Pattern Recognition 000 (2017) 1–15

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Discriminative metric learning for multi-view graph partitioning

Juan-Hui Li, Chang-Dong Wang*, Pei-Zhen Li, Jian-Huang Lai

School of Data and Computer Science, Sun Yat-sen University, Guangzhou, PR China

ARTICLE INFO

Article history:
Received 15 November 2016
Revised 27 May 2017
Accepted 7 June 2017
Available online xxx

Keywords:
Discriminative metric learning
Multi-view
Graph partitioning
Permanence
Modularity

ABSTRACT

In recent years, multi-view graph partitioning has attracted more and more attention, but most efforts have been made to develop graph partitioning approaches directly in the original topological structure. In many real-world applications, graph may contain noisy links and the distance metric may not be so discriminative for revealing cluster structure. This paper addresses the problem of discriminative metric learning for multi-view graph partitioning. In particular, we propose a novel method called *Multi-view DML* (abbr. of Multi-view Discriminative Metric Learning) to transform the metric space in the original graph into a more discriminative metric space, in which better graph partitioning results will be obtained. We envision the multi-view graph as an adaptive dynamic system, where both the intra-view connections and the inter-view couplings are interplayed to gradually update the relation metric among nodes. On the one hand, the inter-view coupling will be influenced by the intra-view connection between two nodes. On the other hand, the inter-view coupling will also affect the intra-view connection. Such interplay eventually makes the whole graph reach a steady state which has a stronger cluster structure than the original graph. Extensive experiments are conducted on both synthetic and real-world graphs to confirm that the proposed method is able to learn more discriminative metric.

1. Introduction

Recently, a huge amount of graphs have been generated such as social networks, citation networks and collaboration networks, where the nodes represent the objects in the world and the edges (connections) represent the relationship (interaction) among the objects [1–4]. The weight of an edge represents the interaction strength between the connected nodes [5]. These graphs may exhibit some cluster structure where nodes in the same cluster are more similar than nodes belonging to different clusters. Discovering the cluster structure not only reveals the graph properties but also has a number of real world applications, such as information influence analysis in social network, market segmentation in product market analysis, customer recommendation in online shopping platform, etc [4]. To achieve this goal, many graph partitioning[1] algorithms have been developed, such as *permanence* [6], *weighted conductance* [7], *modularity* [8,9], *PageRank centrality* [10], *cut-ratio* [11], *normalized cut* [12], and *distance dynamics* [13].

As the rapid development of the web technology, in particular in the social network, there emerge some multi-view graphs [14]. For instance, in social network, it is often the case that each user has various accounts in different social platforms which are coupled by some common bonds, e.g. Email address. Each platform is taken as a view in the social network. It is possible for a user to behave slightly differently in various views and the social behaviours in different views may influence each other, which means that the inter-view coupling should be taken into account for discovering the relationship among nodes in the same view [15]. In two different views, when a user in one view is more similar to itself in another view (i.e., more similar linkage structure and interaction strength), the strength of the inter-view coupling between this pair of views is stronger [16]. Due to the complex interplay between the intra-view connection and the inter-view coupling, multi-view graph partitioning encounters more challenges than that in single-view graph. Although some methods have been developed for multi-view graph partitioning [14,17,18], most of them are only based on the original topological structure without learning more discriminative metric space.

In real-world applications, the cluster structure in the original graph would not be quite clear and would further weaken due to the noisy links [19,20]. For example, in the same university, two users u and v from different departments may know each other.

* Corresponding author at: School of Mobile Information Engineering, School of Mobile Information Engineering, 5th floor of Administrative Bu, Zhuhai, Guangdong 519082, China

E-mail addresses: sysuLiJuanHui@163.com (J.-H. Li), changdongwang@hotmail.com (C.-D. Wang), sysuLiPeizhen@163.com (P.-Z. Li), stsljh@mail.sysu.edu.cn (J.-H. Lai).

