Chapter 6

CLP-MUL: Clustering-based Link Prediction in Multiplex Networks

In the previous chapter, an attempt was made to merge local information-based edge quantification and global information-based node centralities. This centrality calculation process involves possible discerning paths in a graph which is a computationally expensive operation for large graphs. In order to improve upon this issue, in this chapter, we attempt to use community detection to enhance link prediction in multiplex networks. Community detection in social networks minimizes local intra-community similarity and maximizes global inter-community information. Hence, such methods follow the quasi-local information template, which attempts to find a trade-off between local and global information. Community-based link prediction on multiplex networks ($CLP-MUL$) is based on the fact that even if the types of links between nodes may change depending on the specific layer of a multiplex network, the nodes represent the same entities and share some fundamental structures (communities). Some communities may change from layer to layer, but there exist more rigid communities (superimposed on the whole network structure) whose influence is felt across layers. In this chapter, it is suggested that nodes belonging to such a community have a greater likelihood of having connections between them (in various layers) and that this may be exploited for more accurate link prediction.

6.1 Introduction

When all the information contained in different layers is considered in the context of multiplex networks, communities/clusters common across layers of the given network can be identified. The inspiration for the $CLP-MUL$ method is based on the fact that even though some communities may change from layer to layer, there exist other more rigid communities (superimposed on the whole network structure) whose influence is felt across layers. The basic argument in this chapter is that the nodes belonging to such a community have a higher chance of having multiple links between them (in different layers for different interaction types), and this fact can exploit for more efficient link prediction.

The main motivations behind the $CLP-MUL$ method are (1) creating a method for community detection in multiplex networks that uses information from all layers; and (2) using these superimposed communities for link prediction in a manner that is not layer-dependent. This chapter presents a link prediction method based on clustering/community detection on multiplex networks. An additional advantage of the $CLP-MUL$ approach is that after performing community detection and link prediction on a single weighted graph, the link prediction probabilities can directly be transferred to specific layers without any added processing, using a simple scalar multiplication of probability matrix.

6.1.1 Information Diffusion

In this framework, an algorithm is used to detect network communities, which combines the flow of information with detachability and label propagation. Information diffusion

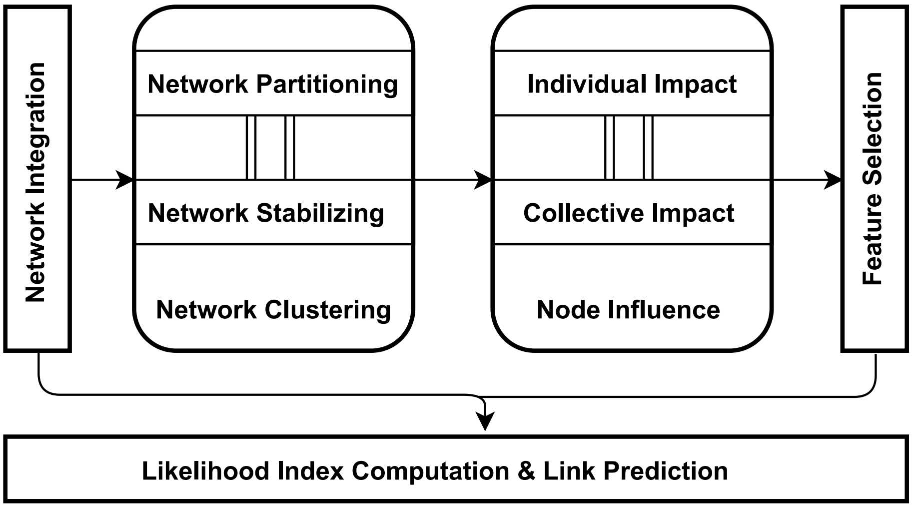


FIGURE 6.1: $CLP-MUL$ Framework

can be defined as the ”study of how a piece of information is propagated in a network through edges and how nodes can influence each other” [75]. Information diffusion helps bring out communities' shape within the network by maximizing influence spread. A probabilistic method to model the diffusion process as a Markov random field was proposed by Domingos and Richardson [166]. Kempe et al. [167] were the first to reformat this problem based on information cascade and linear threshold models. Other concepts such as sub-modularity [168] and influence paths [169] has also been used for enhancing the information diffusion process. A recent survey of information diffusion-based community detection was undertaken by Das and Biswas [170].

6.2 Proposed Work

This section discusses the proposed framework of $CLP-MUL$, which takes multi-interaction networks and uses a clustering-based link prediction framework to predict links in the layers of these multiplex networks. The proposed framework is shown in Figure 6.1. First, the algorithm forms a weighted summarized network from multi-interaction networks then it performs network clustering using information diffusion [162]. It then computes the individual and collective impact of node x to their neighbors y. After this, the algorithm calculates the features (common relevant nodes) for final likelihood computation. Finally, the algorithm computes the likelihood score of each non-existing link to predict missing links. The framework consists of the following modules -

6.2.1 Network Integration

Multi-interaction (multiplex) networks are merged into a single weighted network in the network integration phase. To create such a weighted network, a topological information-based strategy for layer integration is developed, which is also beneficial in implementing any single weighted network link prediction algorithm. Therefore, the weight of an edge between nodes $x\&y$, i.e., $A_M(x,y)$ can be computed using Equation 6.1 (where n is the number of layers in multiplex network).

$$A_M(x,y) \leftarrow \frac{1}{n}\sum_{j=1}^{n}\{A_j | A_j = a_{xy}^j\} \tag{6.1}$$

where (E_j is the edge set of j^{th} layer of multiplex network),

$$a_{xy}^j \leftarrow \begin{cases} 1 & if\ \exists(x,y) \in E_j, j \in [1,n] \\ 0 & otherwise \end{cases} \tag{6.2}$$

6.2.2 Network Clustering

To incorporate the collective impact of an individual on another, a network clustering method [162] using information diffusion and trust formation is utilized to identify community structure. This method works in two phases: partitioning and stabilization.

- **Partitioning.** In this phase, the transformed multiplex network (weighted simple graph) is divided into sub-networks based on information propagation. This depends on trustingness and trustworthiness along with the diffusion model. Trustingness τ_z^p measures the propensity of an individual z to trust someone while trustworthiness τ_{xz}^I measures the user's x outlook about others z. A node x belongs to a community $C_L(x)$ in which node x has the maximum influence (l is the total number of communities). An individual's influence is calculated under the independent cascade model (here $N.C_j$ are neighbors of node x, which belong to the same community k with label C_j).

$$C_L(x) \leftarrow argmsmax_{1 \leq k \leq l}\{1 - \prod_{z \in N.C_k} (1 - \tau_z^p \times \tau_{xz}^I)\} \tag{6.3}$$

- **Stabilization.** Now, the algorithm stabilizes the community structure using a stabilizing index. Therefore, unstable communities will be merged with more suitable and stable communities. Stabilizing index is the ratio of intra-influence to the overall influence of a community, given as follows.

$$S_I(C_i) \leftarrow \frac{\sum_{x,y \in C_i} \tau_x^p \times \tau_{xy}^I}{\sum_{x,y \in C_i} \tau_x^p \times \tau_{xy}^I \times \sum_{x \in C_i}^{y \notin C_i} \tau_x^p \times \tau_{xy}^I} \tag{6.4}$$

6.2.3 Individual Impact Calculation

Some studies suggest that the influence of an individual is limited to their local area, such as small world phenomena, three-degree theory, etc. The three-degree theory [171] states that a node x influence is not propagated beyond the three-hop area. Therefore, an individual influence will decrease exponentially with respect to the number of hops. So, an individual impact $I_I(x,y)$ of x to y can be computed by Equation 6.5 (here $d_s(x,y)$ is shortest distance between $x\&y$).

$$I_I(x,y) \leftarrow \begin{cases} 1, & if \quad d_s(x,y) \leq 1 \\ (\tau^I_{xz} \times \tau^p_x + \tau^I_{zy} \times \tau^p_z) \times \exp^{1-d_s}, & if \quad d_s(x,y) = 2 \\ (\tau^I_{xz_1} \times \tau^p_x + \tau^I_{z_1z_2} \times \tau^p_{z_1} + \tau^I_{z_2y} \times \tau^p_{z_2}) \times \exp^{1-d_s}, & if \quad d_s(x,y) = 3 \\ 0, & if \quad d_s(x,y) \geq 4 \end{cases} \tag{6.5}$$

6.2.4 Measuring Collective Impact

Nodes that are more closely related to each other have more influence on each other. This relationship can be quantified using different approaches, such as the number of common neighbors, shortest path length between nodes, etc. In the community-based approach, a node is more closely related to nodes belonging to the same community than inter-community nodes. So, a node has more influence on nodes that belong to the same community. The collective impact (x,y) of a node x on y can be defined as follows (here α is the weightage assigned to a node pair when they belong to the same community and $C_L(x), C_L(y)$ are the community labels for nodes $x \& y$ which is calculated in network clustering in Section 6.2.2). α has a fixed value of 1 in this chapter.

$$C_I(x,y) \leftarrow \begin{cases} \alpha & \textit{if } C_L(x) = C_L(y) \\ 0 & \textit{otherwise} \end{cases} \tag{6.6}$$

6.2.5 Feature Selection

Now, the proposed algorithm identifies the feature set $\gamma(n_1,n_2)$ for each non-existing edge (n_1,n_2) based on topological features. There are different topological features which are utilized, the proposed variations being *CLP_M_CN*, *CLP_M_PA*, *CLP_M_CAR*, *CLP_M_CC*, defined as follows. This feature set will be used to calculate the final likelihood score between nodes which are not directly connected to each other.

