078. https://doi.org/10.1073/pnas.0908366106 (2009).
4. Li, X. & Chen, H. Recommendation as link prediction in bipartite graphs: A graph kernel-based machine learning approach. *Decis. Support Syst.* **54**(2), 880–890. https://doi.org/10.1016/j.dss.2012.09.019 (2013).
5. Esslimani, I., Brun, A. & Boyer, A. Densifying a behavioral recommender system by social networks link prediction methods. *Soc. Netw. Anal. Min.* **1**(3), 159–172. https://doi.org/10.1007/s13278-010-0004-6 (2011).
6. Yadav, A., Singh, Y. N. & Singh, R. R. Improving routing performance in AODV with link prediction in mobile adhoc networks. *Wireless Pers. Commun.* **83**(1), 603–618. https://doi.org/10.1007/s11277-015-2411-5 (2015).
7. Lin, D. An Information-Theoretic Definition of Similarity. Proceedings of the Fifteenth International Conference on Machine Learning. 657297: Morgan Kaufmann Publishers Inc.; 1998. p. 296–304.
8. Chen, H., Li, X., Huang, Z. (eds) Link prediction approach to collaborative filtering. Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '05); 2005 7–11 June 2005.
9. Adamic, L. A. & Adar, E. Friends and neighbors on the Web. *Soc. Netw.* **25**(3), 211–230. https://doi.org/10.1016/S0378-8733(03)00009-1 (2003).
10. Zhou, T., Lü, L. & Zhang, Y.-C. Predicting missing links via local information. *Eur. Phys. J. B.* **71**(4), 623–630. https://doi.org/10.1140/epjb/e2009-00335-8 (2009).
11. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43. https://doi.org/10.1007/BF02289026 (1953).
12. Brin, S., Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Seventh International World-Wide Web Conference (WWW 1998); Brisbane, Australia1998.
13. Lü, L., Jin, C.-H. & Zhou, T. Similarity index based on local paths for link prediction of complex networks. *Phys. Rev. E* **80**(4), 046122. https://doi.org/10.1103/PhysRevE.80.046122 (2009).
14. Liu, W. & Lü, L. Link prediction based on local random walk. *EPL (Europhys. Lett.).* **89**(5), 58007. https://doi.org/10.1209/0295-5075/89/58007 (2010).
15. Menon, A. K. & Elkan, C. (eds) *Link Prediction via Matrix Factorization* (Springer, 2011).
16. Lü, L. & Zhou, T. Link prediction in complex networks: A survey. *Phys. A* **390**(6), 1150–1170. https://doi.org/10.1016/j.physa.2010.11.027 (2011).
17. Tan, F., Xia, Y. & Zhu, B. Link prediction in complex networks: A mutual information perspective. *PLoS ONE* **9**(9), e107056. https://doi.org/10.1371/journal.pone.0107056 (2014).
18. Xu, Z., Pu, C. & Yang, J. Link prediction based on path entropy. *Phys. A* **456**, 294–301. https://doi.org/10.1016/j.physa.2016.03.091 (2016).
19. Zhu, B. & Xia, Y. An information-theoretic model for link prediction in complex networks. *Sci. Rep.* **5**(1), 13707. https://doi.org/10.1038/srep13707 (2015).
20. Zhu, B. & Xia, Y. Link Prediction in Weighted Networks: A Weighted Mutual Information Model. *PLoS ONE* **11**(2), e0148265. https://doi.org/10.1371/journal.pone.0148265 (2016).
21. Lichtenwalter, R.N., Lussier, J.T., Chawla, N.V. New perspectives and methods in link prediction. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining; Washington, DC, USA. 1835837: ACM; 2010. p. 243–252.
22. Grover, A. & Leskovec, J. node2vec: Scalable feature learning for networks. *KDD* **2016**, 855–864. https://doi.org/10.1145/2939672.2939754 (2016).
23. Velickovic, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D. (eds) Deep Graph Infomax. ICLR (Poster); 2019.
24. Park, C., Han, J. & Yu, H. Deep multiplex graph infomax: Attentive multiplex network embedding using global information. *Knowl.-Based Syst.* **197**, 105861. https://doi.org/10.1016/j.knosys.2020.105861 (2020).