[1] In the literature of data mining research, graph partitioning is also called community detection in network where "community" has the same meaning of "cluster", i.e. a set of similar nodes. In this paper, we will use network and graph, community and cluster alternatively.

http://dx.doi.org/10.1016/j.patcog.2017.06.012
0031-3203/

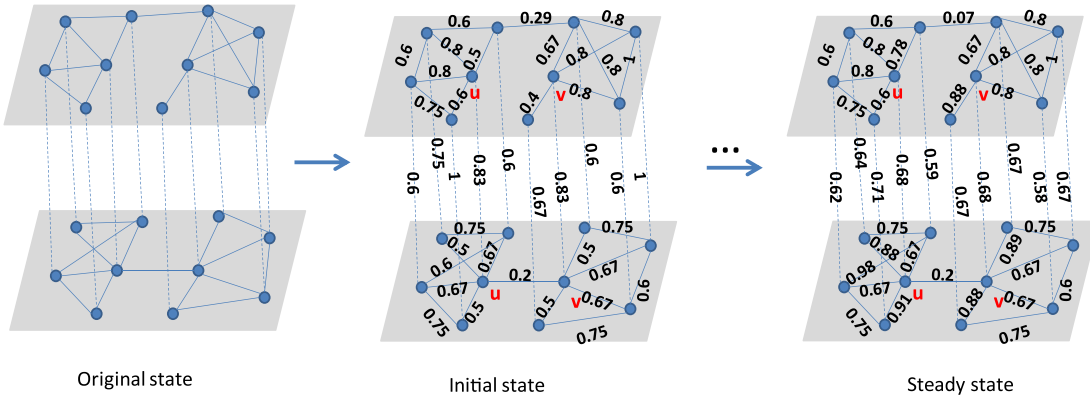


Fig. 1. Illustration of the three states of a two-view graph.

In unweighted graph, they are connected by an edge. However, it is possible that the graph partitioning algorithm assigns these two nodes into the same cluster, i.e. the same department, which is obviously wrong according to the ground-truth. Although recently, some efforts have been made to metric learning in graph so as to enhance the discriminability of cluster structure [13,19,21], they are limited to the single-view graph and there is still a lack of such work on the multi-view graph. The main challenge lies in the mutual influences between the intra-view connection and the inter-view coupling.

In this paper, we propose a new method called *Multi-view DML* (abbr. of Multi-view Discriminative Metric Learning) to address the discriminative metric learning for multi-view graph partitioning. The basic idea is to envision the given multi-view graph as a dynamic system, where three types of states are involved. The first one is the *original state*, which represents the original graph, i.e. the input of our algorithm. Jaccard similarity is utilized to transform the original state into the *initial state*. From the initial state, the intra-view connection strength and the inter-view coupling strength will be gradually updated by the proposed strategies, leading to the *steady state* which is taken as output. Therefore, the main idea of our approach has two steps: 1) Initialization. The intra-view connection strength and the inter-view coupling strength are initialized by the proposed transformation based on Jaccard similarity. 2) Update procedure. Different strategies are used to update the two types of strengths of the multi-view graph in an interplay manner. For the intra-view connection strength, three types of inter-view coupling adjusted influence from *directly linked nodes, common neighbors* and *exclusive neighbors* are all taken into account. On the other hand, the update of the inter-view coupling strength of one node between two views takes into account the interplay of both the intra-view connections associated with the node and the inter-view couplings related with its neighbors. The update procedure continues until convergence, in which the more discriminative metric (i.e. the intra-view connections and the inter-view couplings) will be obtained.

For illustration, Fig. 1 shows the three states of a two-view graph. (a) is the original multi-view graph describing the linking information of all the nodes in two different views and is used as the input of our algorithm. (b) presents the initial social relationship among nodes after transformation based on Jaccard similarity. (c) is the steady state of this two-view graph, which is taken as the output and provides more discriminative metric for graph partitioning.

The proposed method addresses a specific problem in multi-view clustering, which belongs to a more general field of multi-view learning. Compared with the existing metric learning methods for single-view graphs, apart from inheriting the advantages of the method in [13], the proposed Multi-view DML method for the first time provides a strategy for studying the interplay between the intra-view connection and the inter-view coupling, which would be helpful in the research of many multi-view graph analysis tasks.

The rest of the paper will be organized as follows. In Section 2, we briefly review the related work. Section 3 describes our model in detail, and in Section 4, extensive experiments are conducted to demonstrate the effectiveness of our algorithm. We finally conclude our paper in Section 5.