1. **Common Neighbors (CN)**. Nodes in the actual world are heavily clustered locally due to small world phenomena, meaning they are more likely to be connected if they have more common neighbours [16]. The common neighbors feature set $\gamma(n_1,n_2)$ for a non-existing pair (n_1,n_2) is defined as,

$$\gamma(x,y) \leftarrow \{z|z \in \{N(n_1) \cap N(n_2)\}\} \tag{6.7}$$

where $N(n_1)$ and $N(n_2)$ denotes the neighbors of node n_1 and n_2 respectively.

2. **Preferential Attachment (PA)**. [90] formulated that nodes with more connections overall were more likely to get new connections. Therefore, the preferential attachment feature set $\gamma(n_1,n_2)$ for a non-existing pair (n_1,n_2) is defined as,

$$\gamma(n_1,n_2) \leftarrow \{z|z \in \{N(n_1) \cup N(n_2)\}\} \tag{6.8}$$

3. **CAR**. [20] stated that nodes which belong to the same local community are more likely to have a connection. CAR attempts to find connections within the common neighbors themselves. Essentially two common neighbors of nodes $n_1 \& n_2$ ehich have common connections amongst themselves are considered as part of this feature set. Therefore, CAR feature set $\gamma(n_1,n_2)$ for a non-existing pair (n_1,n_2) is defined as follows.

$$\gamma(n_1,n_2) \leftarrow \{z|z \in \{N(n_1) \cap N(n_2) \cap N(z')\}\} \tag{6.9}$$

where z' is another common neighbor of nodes $n_1 \& n_2$,

$$z' \leftarrow \{N(n_1) \cap N(n_2)\} \tag{6.10}$$

4. **Clustering Coefficient (CC)**. The degree to which nodes tend to cluster together is measured by the clustering coefficient [42]. Therefore, The clustering coefficient

feature set $\gamma(x,y)$ for a non-existing pair (n_1,n_2) is defined as the set of neighbor nodes that tend to form triangles.

$$\gamma(n_1,n_2) \leftarrow \{z|z \in \Delta_m\} \tag{6.11}$$

where Δ_m denotes set of nodes which forms triangles passing through node m, where $m \in \{N(n_1) \cap N(n_2)\}$.

6.2.6 Likelihood Score Computation

The likelihood can be determined by individual and collective impact of a node onto other. The likelihood score $L_I(x,y)$ of non-existing link x,y to predict missing link can be computed as follows (here $\gamma(x,y)$ is calculated using different measures as shown in Section 6.2.5).

$$\begin{gathered} L_I(x,y) \leftarrow \sum_{z \in \gamma(x,y)} \frac{C_I(x,z) \times I_I(x,z) + C_I(z,y) \times I_I(z,y)}{\sum_{w \in N(z)} C_I(z,w) \times I_I(z,w)} \\ iff \sum_{w \in N(z)} C_I(z,w) \times I_I(z,w) \neq 0 \end{gathered} \tag{6.12}$$

6.2.7 $CLP-MUL$ Algorithm with an illustrative example

The **Algorithm 4** takes multi interaction networks as input and produce likelihood scores of non-existing links to predict missing links. Line 2 generates a summarized and weighted multiplex network representation from multi interaction networks of common nodes using topological coupling. Line 4 identifies the community structure based on independent cascade propagation model using trustingness and trustworthiness of an individual. The algorithm, by applying lines 6-8, iteratively computes importance of a node and community. Line 7 computes individual impact to others using three degree theory. Line 8 computes collective impact of an individual corresponding to associated

Algorithm 4: $CLP-MUL$: Clustering-based Link Prediction in Multiplex Network

Input: Social Networks: $G_i(V_i, E_i)$
Output: Likelihood Index: L_I

1 ▷ Network Integration
2 Create a multiplex network A_M from n different interaction networks on same user set using $A_M(x,y) \leftarrow \frac{1}{n}\sum_{j=1}^{n} A_j$
3 ▷ Network Clustering
4 Identify community structure based on independent cascade propagation model using propensity to trust and influence probabilities
5 ▷ Individual & Collective Impact Computation
6 **for** *each edge* $(x,y) \in E$ **do**
7   Compute individual influence of x to others y using $I_I(x,y)$
8   Compute collective influence of x to y using $C_I(x,y)$
9 ▷ Feature Selection
10 **for** *each non-existing edge* $(u,v) \notin E$ **do**
11   Select features set of a pair of individuals based on CN, PA, CAR, & CC
12 ▷ Likelihood Index Computation
13 **for** *each non-existing link* $(u,v) \notin E$ **do**
14   Compute likelihood score of each non-existing pair (u,v) using $L_I(u,v)$
15 **Return** L_I;

community. The algorithm, by applying lines 10-11, iteratively obtains the features of a pair of individuals using CN, PA, CAR, and CC. The loop in lines 13-14 computes the likelihood score of each non-existing link to predict missing links. Finally, line 15 returns the likelihood index of non-existing links.

To explain the working of $CLP-MUL$, a multi interaction network consisting of graphs G_1, G_2, and G_3 is used for demonstration purposes as shown in Figure 6.2. Firstly, algorithm generates a summarized multiplex network representation by integrating multi interaction networks and computes connection strength as $A_M(x,y) \leftarrow \frac{1}{n}\sum_{j=1}^{n} A_j$. For examples, connection strength of (A,B) in multiplex can be calculated as $A_M(A,B) \leftarrow \frac{1}{3}(1+1+1) = 1$. Individual impact and collective impact both will be 1 because A and B have direct connection to each other and belongs to same community. Similarly, we can compute for other existing edges as shown in Table 6.1. After computing individual impact and collective impact for existing edges, the likelihood for

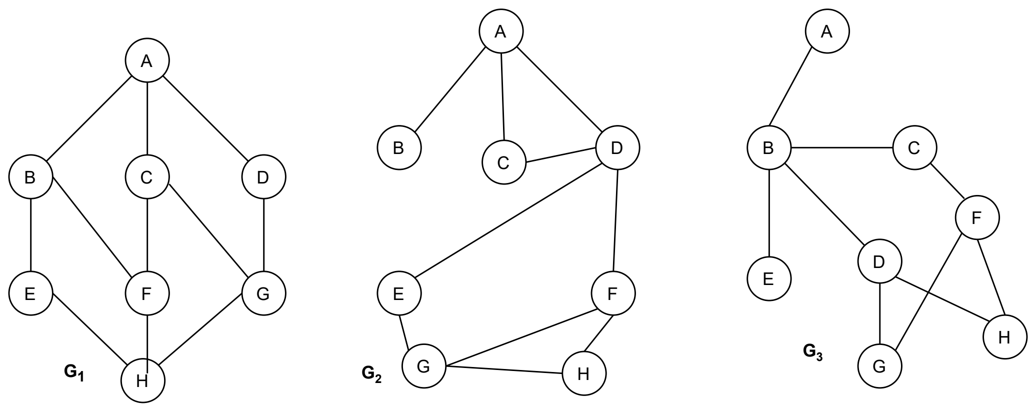


FIGURE 6.2: Example Graph for $CLP-MUL$ framework

non-existing edges to predict missing links is computed. For example, non-existing link (A,E) can compute individual and collective impact as 0.588 and 1 by Equation 6.5 and 6.6 respectively. Next feature set is evaluated using CN, PA, CAR, and CC. Therefore. $CN_z = N(A) \cap N(E = \{B,C,D\} \cap \{B,H,D\}) = \{B,D\}$ and $L_I - CN = 0.163$. Similarly, likelihood scores for other non-existing edges can be computed (as shown in Table 6.1).

6.2.8 Complexity Analysis

In this section, the time complexity of the proposed algorithm $CLP-MUL$ is analyzed. Here D_{avg} is the average degree of graph and C_{avg} is the average clustering coefficient of graph. Line 1 generates multiplex network in $\mathcal{O}(V+E)$ time. Line 2 identify the community structure of multiplex network using label propagation based clustering [75] in $\mathcal{O}(D_{avg}(E+\tau V+l_c^2 C_{avg}))$ time. The for loop in lines 3-5 computes individual and collective impact in $\mathcal{O}(VD_{avg}+E)$ and $\mathcal{O}(1)$ time respectively. The for loop in lines 6-7 computes common neighbors feature set in $\mathcal{O}(VD_{avg}+E)$ time. Finally, algorithm in lines 8-9 computes likelihood score of non-existing links in $\mathcal{O}(E)$ time. The combined complexity of $CLP - MUL$ approach would be $\mathcal{O}(V+E+D_{avg}(E+\tau V+l_c^2 C_{avg})+2(VD_{avg}+E)+E)$. Taking the most significant term into account, overall time complexity of $CLP-MUL$ is $\mathcal{O}(D_{avg}E)$.