25. Wang, P., Xu, B., Wu, Y. & Zhou, X. Link prediction in social networks: The state-of-the-art. *SCIENCE CHINA Inf. Sci.* **58**(1), 1–38. https://doi.org/10.1007/s11432-014-5237-y (2015).
26. Kivelä, M. *et al.* Multilayer networks. *J. Compl. Netw.* **2**(3), 203–271. https://doi.org/10.1093/comnet/cnu016 (2014).
27. Aleta, A. & Moreno, Y. Multilayer networks in a nutshell. *Ann. Rev. Condens. Matter Phys.* **10**(1), 45–62. https://doi.org/10.1146/annurev-conmatphys-031218-013259 (2019).
28. Kleineberg, K.-K., Boguñá, M., Ángeles Serrano, M. & Papadopoulos, F. Hidden geometric correlations in real multiplex networks. *Nat. Phys.* **12**, 1076. https://doi.org/10.1038/nphys3812 (2016).
29. Papadopoulos, F. & Kleineberg, K.-K. Link persistence and conditional distances in multiplex networks. *Phys. Rev. E* **99**(1), 012322. https://doi.org/10.1103/PhysRevE.99.012322 (2019).
30. Sun, Y., Han, J., Yan, X., Yu, P. S. & Wu, T. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proc. VLDB Endowment.* **4**(11), 992–1003 (2011).
31. Shi, C., Kong, X., Huang, Y., Philip, S. Y. & Wu, B. Hetesim: A general framework for relevance measure in heterogeneous networks. *IEEE Trans. Knowl. Data Eng.* **26**(10), 2479–2492. https://doi.org/10.1109/TKDE.2013.2297920 (2014).
32. Shakibian, H. & Moghadam, C. N. Mutual information model for link prediction in heterogeneous complex networks. *Sci. Rep.* **7**, 44981. https://doi.org/10.1038/srep44981 (2017).
33. 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.* **5**(1), 24. https://doi.org/10.1140/epjds/s13688-016-0087-z (2016).
34. Jalili, M., Orouskhani, Y., Asgari, M., Alipourfard, N. & Perc, M. Link prediction in multiplex online social networks. *R Soc Open Sci.* **4**(2), 160863. https://doi.org/10.1098/rsos.160863 (2017).
35. Najari, S., Salehi, M., Ranjbar, V. & Jalili, M. Link prediction in multiplex networks based on interlayer similarity. *Phys. A Stat. Mech. Appl.* https://doi.org/10.1016/j.physa.2019.04.214 (2019).
36. Pujari, M. & Kanawati, R. Link prediction in multiplex networks. *Netw. Heterogen. Med.* **10**, 17–35. https://doi.org/10.3934/nhm.2015.10.17 (2015).
37. Hajibagheri, A., Sukthankar, G., Lakkaraju, K. A holistic approach for predicting links in coevolving multiplex networks. Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; Davis, California. 3192628: IEEE Press; 2016. p. 1079–1086.
38. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* **99**(12), 7821–7826. https://doi.org/10.1073/pnas.122653799 (2002).
39. Samei, Z. & Jalili, M. Application of hyperbolic geometry in link prediction of multiplex networks. *Sci. Rep.* **9**(1), 12604. https://doi.org/10.1038/s41598-019-49001-7 (2019).
40. Abdolhosseini-Qomi, A. M., Yazdani, N. & Asadpour, M. Overlapping communities and the prediction of missing links in multiplex networks. *Phys. A* **554**, 124650. https://doi.org/10.1016/j.physa.2020.124650 (2020).
41. Abdolhosseini-Qomi, A. M. *et al.* Link prediction in real-world multiplex networks via layer reconstruction method. *R Soc Open Sci.* **7**(7), 191928. https://doi.org/10.1098/rsos.191928 (2020).
42. Davis, D., Lichtenwalter, R., Chawla, N.V., editors. Multi-relational Link Prediction in Heterogeneous Information Networks. 2011 International Conference on Advances in Social Networks Analysis and Mining; 2011 25–27 July 2011.
43. Aleta, A., Tuninetti, M., Paolotti, D., Moreno, Y. & Starnini, M. Link prediction in multiplex networks via triadic closure. *Phys. Rev. Res.* **2**(4), 042029. https://doi.org/10.1103/PhysRevResearch.2.042029 (2020).