2. Related work

Recently, some methods have been developed for multi-view graph partitioning [14,17,18,22–25]. One of the earliest methods is [14], which presents a framework of node cluster quality evaluators that allows studying the cluster structure of multi-view graphs. In a more general manner, Zhang et al. [17] combine the information from various views and extend modularity into the multi-view weighted signed version, where the extended modularity is used for measuring the quality of dividing a graph into clusters. Higher value of the extended modularity indicates better division and higher separability of graph structure. In [18], Li et al. propose a model for multi-view weighted graph partitioning based on permanence [6]. In [22], multi-view modularity optimization has been applied to analyze the Los Angeles Police Department field interview card data set. It combines the information of two views, i.e. geographic and social information about stops to find clusters. Hu et al. [23] use multi-view modularity combining graphs and image data for graph partitioning. In [24], the graph is time-dependent and the temporal clusters to be discovered consist of nodes across multiple snapshots. A measure of partition distance called *estrangement* is proposed to find meaningful temporal clusters at various degrees of temporal smoothness in real-world graphs. Although the above methods work well in some relatively simple multi-view graphs, when handling complex multi-view graphs with less obvious cluster structure, their performance may degenerate seriously. One remedy is to enhance the discriminability before performing graph partitioning.

In the single-view graph case, some efforts have been made in enhancing the discriminability of cluster structure [13,19,21,26–31]. For instance, Yang et al. [28] propose a boosting framework for preserving both visual and semantic similarities for image retrieval systems. Yeung and Chang [29] propose an extension of RCA using both positive and negative equivalent constraints. In [30], the distance metric is learned for addressing the text line segmentation problem. In [31], an adaptive Semi-supervised Clustering Kernel Method based on Metric learning (SCKMM) is proposed to solve

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012

Table 1
Notations used in this paper.

Notation	Description
$\mathcal{V}$	The set of nodes
$\mathcal{E}$	The set of weighted intra-view connections
$\mathcal{C}$	The set of weighted inter-view couplings
$\mathcal{W}$	The weights of the intra-view connections
S	The number of the views
N	The number of nodes
w_{uvs}	The weight for the intra-view connection between node u and node v in view s
C_{usr}	The weight for the inter-view coupling of node u between view s and view r
A	The intra-view connection matrix $A = [A_{uvs}] \in \mathbb{R}^{N\times N\times S}$ with $A_{uvs} = w_{uvs}$
c_{us}	The sum of the inter-view coupling strengths of node u in view s
k_{us}	The total interaction strength of node u in view s
$\Gamma_s(u)$	The neighbors of node u in view s
$J_s(u, v)$	The initial value of the intra-view connection between node u and node v in view s
$J_{sr}(u)$	The initial value of the inter-view coupling of node u between view s and r
$\mathcal{CN}_s(u, v)$	The common neighbors for node u and node v in view s
$\mathcal{CN}_{sr}(u)$	The common neighbors of node u across view s and view r
$\mathcal{EN}_s(u\backslash v)$	The exclusive neighbors of node u and node v w.r.t. node v in view s
$\rho_s(x, u)$	The degree of the influence on the interaction between node u and node v from node x
λ	The cohesion parameter which is used to obtain $\rho_s(x, u)$
$DC_{sr}(u)$	The similarity of node u's behavior in view s and view r
DI_{uvs}	The inter-view coupling adjusted influence from directly linked nodes on the connection between node u and node v in view s
CI_{uvs}	The inter-view coupling adjusted influence from common neighbors on the connection between node u and node v in view s
EI_{uvs}	The inter-view coupling adjusted influence from exclusive neighbors on the connection between node u and node v in view s
$J_s(u, v, t)$	The intra-view connection between node u and node v in view s in the tth iteration
$J_{sr}(u, t)$	The inter-view coupling of node u between view s and r in the tth iteration
ϵ	A parameter used to stop the updating procedure

some problems in semi-supervised clustering, e.g. violation problem of pairwise constraints. More recently, Tao et al. [26] propose the Manifold Ranking-Based Matrix Factorization (MRMF) model focusing on the feature information to solve the saliency detection problem. Another work Dual-Regularized KISS (DR-KISS) [27] improves the accuracy of person re-identification problem by developing a more discriminative distance metric learning. In [32], the distance metric is learned based on the commute time. Zhao et al. [19] explore the possibility of using a geometric space embedding to learn the random-walk distances, including: 1) hitting time; 2) commute time; 3) personalized PageRank. Another research work that takes into account the metric learning is the distance dynamics [13], which regards the given graph as a dynamic system where the interaction strength of the nodes changes gradually, leading to more discriminative metric. Very recently, motif is utilized to enhance the cluster structure [21], which is one of the most common higher-order structures, i.e., the small subgraphs describing different features in graph. However, all the above methods are limited to the single-view graph, and due to the complex topological structure consisting of the intra-view connection and the inter-view coupling, it remains a challenging issue to learn discriminative metric in multi-view graph.