TABLE 6.1: The computation of likelihood score of (x,y) under $CLP - MUL$ for Example Fig.6.2

$x-y$	$A_M(x,y)$	$I_I(x,y)$	$C_I(x,y)$	CN_z	$L_I - CN$	PA_z	$L_I - PA$	CAR_z	$L_I - CAR$	CC_z	$L_I - CC$
A-B	1	1	1	-	-	-	-	-	-	-	-
A-C	0.66667	1	1	-	-	-	-	-	-	-	-
A-D	0.66667	1	1	-	-	-	-	-	-	-	-
A-E	-	0.58893	1	B,D	0.16327	B,C,D,G,H	0.52468	B,D	0.16327	A,B,C,D,E,F,G,H	0.72136
A-F	-	0.35281	1	B,C,D	0.2449	B,C,D,G,H	0.56881	B,C,D	0.2449	A,B,C,D,E,F,G,H	0.73935
A-G	-	0.17404	1	C,D	0.16327	B,C,D,E,F,H	0.63058	C,D	0.16327	A,B,C,D,E,F,G,H	0.72174
A-H	-	0.12445	1	D	0.08163	B,C,D,E,F,G	0.45255	-	0	A,B,C,D,E,F,G,H	0.65637
B-C	0.33333	1	1	-	-	-	-	-	-	-	-
B-D	0.33333	1	1	-	-	-	-	-	-	-	-
B-E	0.66667	1	1	-	-	-	-	-	-	-	-
B-F	0.33333	1	1	-	-	-	-	-	-	-	-
B-G	-	0.1336	1	C,D,E,F	0.45714	A,C,D,E,F,H	0.70207	C,D,E,F	0.45714	A,B,C,D,E,F,G,H	0.78187
B-H	-	0.06861	1	D,E,F	0.34286	A,C,D,E,F,G	0.51142	D,E,F	0.34286	A,B,C,D,E,F,G,H	0.71649
C-D	0.33333	1	1	-	-	-	-	-	-	-	-
C-E	-	0.35628	1	B,D,G	0.34286	A,B,D,F,G,H	0.68638	B,D,G	0.34286	A,B,C,D,E,F,G,H	0.7954
C-F	0.66667	1	1	-	-	-	-	-	-	-	-
C-G	0.33333	1	1	-	-	-	-	-	-	-	-
C-H	-	0.06861	1	D,F,G	0.34286	A,B,D,E,F,G	0.52364	D,F,G	0.34286	A,B,C,D,E,F,G,H	0.73041
D-E	0.33333	1	1	-	-	-	-	-	-	-	-
D-F	0.33333	1	1	-	-	-	-	-	-	-	-
D-G	0.66667	1	1	-	-	-	-	-	-	-	-
D-H	0.33333	1	1	-	-	-	-	-	-	-	-
E-F	-	0.23649	1	B,D,G,H	0.57143	B,C,D,G,H	0.63561	B,D,G,H	0.57143	A,B,C,D,E,F,G,H	0.77232
E-G	0.33333	1	1	-	-	-	-	-	-	-	-
E-H	0.33333	1	1	-	-	-	-	-	-	-	-
F-G	0.66667	1	1	-	-	-	-	-	-	-	-
F-H	0.66667	1	1	-	-	-	-	-	-	-	-
G-H	0.66667	1	1	-	-	-	-	-	-	-	-

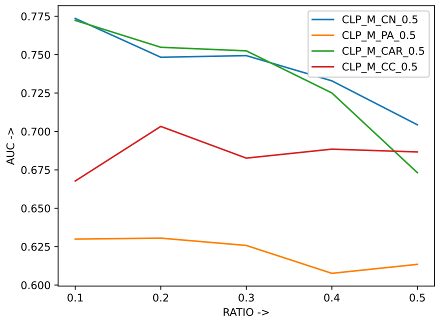
6.3 Performance Analysis

6.3.1 Feature Set Comparison

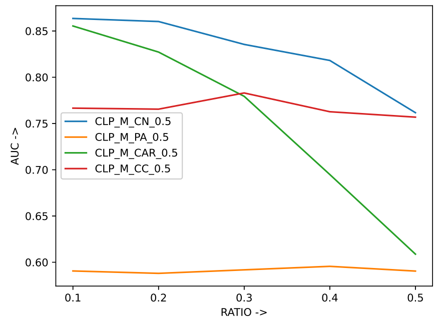
The relationship between the algorithm's performance based on different feature sets (*CLP_M_CN*, *CLP_M_PA*, *CLP_M_CAR*, *CLP_M_CC*) is investigated in this section. Three metrics in these experiments: AUC, F1 Score and Balanced Accuracy Score have been used. Five different ratios (0.1, 0.2, 0.3, 0.4, 0.5) are also used where each ratio is of testing set edges to total edges of graph datasets while the remaining ones are considered training data sets. Each of these tests is performed on six real-world networks. For the sake of simplicity in performing comparisons between different feature sets, the value of α at 1 is fixed. This helps in streamlining the results such that only differences caused due to change of feature sets can be measured.

6.3.1.1 AUC Pattern among different feature sets

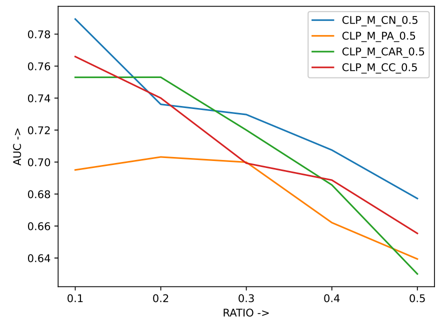
Fig. 6.3 presents the comparison of different feature sets on six datasets. In first five datasets it is observed that *CLP_M_CN* is either the best performing algorithm or it narrowly misses the best position. The only sizable difference is observed in the dataset Xenopus-Genetic where *CLP_M_PA* is the best performing algorithm. But this is not a general pattern as in all other datasets as the PA based variation can be considered the worst performing one. *CLP_M_CAR* is the second best performing algorithm in 4 datasets. The exceptions are CKM-Physicians-Innovation and Xenopus-Genetic where it becomes close to the worst performing algorithm. *CLP_M_CC* can be considered to be algorithm with the most middle-of-the-pack performance. The exception is CKM-Physicians-Innovation where it performs just worse than *CLP_M_CN*.



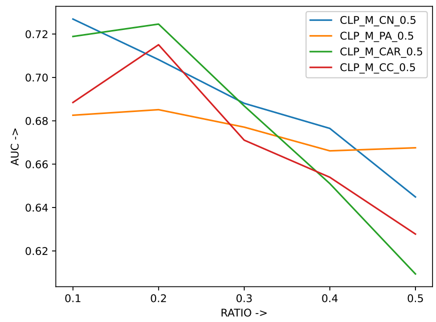
(A) Lazega-Law-Firm



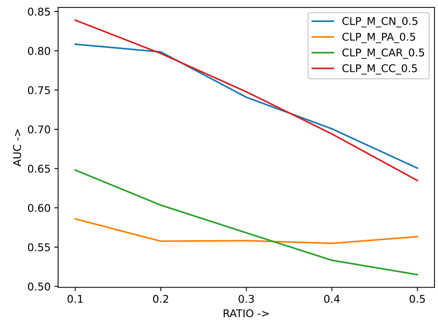
(B) CS-Aarhus



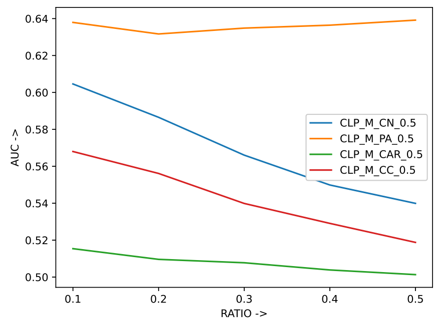
(C) Vickers-Chan-7thGraders



(D) Kapferer-Tailor-Shop

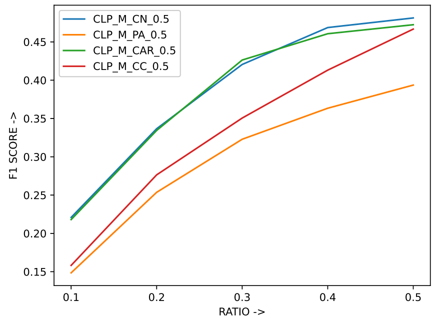


(E) CKM-Physicians-Innovation

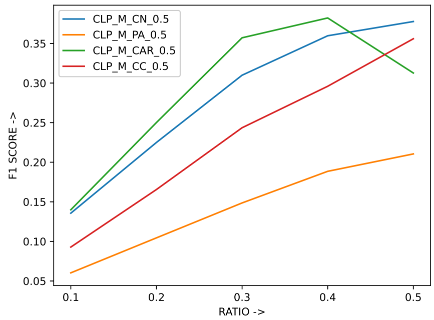


(F) Xenopus-Genetic

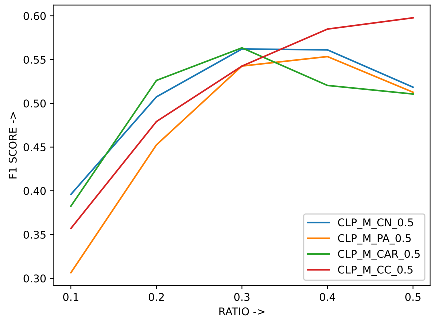
FIGURE 6.3: AUC comparison of $CLP-MUL$ algorithm for different feature sets on datasets