44. Xiao Fan, W. & Guanrong, C. Complex networks: Small-world, scale-free and beyond. *IEEE Circ. Syst. Mag.* **3**(1), 6–20. https://doi.org/10.1109/MCAS.2003.1228503 (2003).
45. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747 (1982).
46. Halu, A., Mukherjee, S. & Bianconi, G. Emergence of overlap in ensembles of spatial multiplexes and statistical mechanics of spatial interacting network ensembles. *Phys. Rev. E* **89**(1), 012806. https://doi.org/10.1103/PhysRevE.89.012806 (2014).
47. Chen, B. L., Hall, D. H. & Chklovskii, D. B. Wiring optimization can relate neuronal structure and function. *Proc. Natl. Acad. Sci. U.S.A.* **103**(12), 4723–4728. https://doi.org/10.1073/pnas.0506806103 (2006).
48. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539. https://doi.org/10.1093/nar/gkj109 (2006).
49. De Domenico, M., Nicosia, V., Arenas, A. & Latora, V. Structural reducibility of multilayer networks. *Nat. Commun.* **6**, 6864. https://doi.org/10.1038/ncomms7864 (2015).
50. Simas, T., Chavez, M., Rodriguez, P. R. & Diaz-Guilera, A. An algebraic topological method for multimodal brain networks comparisons. *Front. Psychol.* **6**, 904. https://doi.org/10.3389/fpsyg.2015.00904 (2015).
51. Coleman, J., Katz, E. & Menzel, H. The diffusion of an innovation among physicians. *Sociometry.* **20**(4), 253–270. https://doi.org/10.2307/2785979 (1957).
52. Battiston, F., Nicosia, V. & Latora, V. Structural measures for multiplex networks. *Phys. Rev. E* **89**(3), 032804. https://doi.org/10.1103/PhysRevE.89.032804 (2014).
53. De Domenico, M., Solé-Ribalta, A., Gómez, S. & Arenas, A. Navigability of interconnected networks under random failures. *Proc. Natl. Acad. Sci.* **111**(23), 8351–8356. https://doi.org/10.1073/pnas.1318469111 (2014).
54. Magnani, M., Micenkova, B., Rossi, L. Combinatorial analysis of multiple networks. arXiv preprint arXiv:1303.4986. 2013.
55. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x (1948).
56. Cover, T.M., Thomas, J.A. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing): Wiley-Interscience; 2006.
57. Fouss, F., Pirotte, A., Renders, J. & Saerens, M. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Trans. Knowl. Data Eng.* **19**(3), 355–369. https://doi.org/10.1109/TKDE.2007.46 (2007).
58. Martínez, V., Berzal, F. & Cubero, J.-C. A Survey of Link Prediction in Complex Networks. *ACM Comput. Surv.* **49**(4), 69. https://doi.org/10.1145/3012704 (2016).
59. Lü, L., Pan, L., Zhou, T., Zhang, Y.-C. & Stanley, H. E. Toward link predictability of complex networks. *Proc. Natl. Acad. Sci. U.S.A.* **112**(8), 2325–2330. https://doi.org/10.1073/pnas.1424644112 (2015).
60. Garcia, D. Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Stat. Data Anal.* **54**(4), 1167–1178. https://doi.org/10.1016/j.csda.2009.09.020 (2010).

Acknowledgements

We express our thanks to Dr. Behnam Bahrak for reviewing the manuscript and providing helpful comments and insights.

Author contributions

S.H.J. abd A.M.A. conceived the original idea. A.M.A. and S.H.J analyzed the data. A.M.A. and S.H.J. did the mathematical modeling. S.H.J. performed the coding. A.M.A. and S.H.J. designed experiments. A.M.A. and

S.H.J. conducted the experiments. S.H.J. wrote the paper. M.A., M.R. and N.S. supervised the study. All authors did proof-reading and commenting.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to S.H.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.