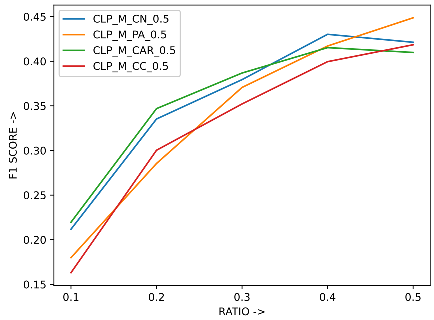
(A) Lazega-Law-Firm



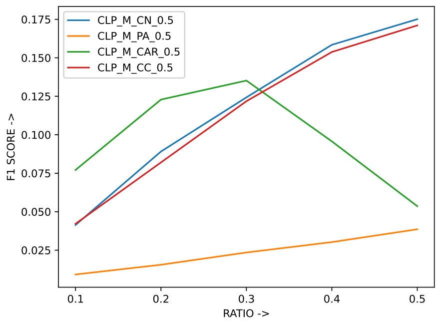
(B) CS-Aarhus



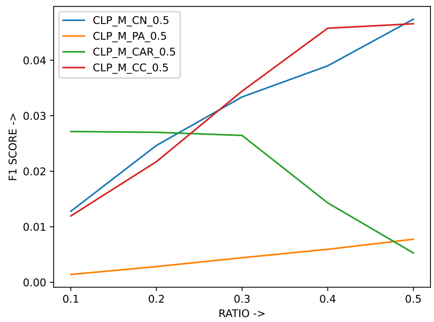
(C) Vickers-Chan-7thGraders



(D) Kapferer-Tailor-Shop



(E) CKM-Physicians-Innovation



(F) Xenopus-Genetic

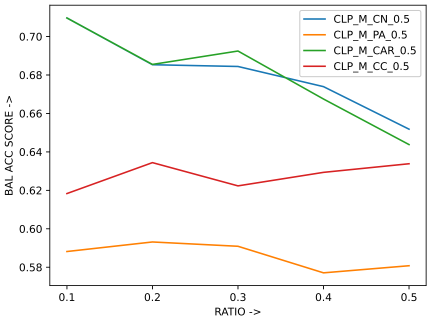
FIGURE 6.4: F1 Score comparison of $CLP-MUL$ algorithm for different feature sets on datasets

6.3.1.2 F1 Score Pattern among different feature sets

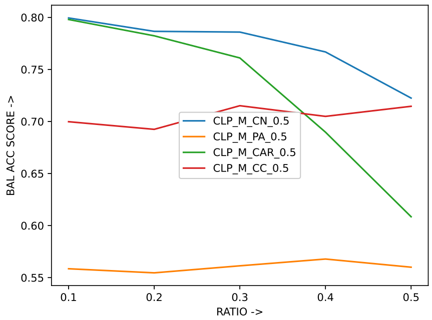
Fig. 6.4 presents the comparison of different feature sets on six datasets. *CLP_M_CN* is observed to be one of the best performing algorithm across all datasets. The overall pattern of variation is increase of F1 score as the training edge set becomes smaller and the testing edge set becomes bigger i.e., increase in *Ratio* variable. But it is evident that the quantum of increase of F1 score decreases with the increase of *Ratio* variable. *CLP_M_CAR* shows the most erratic behavior unlike the other three algorithms especially in CKM-Physicians-Innovation and Xenopus-Genetic dataset. In these cases it doesn't follow the gradual increasing order pattern. *CLP_M_PA* can be seen to be the worst performing algorithm in all six datasets while *CLP_M_CC* can be considered to be the middle-of-the-pack one.

6.3.1.3 Balanced Accuracy Score Pattern among different feature sets

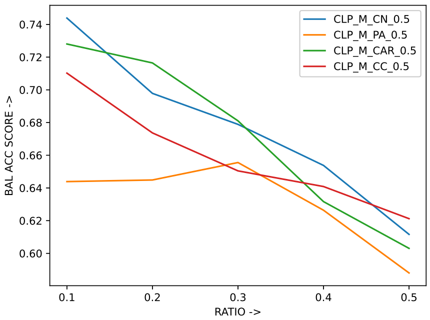
Fig. 6.5 presents the comparison of different feature sets on six datasets. *CLP_M_CN* can be seen to be either the best perming algorithm or its performance is very close to the others. The only exception is Xenopus-Genetic dataset where *CLP_M_PA* is the best performing one and shows a minuscule increase in performance. This is contrary to the gradual decreasing pattern in performance followed by other algorithms. But in all other datasets, *CLP_M_PA* is the worst performing algorithm. *CLP_M_CC* shows a middle-of-the-pack performance in all datasets except Kapferer-Tailor-Shop and CKM-Physicians-Innovation. In Kapferer-Tailor-Shop it is the worst performing algorithm while in CKM-Physicians-Innovation it is the best performing one. *CLP_M_CAR* shows very good performance in all datasets except in CKM-Physicians-Innovation and Xenopus-Genetic. In both these datasets it becomes the worst performing algorithm.



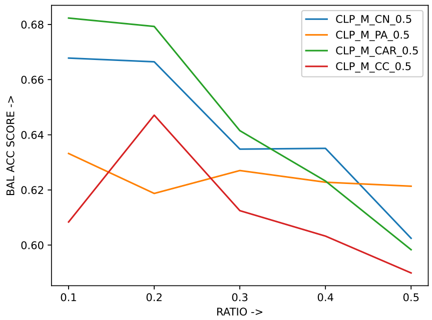
(A) Lazega-Law-Firm



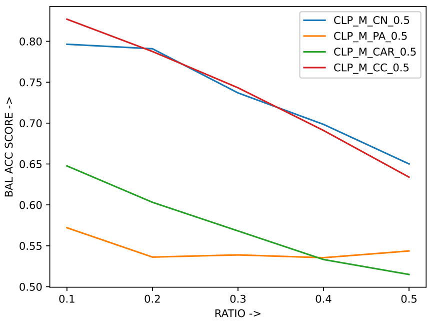
(B) CS-Aarhus



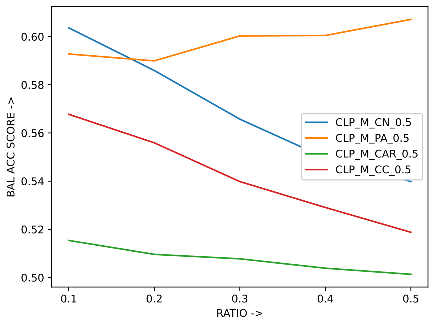
(C) Vickers-Chan-7thGraders



(D) Kapferer-Tailor-Shop



(E) CKM-Physicians-Innovation



(F) Xenopus-Genetic

FIGURE 6.5: Balanced Accuracy Score comparison of $CLP-MUL$ algorithm for different feature sets on datasets

6.3.2 $CLP-MUL$ comparison with link prediction methods on summarized weighted graph

In this section, the performance of the proposed algorithm with different baseline algorithms on the weighted graph is compared. From section 6.3.1 it can be concluded that the best variation of $CLP-MUL$ algorithm with the most consistent performance across datasets is CLP_M_CN, from hereon referred as $CLP-MUL$. Table 6.2 shows the comparison of the proposed $CLP-MUL$ algorithm with baseline methods with respect to AUC metric. $CLP-MUL$ is the best performing algorithm in all six datasets. In CS-Aarhus and CKM-Physicians-Innovation the improvement is quiet drastic. For Lazega-Law-Firm, Vickers-Chan-7thGraders and Kapferer-Tailor-Shop the performance improvement is significant. The least improvement is observed in Xenopus-Genetic dataset. A point to be noted here is that Xenopus-Genetic is the dataset in which the PA based feature set showed the most promise in Fig. 6.3, 6.4 and 6.5. Table 6.3 shows the comparison of the proposed $CLP-MUL$ algorithm with baseline methods with respect to F1 score. $CLP-MUL$ is the best performing algorithm in five datasets. The exception is Xenopus-Genetic where it is narrowly pushed to second place by JC-WT with CN-WT as a close third. Table 6.4 shows the comparison of the proposed $CLP-MUL$ algorithm with baseline methods with respect to Balanced Accuracy score. $CLP-MUL$ is the best performing algorithm in all six datasets. The least improvement is seen for Kapferer-Tailor-Shop and Xenopus-Genetic dataset while all others show significant improvement.

6.3.3 $CLP-MUL$ comparison with multiplex link prediction methods on individual layers

This section presents the results of the $CLP-MUL$ algorithm on application to specific multiplex network layers. In order to convert the probability matrix obtained after processing the summarized weighted graph, for each layer, the probability matrix is

TABLE 6.2: Comparison of the proposed algorithm $CLP-MUL$ with baseline algorithms in terms of AUC

DATASET	Ratio	CN-WT	PA-WT	JC-WT	AA-WT	RA-WT	LOCALP-WT	CC-WT	CLP-MUL
Lazega-Law-Firm	0.1	0.58119	0.54786	0.58759	0.58379	0.58564	0.58257	0.5833	0.77357
	0.2	0.58279	0.54804	0.58557	0.58387	0.58477	0.58164	0.56426	0.74831
	0.3	0.58062	0.54981	0.58363	0.58356	0.58426	0.57756	0.52795	0.74938
	0.4	0.58065	0.54609	0.57991	0.58235	0.58309	0.57272	0.55371	0.73296
	0.5	0.58063	0.54489	0.57832	0.58144	0.57856	0.56801	0.53739	0.70435
CS-Aarhus	0.1	0.64594	0.54625	0.64706	0.64938	0.65215	0.61883	0.47337	0.86357
	0.2	0.64888	0.54561	0.64867	0.64585	0.65074	0.61696	0.48034	0.86026
	0.3	0.64966	0.54537	0.64859	0.65129	0.64411	0.61727	0.48162	0.83551
	0.4	0.64151	0.54341	0.64551	0.64345	0.64404	0.61916	0.50509	0.81817
	0.5	0.62513	0.544	0.63206	0.63564	0.63721	0.61488	0.48388	0.7617
Vickers-Chan-7thGraders	0.1	0.59721	0.58669	0.59699	0.5979	0.60533	0.6007	0.60887	0.78945
	0.2	0.59586	0.5886	0.58853	0.59535	0.60059	0.59615	0.60457	0.73613
	0.3	0.59047	0.58178	0.58859	0.59319	0.59867	0.5905	0.57297	0.72974
	0.4	0.59428	0.57948	0.58143	0.59417	0.59438	0.58812	0.57714	0.70754
	0.5	0.59007	0.57489	0.58028	0.5948	0.59054	0.58235	0.55529	0.67723
Kapferer-Tailor-Shop	0.1	0.57943	0.55943	0.5763	0.5822	0.5818	0.58412	0.58365	0.72692
	0.2	0.57931	0.56609	0.56945	0.57731	0.57855	0.57708	0.55097	0.70821
	0.3	0.58036	0.56374	0.56634	0.58059	0.57898	0.57239	0.55464	0.68807
	0.4	0.58021	0.56183	0.56497	0.57897	0.58051	0.56613	0.53907	0.67649
	0.5	0.57421	0.56094	0.56364	0.57806	0.57816	0.56125	0.55584	0.64488
CKM-Physicians-Innovation	0.1	0.664	0.52145	0.66303	0.66485	0.66876	0.68705	0.43969	0.80837
	0.2	0.65029	0.52094	0.65666	0.65673	0.65276	0.67604	0.45768	0.79849
	0.3	0.63292	0.52048	0.63456	0.63465	0.63757	0.65673	0.48195	0.74078
	0.4	0.61458	0.51916	0.60769	0.61446	0.61231	0.64224	0.49062	0.70065
	0.5	0.58946	0.52064	0.58859	0.59162	0.58818	0.61815	0.50531	0.6505
Xenopus-Genetic	0.1	0.55662	0.4446	0.55587	0.54642	0.55731	0.57864	0.5511	0.60461
	0.2	0.54381	0.45663	0.54208	0.54406	0.54237	0.57165	0.52706	0.58655
	0.3	0.53362	0.46825	0.53855	0.53424	0.53369	0.55356	0.52827	0.56606
	0.4	0.52884	0.48459	0.52782	0.53125	0.52653	0.54864	0.52559	0.54989
	0.5	0.52023	0.49288	0.52097	0.51965	0.51968	0.53381	0.52662	0.53993

TABLE 6.3: Comparison of the proposed algorithm $CLP - MUL$ with baseline algorithms in terms of F1 Score

DATASET	Ratio	CN-WT	PA-WT	JC-WT	AA-WT	RA-WT	LOCALP-WT	CC-WT	CLP-MUL
Lazega-Law-Firm	0.1	0.14043	0.12675	0.13098	0.14115	0.14208	0.13721	0.13549	0.22086
	0.2	0.23764	0.21344	0.22198	0.23689	0.23763	0.23288	0.22247	0.33668
	0.3	0.30345	0.27433	0.29009	0.30383	0.30484	0.29954	0.25496	0.42053
	0.4	0.34998	0.31699	0.33592	0.35177	0.35285	0.34659	0.32492	0.46872
	0.5	0.37949	0.35079	0.37061	0.37955	0.37535	0.37679	0.34784	0.48126
CS-Aarhus	0.1	0.07636	0.05362	0.07152	0.07556	0.08004	0.07399	0.03638	0.13564
	0.2	0.14061	0.09601	0.13436	0.13422	0.14499	0.13851	0.06723	0.22498
	0.3	0.19306	0.13771	0.1942	0.19479	0.19547	0.1897	0.10731	0.30993
	0.4	0.24597	0.16923	0.24958	0.24502	0.24893	0.24872	0.13977	0.35976
	0.5	0.28022	0.19746	0.29053	0.29139	0.29505	0.28309	0.15252	0.37776
Vickers-Chan-7thGraders	0.1	0.24514	0.23654	0.22808	0.25031	0.26436	0.24203	0.25548	0.39588
	0.2	0.3757	0.35582	0.34592	0.36785	0.38542	0.36449	0.3791	0.50737
	0.3	0.43206	0.42135	0.41961	0.43882	0.44901	0.42953	0.43259	0.56212
	0.4	0.46758	0.45961	0.45583	0.47463	0.47674	0.47075	0.47236	0.56112
	0.5	0.48421	0.48202	0.47258	0.49161	0.47803	0.48952	0.4842	0.51848
Kapferer-Tailor-Shop	0.1	0.14107	0.13683	0.13306	0.1442	0.15047	0.14069	0.13601	0.21176
	0.2	0.23426	0.23131	0.21981	0.23509	0.24475	0.23412	0.21469	0.33528
	0.3	0.30629	0.29341	0.28531	0.30231	0.31218	0.29931	0.28854	0.37939
	0.4	0.34668	0.3388	0.33192	0.34564	0.35185	0.34371	0.31644	0.43017
	0.5	0.36805	0.37098	0.35902	0.37154	0.37638	0.36375	0.35105	0.42124
CKM-Physicians-Innovation	0.1	0.03974	0.00757	0.03949	0.03958	0.04048	0.03576	0.00508	0.04138
	0.2	0.07713	0.01459	0.07988	0.0794	0.07783	0.06706	0.0107	0.08911
	0.3	0.11108	0.02193	0.11271	0.11223	0.11466	0.09254	0.01715	0.1243
	0.4	0.14073	0.02844	0.13474	0.13964	0.13853	0.11934	0.0233	0.15845
	0.5	0.15339	0.03519	0.15266	0.15608	0.15184	0.1394	0.02976	0.1751
Xenopus-Genetic	0.1	0.0143	0.00106	0.0141	0.01179	0.0145	0.00967	0.00105	0.01278
	0.2	0.02496	0.00165	0.02358	0.02617	0.02366	0.0214	0.0018	0.02465
	0.3	0.03221	0.00268	0.03789	0.03349	0.03393	0.03006	0.00309	0.03338
	0.4	0.04132	0.00416	0.04182	0.0446	0.03958	0.04723	0.00498	0.03899
	0.5	0.0417	0.00396	0.04273	0.04024	0.04069	0.04875	0.00786	0.04739

TABLE 6.4: Comparison of the proposed algorithm $CLP-MUL$ with baseline algorithms in terms of Balanced Accuracy Score

DATASET	Ratio	CN-WT	PA-WT	JC-WT	AA-WT	RA-WT	LOCALP-WT	CC-WT	CLP-MUL
Lazega-Law-Firm	0.1	0.5615	0.53646	0.54503	0.56249	0.56419	0.55614	0.55657	0.70967
	0.2	0.56908	0.53971	0.54971	0.5682	0.56907	0.5635	0.54852	0.68534
	0.3	0.57191	0.5403	0.55604	0.57204	0.57311	0.56729	0.51512	0.68443
	0.4	0.57415	0.53998	0.55704	0.57521	0.5763	0.56956	0.53591	0.67394
	0.5	0.57288	0.54144	0.56032	0.57254	0.5706	0.56938	0.5241	0.65181
CS-Aarhus	0.1	0.6079	0.5407	0.59873	0.60742	0.61485	0.60467	0.45819	0.79948
	0.2	0.61242	0.53731	0.6061	0.60458	0.61784	0.61127	0.45942	0.78661
	0.3	0.61539	0.54551	0.61536	0.61618	0.61383	0.61167	0.4852	0.78591
	0.4	0.61834	0.54343	0.62198	0.61801	0.62008	0.62155	0.49376	0.76682
	0.5	0.61216	0.5445	0.61885	0.61974	0.62189	0.61359	0.47936	0.7225
Vickers-Chan-7thGraders	0.1	0.56084	0.55122	0.53842	0.56694	0.58246	0.55709	0.5727	0.7439
	0.2	0.58069	0.56213	0.54738	0.57646	0.59167	0.56928	0.57518	0.69784
	0.3	0.58228	0.57004	0.56163	0.58828	0.59585	0.57643	0.55499	0.67889
	0.4	0.5828	0.57044	0.55981	0.58911	0.59632	0.58088	0.55292	0.65378
	0.5	0.58206	0.57904	0.56112	0.59386	0.58615	0.58334	0.53539	0.61161
Kapferer-Tailor-Shop	0.1	0.54936	0.54222	0.53539	0.55506	0.56487	0.54926	0.54167	0.66782
	0.2	0.55652	0.55327	0.53778	0.55802	0.56854	0.5563	0.52739	0.66647
	0.3	0.5686	0.55616	0.54465	0.56497	0.57525	0.56105	0.53965	0.6348
	0.4	0.56843	0.55963	0.5498	0.56694	0.57457	0.56444	0.52335	0.63508
	0.5	0.56411	0.56305	0.54868	0.56669	0.57349	0.5605	0.52094	0.60248
CKM-Physicians-Innovation	0.1	0.65981	0.52769	0.65872	0.66035	0.66422	0.67368	0.45523	0.7964
	0.2	0.64732	0.52369	0.65402	0.65327	0.64963	0.66186	0.46528	0.79089
	0.3	0.63092	0.52545	0.63289	0.63248	0.63545	0.64863	0.48031	0.73693
	0.4	0.61342	0.52267	0.60692	0.61312	0.6112	0.63879	0.48703	0.69842
	0.5	0.58892	0.52267	0.58829	0.59094	0.58761	0.61631	0.49372	0.65008
Xenopus-Genetic	0.1	0.55648	0.51128	0.55599	0.54617	0.55701	0.57796	0.50966	0.60364
	0.2	0.54372	0.48181	0.54216	0.54389	0.54217	0.57126	0.49677	0.58594
	0.3	0.53357	0.49364	0.53862	0.53413	0.53358	0.55335	0.50601	0.56574
	0.4	0.52882	0.50837	0.52785	0.53118	0.52647	0.54857	0.51805	0.54968
	0.5	0.52022	0.48308	0.521	0.51961	0.51964	0.53378	0.52016	0.53986

multiplied by the relative density of the current layer with respect to the overall density of the summarized graph. This factor is $\frac{|E_M|-|E_j|}{|E_M|}$. Tables 6.5 and 6.6 show the comparison of the proposed $CLP-MUL$ algorithm with baseline methods for the AUC metric. $CLP-MUL$ is the best performing algorithm in datasets Lazega-Law-Firm, CS-Aarhus, Vickers-Chan-7thGraders and Kapferer-Tailer-Shop for ratios 0.1&0.2 in most cases. The exceptions are layer 1 of Lazega-Law-Firm, layer 3 of CS-Aarhus, and layer 1 of Kapferer-Tailer-Shop. Out of these exceptions, the performance of the proposed $CLP-MUL$ is third best only in layer 3 of the CS-Aarhus dataset, where it is outperformed by both $NSILR-MUL$ and $MADM-MUL$ because of extremely low average connectivity. Besides this exception, the proposed algorithm has the second-best performance in other cases, only marginally behind $MADM-MUL$. In the CKM-Physicians-Innovation dataset, the $CLP-MUL$ algorithm is the third best behind *NSIRLP* and *MADMLP*, which are designed specifically for link prediction in multiplex networks. This dataset is different from all others because of its high average shortest path length and extremely low average connectivity (which represents the cohesion in graph structure or relative difficulty of breaking the graph structure) in all layers. In the Xenopus-Genetic dataset, the $CLP-MUL$ algorithm performs best in layer 3. In layer 2, all link prediction algorithms have comparable results. In contrast, layer 1 is an exceptional case where the structure of the actual layer is drastically different from the summarized weighted graph because of a low number of nodes and edges. $PA-WT$ and $MADM-MUL$ algorithms show the best performance in such a case.

Tables 6.7 and 6.8 show the comparison of the proposed $CLP-MUL$ algorithm with baseline methods for F1 score. $CLP-MUL$ is the best performing algorithm in datasets Lazega-Law-Firm, CS-Aarhus, Vickers-Chan-7thGraders, and Kapferer-Tailer-Shop for ratios 0.1&0.2. For the CKM-Physicians dataset, the proposed algorithm shows the best results for layer 3 and is second best in layers 1 and 2. This dataset is characterized by high average shortest paths and low average connectivity. In the Xenopus-Genetic dataset, $CLP-MUL$ shows the best performance in layer 2 and is second best in layer 3. Layer 1 of this dataset presents an interesting edge case where the performance of all

TABLE 6.5: Comparison of the proposed algorithm $CLP-MUL$ with baseline algorithms in terms of AUC layer-wise

DATASET	LAYER	Ratio	CN-WT	PA-WT	JC-WT	AA-WT	RA-WT	CC-WT	LP-WT	NSILRLP	MADMLP	CLP-MUL
Lazega-Law-Firm	1	0.1	0.6022	0.5505	0.6054	0.6035	0.6040	0.5502	0.5892	0.7802	0.8451	0.8229
		0.2	0.6023	0.5487	0.6052	0.6038	0.6043	0.5439	0.5853	0.8073	0.8573	0.8115
		0.3	0.6037	0.5481	0.6054	0.6048	0.6046	0.5396	0.5836	0.8491	0.9060	0.7962
	2	0.1	0.6499	0.5671	0.6501	0.6486	0.6533	0.5918	0.6199	0.8253	0.8210	0.8743
		0.2	0.6512	0.5677	0.6509	0.6520	0.6521	0.5870	0.6204	0.8324	0.8204	0.8584
		0.3	0.6488	0.5671	0.6491	0.6496	0.6509	0.5750	0.6222	0.8733	0.8497	0.8373
	3	0.1	0.5912	0.5479	0.5897	0.5896	0.5918	0.4910	0.5833	0.7035	0.7283	0.7931
		0.2	0.5901	0.5466	0.5889	0.5918	0.5918	0.4976	0.5804	0.7409	0.7535	0.7837
		0.3	0.5913	0.5462	0.5885	0.5921	0.5927	0.5001	0.5791	0.8002	0.8099	0.7714
CS-Aarhus	1	0.1	0.7035	0.5309	0.7045	0.7054	0.7054	0.4339	0.6881	0.8976	0.8685	0.9474
		0.2	0.6990	0.5312	0.6998	0.6970	0.6988	0.4362	0.6882	0.9063	0.8927	0.9249
		0.3	0.6881	0.5292	0.6878	0.6884	0.6901	0.4509	0.6845	0.9407	0.9313	0.8930
	2	0.1	0.7056	0.7039	0.7079	0.7056	0.7069	0.7510	0.7124	0.8637	0.8161	0.9299
		0.2	0.7007	0.7017	0.7025	0.6989	0.7008	0.7347	0.7087	0.8932	0.8423	0.8940
		0.3	0.6879	0.7017	0.6860	0.6873	0.6880	0.7154	0.7009	0.9211	0.8700	0.8696
	3	0.1	0.5985	0.5924	0.6135	0.6078	0.6053	0.5227	0.6123	0.8938	0.8503	0.7326
		0.2	0.5828	0.5784	0.5773	0.5899	0.5864	0.5241	0.5820	0.9146	0.8914	0.6708
		0.3	0.5623	0.5695	0.5662	0.5680	0.5606	0.5348	0.5685	0.9056	0.8943	0.6226
Vickers-Chan-7thGraders	1	0.1	0.5941	0.5805	0.5898	0.5961	0.5987	0.5748	0.5931	0.7282	0.6960	0.8167
		0.2	0.5914	0.5785	0.5868	0.5920	0.5965	0.5658	0.5900	0.7679	0.7202	0.8003
		0.3	0.5880	0.5757	0.5828	0.5911	0.5931	0.5533	0.5849	0.8668	0.8820	0.7862
	2	0.1	0.6214	0.5638	0.6142	0.6266	0.6256	0.4927	0.5956	0.8200	0.7664	0.8539
		0.2	0.6194	0.5659	0.6157	0.6257	0.6276	0.4922	0.5942	0.8175	0.7862	0.8403
		0.3	0.6216	0.5657	0.6122	0.6246	0.6227	0.4852	0.5928	0.8531	0.8458	0.8166
	3	0.1	0.6200	0.5786	0.6186	0.6213	0.6250	0.5831	0.6029	0.7440	0.7525	0.8437
		0.2	0.6224	0.5755	0.6147	0.6224	0.6251	0.5572	0.6031	0.7620	0.7548	0.8415
		0.3	0.6199	0.5755	0.6129	0.6235	0.6229	0.5487	0.6020	0.8620	0.8486	0.8141

TABLE 6.6: Comparison of the proposed algorithm $CLP-MUL$ with baseline algorithms in terms of AUC layer-wise (contd..)

DATASET	LAYER	Ratio	CN-WT	PA-WT	JC-WT	AA-WT	RA-WT	CC-WT	LP-WT	NSILRLP	MADMLP	CLP-MUL
Kapferer-Tailor-Shop	1	0.1	0.6097	0.5619	0.5990	0.6085	0.6103	0.5084	0.5853	0.6784	0.8181	0.7667
		0.2	0.6039	0.5610	0.5951	0.6067	0.6066	0.5159	0.5853	0.7351	0.7811	0.7392
		0.3	0.5991	0.5622	0.5913	0.5992	0.5992	0.5169	0.5841	0.7942	0.8308	0.7298
	2	0.1	0.5942	0.5604	0.5912	0.5952	0.5937	0.5911	0.5766	0.7060	0.7200	0.7721
		0.2	0.5946	0.5604	0.5888	0.5940	0.5955	0.5802	0.5782	0.7160	0.7512	0.7592
		0.3	0.5925	0.5592	0.5885	0.5931	0.5963	0.5740	0.5762	0.7825	0.7713	0.7360
	3	0.1	0.5959	0.5832	0.5941	0.6037	0.6046	0.5794	0.5824	0.6877	0.6864	0.7086
		0.2	0.5899	0.5793	0.5854	0.5883	0.5909	0.5623	0.5858	0.7277	0.7172	0.6845
		0.3	0.5823	0.5753	0.5768	0.5809	0.5829	0.5635	0.5715	0.7576	0.7559	0.6552
CKM-Physicians-Innovation	1	0.1	0.6079	0.5566	0.6079	0.6107	0.6121	0.4787	0.6370	0.8083	0.8162	0.7054
		0.2	0.5925	0.5543	0.5912	0.5920	0.5910	0.4855	0.6250	0.8123	0.8290	0.6719
		0.3	0.5731	0.5508	0.5748	0.5754	0.5741	0.4991	0.6036	0.8317	0.8416	0.6395
	2	0.1	0.6266	0.5240	0.6242	0.6201	0.6224	0.4385	0.6469	0.7414	0.7746	0.7321
		0.2	0.6046	0.5248	0.6043	0.6035	0.6047	0.4562	0.6298	0.7452	0.8066	0.6902
		0.3	0.5840	0.5219	0.5838	0.5831	0.5838	0.4671	0.6066	0.7848	0.8231	0.6534
	3	0.1	0.6006	0.5139	0.5995	0.5992	0.5995	0.4937	0.6236	0.6213	0.6919	0.6858
		0.2	0.5857	0.5164	0.5854	0.5846	0.5838	0.4814	0.6045	0.6641	0.6973	0.6539
		0.3	0.5702	0.5159	0.5700	0.5687	0.5699	0.4908	0.5893	0.7054	0.7318	0.6286
Xenopus-Genetic	1	0.1	0.5000	0.6198	0.5000	0.5000	0.5000	0.5000	0.5437	0.4932	0.6244	0.4999
		0.2	0.5000	0.6056	0.5000	0.5000	0.5000	0.5000	0.5383	0.4940	0.5837	0.4999
		0.3	0.5000	0.5913	0.5000	0.5000	0.5000	0.5000	0.5170	0.4931	0.6066	0.5000
	2	0.1	0.5198	0.5495	0.5200	0.5186	0.5206	0.5405	0.5685	0.5437	0.5558	0.5368
		0.2	0.5165	0.5500	0.5149	0.5164	0.5139	0.5401	0.5599	0.5462	0.5787	0.5311
		0.3	0.5128	0.5474	0.5120	0.5121	0.5114	0.5351	0.5487	0.5526	0.5852	0.5222
	3	0.1	0.5615	0.5412	0.5610	0.5569	0.5588	0.5922	0.5646	0.5453	0.5959	0.6148
		0.2	0.5504	0.5469	0.5490	0.5508	0.5508	0.5741	0.5581	0.5578	0.5766	0.5982
		0.3	0.5416	0.5482	0.5411	0.5425	0.5446	0.5586	0.5494	0.5941	0.6001	0.5802

TABLE 6.7: Comparison of the proposed algorithm $CLP-MUL$ with baseline algorithms in terms of F1 Score layer-wise

DATASET	LAYER	Ratio	CN-WT	PA-WT	JC-WT	AA-WT	RA-WT	CC-WT	LP-WT	NSILRLP	MADMLP	CLP-MUL
Lazega-Law-Firm	1	0.1	0.1025	0.0845	0.0954	0.1015	0.1021	0.0820	0.0988	0.0712	0.0950	0.1569
		0.2	0.1765	0.1486	0.1691	0.1771	0.1801	0.1433	0.1746	0.1743	0.2259	0.2655
		0.3	0.2422	0.2002	0.2286	0.2412	0.2400	0.1915	0.2358	0.3592	0.4510	0.3427
	2	0.1	0.0652	0.0483	0.0641	0.0647	0.0686	0.0457	0.0651	0.0464	0.0493	0.0994
		0.2	0.1258	0.0900	0.1244	0.1252	0.1277	0.0849	0.1241	0.1090	0.1121	0.1811
		0.3	0.1806	0.1267	0.1813	0.1801	0.1820	0.1160	0.1799	0.2215	0.2333	0.2472
	3	0.1	0.0991	0.0854	0.0923	0.1001	0.1024	0.0691	0.0972	0.0637	0.0903	0.1592
		0.2	0.1762	0.1510	0.1654	0.1760	0.1797	0.1275	0.1715	0.1612	0.2061	0.2688
		0.3	0.2321	0.2019	0.2202	0.2330	0.2377	0.1746	0.2310	0.3378	0.4201	0.3420
CS-Aarhus	1	0.1	0.0988	0.0264	0.0983	0.0984	0.0983	0.0175	0.0937	0.0422	0.0519	0.1125
		0.2	0.1849	0.0500	0.1858	0.1827	0.1845	0.0345	0.1749	0.0976	0.1280	0.2195
		0.3	0.2569	0.0695	0.2553	0.2559	0.2572	0.0520	0.2450	0.2027	0.2758	0.3180
	2	0.1	0.0849	0.0639	0.0860	0.0850	0.0855	0.0282	0.0815	0.0288	0.0304	0.0937
		0.2	0.1564	0.1135	0.1573	0.1554	0.1557	0.0512	0.1463	0.0737	0.0830	0.1764
		0.3	0.2207	0.1591	0.2187	0.2198	0.2208	0.0729	0.2067	0.1386	0.1730	0.2615
	3	0.1	0.1274	0.0117	0.1452	0.1428	0.1378	0.0040	0.1211	0.0070	0.0125	0.1807
		0.2	0.1612	0.0195	0.1545	0.1762	0.1675	0.0069	0.1362	0.0107	0.0205	0.2147
		0.3	0.1542	0.0282	0.1646	0.1686	0.1506	0.0116	0.1523	0.0255	0.0481	0.2231
Vickers-Chan-7thGraders	1	0.1	0.2315	0.2205	0.2137	0.2377	0.2484	0.2267	0.2273	0.1586	0.1424	0.3699
		0.2	0.3533	0.3338	0.3309	0.3562	0.3687	0.3391	0.3462	0.3492	0.3278	0.5229
		0.3	0.4135	0.3995	0.4000	0.4234	0.4310	0.3958	0.4119	0.6718	0.6998	0.5903
	2	0.1	0.1072	0.0935	0.1039	0.1207	0.1292	0.0754	0.1031	0.0753	0.0785	0.2026
		0.2	0.1846	0.1657	0.1872	0.1927	0.2265	0.1360	0.1853	0.1871	0.1998	0.3356
		0.3	0.2518	0.2186	0.2495	0.2509	0.2782	0.1725	0.2519	0.3877	0.4290	0.3951
	3	0.1	0.1603	0.1320	0.1408	0.1552	0.1533	0.1274	0.1461	0.0919	0.0926	0.2468
		0.2	0.2431	0.2152	0.2262	0.2489	0.2525	0.1970	0.2337	0.2114	0.2080	0.3834
		0.3	0.2935	0.2774	0.2914	0.2991	0.3143	0.2511	0.2951	0.4704	0.4645	0.4533

TABLE 6.8: Comparison of the proposed algorithm $CLP-MUL$ with baseline algorithms in terms of F1 Score layer-wise (contd..)

DATASET	LAYER	Ratio	CN-WT	PA-WT	JC-WT	AA-WT	RA-WT	CC-WT	LP-WT	NSILRLP	MADMLP	CLP-MUL
Kapferer-Tailor-Shop	1	0.1	0.0764	0.0707	0.0706	0.0749	0.0797	0.0490	0.0720	0.0453	0.0578	0.1078
		0.2	0.1315	0.1269	0.1301	0.1340	0.1407	0.0923	0.1308	0.1132	0.1297	0.1802
		0.3	0.1830	0.1712	0.1802	0.1812	0.1868	0.1261	0.1807	0.2586	0.2772	0.2408
	2	0.1	0.1032	0.0981	0.0966	0.1101	0.1107	0.0960	0.1027	0.0753	0.0899	0.1585
		0.2	0.1806	0.1681	0.1684	0.1835	0.1914	0.1607	0.1777	0.1645	0.2086	0.2619
		0.3	0.2395	0.2200	0.2244	0.2345	0.2467	0.2111	0.2342	0.3402	0.3795	0.3272
	3	0.1	0.0474	0.0345	0.0487	0.0493	0.0500	0.0291	0.0455	0.0299	0.0344	0.0545
		0.2	0.0910	0.0658	0.0915	0.0906	0.0927	0.0516	0.0870	0.0640	0.0721	0.1033
		0.3	0.1245	0.0887	0.1222	0.1235	0.1255	0.0724	0.1074	0.1314	0.1558	0.1424
CKM-Physicians-Innovation	1	0.1	0.0295	0.0042	0.0295	0.0300	0.0304	0.0025	0.0219	0.0138	0.0332	0.0339
		0.2	0.0557	0.0083	0.0552	0.0553	0.0548	0.0048	0.0446	0.0328	0.0755	0.0638
		0.3	0.0740	0.0122	0.0753	0.0757	0.0745	0.0076	0.0637	0.0691	0.1299	0.0901
	2	0.1	0.0375	0.0038	0.0370	0.0358	0.0363	0.0027	0.0269	0.0166	0.0337	0.0413
		0.2	0.0681	0.0078	0.0680	0.0674	0.0682	0.0055	0.0528	0.0400	0.0839	0.0763
		0.3	0.0904	0.0114	0.0901	0.0894	0.0900	0.0077	0.0731	0.0879	0.1470	0.1057
	3	0.1	0.0392	0.0032	0.0387	0.0386	0.0388	0.0029	0.0307	0.0111	0.0290	0.0435
		0.2	0.0703	0.0064	0.0700	0.0692	0.0689	0.0052	0.0567	0.0340	0.0649	0.0786
		0.3	0.0916	0.0094	0.0910	0.0899	0.0911	0.0081	0.0802	0.0768	0.1266	0.1092
Xenopus-Genetic	1	0.1	0.0000	0.0167	0.0000	0.0000	0.0000	0.0000	0.0250	0.0000	0.0000	0.0000
		0.2	0.0000	0.0241	0.0000	0.0000	0.0000	0.0001	0.0363	0.0000	0.0000	0.0000
		0.3	0.0000	0.0420	0.0000	0.0000	0.0000	0.0001	0.0289	0.0000	0.0001	0.0000
	2	0.1	0.0085	0.0007	0.0086	0.0080	0.0088	0.0013	0.0192	0.0018	0.0002	0.0085
		0.2	0.0148	0.0016	0.0135	0.0147	0.0125	0.0028	0.0362	0.0045	0.0006	0.0158
		0.3	0.0183	0.0026	0.0173	0.0175	0.0166	0.0046	0.0505	0.0101	0.0012	0.0197
	3	0.1	0.0150	0.0008	0.0149	0.0139	0.0144	0.0009	0.0091	0.0068	0.0176	0.0148
		0.2	0.0284	0.0017	0.0278	0.0286	0.0285	0.0018	0.0205	0.0166	0.0301	0.0307
		0.3	0.0403	0.0028	0.0395	0.0413	0.0421	0.0030	0.0317	0.0424	0.0594	0.0453

TABLE 6.9: Comparison of the proposed algorithm $CLP-MUL$ with baseline algorithms in terms of Balanced Accuracy Score layer-wise

DATASET	LAYER	Ratio	CN-WT	PA-WT	JC-WT	AA-WT	RA-WT	CC-WT	LP-WT	NSILRLP	MADMLP	CLP-MUL
Lazega-Law-Firm	1	0.1	0.5834	0.5416	0.5701	0.5814	0.5825	0.5375	0.5766	0.7064	0.7875	0.7394
		0.2	0.5818	0.5412	0.5729	0.5823	0.5857	0.5312	0.5793	0.7215	0.7912	0.7315
		0.3	0.5901	0.5428	0.5771	0.5894	0.5881	0.5277	0.5842	0.7545	0.8232	0.7221
	2	0.1	0.6140	0.5620	0.6114	0.6123	0.6207	0.5632	0.6140	0.7561	0.7325	0.7938
		0.2	0.6207	0.5621	0.6191	0.6202	0.6224	0.5592	0.6183	0.7581	0.7224	0.7864
		0.3	0.6246	0.5632	0.6257	0.6242	0.6259	0.5497	0.6241	0.7726	0.7375	0.7730
	3	0.1	0.5735	0.5404	0.5590	0.5751	0.5789	0.4893	0.5698	0.6518	0.6591	0.7244
		0.2	0.5780	0.5416	0.5641	0.5779	0.5819	0.4962	0.5722	0.6808	0.6791	0.7155
		0.3	0.5780	0.5425	0.5650	0.5788	0.5833	0.4990	0.5770	0.7204	0.7319	0.7055
CS-Aarhus	1	0.1	0.6871	0.5268	0.6873	0.6874	0.6871	0.4447	0.6835	0.8180	0.8256	0.8818
		0.2	0.6872	0.5292	0.6875	0.6839	0.6855	0.4517	0.6842	0.8201	0.8454	0.8773
		0.3	0.6805	0.5238	0.6795	0.6794	0.6806	0.4619	0.6805	0.8431	0.8765	0.8604
	2	0.1	0.6957	0.6915	0.6980	0.6949	0.6966	0.6795	0.6994	0.8076	0.7033	0.8926
		0.2	0.6931	0.6903	0.6946	0.6911	0.6928	0.6696	0.6948	0.8300	0.7306	0.8688
		0.3	0.6829	0.6900	0.6810	0.6818	0.6826	0.6501	0.6879	0.8307	0.7471	0.8519
	3	0.1	0.5984	0.5896	0.6135	0.6077	0.6051	0.5317	0.6122	0.8091	0.8379	0.7320
		0.2	0.5828	0.5771	0.5773	0.5898	0.5863	0.5273	0.5819	0.8125	0.8768	0.6705
		0.3	0.5623	0.5687	0.5662	0.5680	0.5605	0.5354	0.5684	0.8193	0.8825	0.6224
Vickers-Chan-7thGraders	1	0.1	0.5625	0.5492	0.5387	0.5698	0.5817	0.5552	0.5572	0.6927	0.6492	0.7415
		0.2	0.5747	0.5566	0.5484	0.5781	0.5900	0.5531	0.5669	0.7094	0.6731	0.7319
		0.3	0.5749	0.5626	0.5551	0.5844	0.5932	0.5396	0.5718	0.7856	0.7985	0.7168
	2	0.1	0.5753	0.5442	0.5694	0.5980	0.6080	0.4915	0.5680	0.7202	0.7310	0.7784
		0.2	0.5753	0.5492	0.5788	0.5850	0.6158	0.4959	0.5763	0.7320	0.7531	0.7716
		0.3	0.5885	0.5534	0.5860	0.5873	0.6087	0.4837	0.5884	0.7668	0.8037	0.7496
	3	0.1	0.6053	0.5663	0.5822	0.6002	0.5982	0.5659	0.5882	0.6793	0.6550	0.7659
		0.2	0.5976	0.5675	0.5803	0.6020	0.6051	0.5425	0.5885	0.6903	0.6588	0.7625
		0.3	0.5865	0.5713	0.5845	0.5916	0.6039	0.5348	0.5880	0.7803	0.7427	0.7424

TABLE 6.10: Comparison of the proposed algorithm $CLP-MUL$ with baseline algorithms in terms of Balanced Accuracy Score layer-wise

DATASET	LAYER	Ratio	CN-WT	PA-WT	JC-WT	AA-WT	RA-WT	CC-WT	LP-WT	NSILRLP	MADMLP	CLP-MUL
Kapferer-Tailor-Shop	1	0.1	0.5892	0.5749	0.5756	0.5855	0.5925	0.5026	0.5787	0.6440	0.7315	0.7045
		0.2	0.5796	0.5742	0.5779	0.5834	0.5897	0.5064	0.5789	0.6684	0.7142	0.6842
		0.3	0.5819	0.5716	0.5793	0.5799	0.5844	0.5028	0.5798	0.7222	0.7494	0.6796
	2	0.1	0.5692	0.5592	0.5581	0.5814	0.5815	0.5655	0.5678	0.6827	0.6693	0.7077
		0.2	0.5766	0.5612	0.5622	0.5794	0.5877	0.5528	0.5730	0.6700	0.6833	0.6982
		0.3	0.5800	0.5607	0.5654	0.5754	0.5864	0.5472	0.5753	0.7104	0.6941	0.6846
	3	0.1	0.5860	0.5704	0.5905	0.5931	0.5948	0.5598	0.5833	0.6771	0.6611	0.6887
		0.2	0.5829	0.5709	0.5829	0.5815	0.5848	0.5383	0.5826	0.6912	0.6832	0.6692
		0.3	0.5776	0.5670	0.5758	0.5762	0.5789	0.5378	0.5665	0.7132	0.7210	0.6470
CKM-Physicians-Innovation	1	0.1	0.6072	0.5456	0.6077	0.6095	0.6110	0.4670	0.6349	0.8039	0.8102	0.7024
		0.2	0.5920	0.5450	0.5911	0.5914	0.5903	0.4660	0.6234	0.8046	0.8224	0.6700
		0.3	0.5728	0.5422	0.5748	0.5750	0.5736	0.4751	0.6027	0.8199	0.8323	0.6384
	2	0.1	0.6258	0.5151	0.6237	0.6192	0.6214	0.4605	0.6445	0.7400	0.7695	0.7294
		0.2	0.6041	0.5198	0.6041	0.6028	0.6041	0.4642	0.6283	0.7422	0.8010	0.6886
		0.3	0.5837	0.5167	0.5837	0.5827	0.5834	0.4592	0.6055	0.7789	0.8152	0.6523
	3	0.1	0.6002	0.5091	0.5990	0.5988	0.5992	0.4950	0.6224	0.6201	0.6899	0.6849
		0.2	0.5855	0.5133	0.5851	0.5844	0.5836	0.4832	0.6038	0.6618	0.6946	0.6534
		0.3	0.5701	0.5113	0.5698	0.5686	0.5697	0.4914	0.5889	0.7020	0.7286	0.6283
Xenopus-Genetic	1	0.1	0.5000	0.6198	0.5000	0.5000	0.5000	0.5000	0.5437	0.4932	0.5011	0.4999
		0.2	0.5000	0.6056	0.5000	0.5000	0.5000	0.5000	0.5383	0.4940	0.5011	0.4999
		0.3	0.5000	0.5913	0.5000	0.5000	0.5000	0.5000	0.5170	0.4931	0.5011	0.5000
	2	0.1	0.5198	0.5421	0.5200	0.5186	0.5206	0.5409	0.5685	0.5433	0.5011	0.5368
		0.2	0.5165	0.5444	0.5150	0.5164	0.5139	0.5404	0.5600	0.5458	0.5011	0.5311
		0.3	0.5128	0.5434	0.5120	0.5121	0.5114	0.5354	0.5487	0.5522	0.5011	0.5222
	3	0.1	0.5614	0.5306	0.5611	0.5567	0.5586	0.5412	0.5642	0.5452	0.5950	0.6141
		0.2	0.5503	0.5389	0.5491	0.5506	0.5507	0.5375	0.5579	0.5577	0.5757	0.5977
		0.3	0.5416	0.5425	0.5411	0.5424	0.5445	0.5351	0.5493	0.5937	0.5986	0.5799

algorithms falls drastically. This is because of an extremely small number of edges in this layer which makes it significantly different from the summarized weighted graph. Tables 6.9 and 6.10 show the comparison of the proposed $CLP-MUL$ algorithm with baseline methods for the Balanced Accuracy score. $CLP-MUL$ is the best performing algorithm in datasets Lazega-Law-Firm, Vickers-Chan-7thGraders and Kapferer-Tailer-Shop for ratios 0.1&0.2. In the CKM-Physicians-Innovation dataset, the $CLP-MUL$ algorithm is the third best behind *NSIRLP* and *MADMLP*, which are designed specifically for link prediction in multiplex networks. This dataset differs from all others because of its high average shortest path length in all layers. For $Ratio = 0.3$ $CLP-MUL$ algorithm shows worse performance than others because the performance of this algorithm is directly correlated with the task of community detection, which becomes cumbersome if a complete overall view of the graph and its relationships is not available. These tables only show networks in which the number of layers is less than five due to space constraints. This is because it can be assumed that the probabilities for links on the weighted graph are the same as links on layers for the same pair of nodes, as in rigid core communities.

6.4 Concluding Remarks

This chapter presents a novel method for link prediction in multiplex networks based on community detection, $CLP-MUL$. The proposed algorithm predicts links that are not specific to a particular layer but are based on communities detected using the summarized information of all layers. In this approach, a clustering method that uses information diffusion for label propagation to fit our needs on weighted networks is formulated. This method determines the region of influence of different central nodes. These regions are the communities/clusters that have high rigidity across layers. This approach considers these communities to stretch across layers even if the edge structure of a particular layer may not agree entirely with it. The detected clusters are used for calculating intra-cluster and inter-cluster similarity between node pairs for link prediction. The experiments are performed on six real-world datasets. The results indicate that the argument was justified for datasets with low average shortest path length and relatively higher edge density. $CLP-MUL$ method is compared with the classical link prediction methods for weighted graphs, demonstrating its superior performance both on the summarized weighted graph and on the original layers. The algorithm performance shows a slight deterioration for datasets with a high average shortest path length compared with the best algorithms in those cases.