To address the above issues, inspired by [13], we propose a discriminative metric learning method for multi-view graph called *Multi-view DML*, which can find a more discriminative metric for both the intra-view connection and the inter-view coupling.

3. The proposed model

In this section, we will introduce the proposed model starting with some preliminary definitions followed by the two main steps (i.e. initialization and update procedure). The notations used in this paper are summarized in Table 1.

3.1. Preliminaries

Given a multi-view graph $G(\mathcal{V}, \mathcal{E}, \mathcal{C}, \mathcal{W})$ consisting of S views with each view containing N nodes, where $\mathcal{V}$ is the set of nodes, $\mathcal{E}$ denotes the set of weighted intra-view connections with the weights stored in $\mathcal{W}$, and $\mathcal{C}$ is the set of weighted inter-view couplings. We use w_{uvs} $(0 \le w_{uvs} \le 1)$ to represent the weight for the intra-view connection between node u and node v in view s. In the unweighted case, $w_{uvs} = 1$ if node u and node v are interconnected in view s, and $w_{uvs} = 0$ otherwise. Similarly, we use C_{usr} $(0 \le C_{usr} \le 1)$ to denote the weight for the inter-view coupling of node u between view s and view r. In the unweighted case, $C_{usr} = 1$. The multi-view graph can be represented by an intra-view connection matrix $A = [A_{uvs}] \in \mathbb{R}^{N\times N\times S}$ with $A_{uvs} = w_{uvs}$ and an inter-view coupling matrix $C = [C_{usr}] \in \mathbb{R}^{N\times S\times S}$. Notation $c_{us} = \sum_r C_{usr}$ is used to denote the sum of the inter-view coupling strengths of node u in view s, and $k_{us} = \sum_v A_{uvs}$ denotes the total interaction strength of node u in view s which is also the degree in the unweighted case.

To better describe the following sections, we make a definition following [13]:

Definition 1 (Neighbors of node u in view s). For a given multi-view graph $G(\mathcal{V}, \mathcal{E}, \mathcal{C}, \mathcal{W})$, each node may have various neighbors in different views. For node u in view s, the neighborhood contains itself and its adjacent nodes

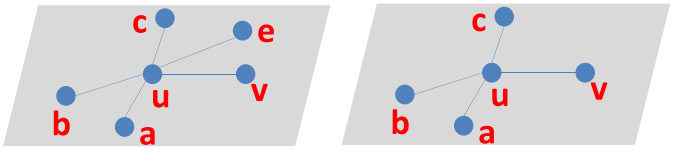
$$\Gamma_s(u) = \{v \in \mathcal{V} | (u, v) \in \mathcal{E}\} \cup \{u\}. \quad (1)$$

Notice that in the above definition, since $\mathcal{E}$ contains only intra-view connections, for node u in view s, $(u, v) \in \mathcal{E}$ implies that only intra-view connections within view s are considered in finding neighbor nodes. Therefore, as expected, Eq. (1) returns only neighbors of node u in view s.

3.2. Initialization

Before performing the update of the relation metric, some initializations are required to transform the original graph into the initial metric space based on the Jaccard similarity [33]. For the sake of simplicity and without loss of generality, we consider the original value of the inter-view coupling as 1. For the given multi-view graph, the intra-view connection can be either weighted or unweighted, where different initialization strategies are used.

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012



(a) Neighborhood of node u in View s (b) Neighborhood of node u in View r

Fig. 2. Example to show how to obtain the initial value for the inter-view coupling from the original graph with unweighted intra-view connection.

3.2.1. Unweighted graph

Following [13], we define the intra-view connection strength as follows.

Definition 2 (Initialization of the intra-view connection strength in unweighted graph)**.** In the initial state, the intra-view connection strength is assigned using the unweighted Jaccard similarity

$$J_s(u,v) = \frac{|\Gamma_s(u) \cap \Gamma_s(v)|}{|\Gamma_s(u) \cup \Gamma_s(v)|} \tag{2}$$

where $\Gamma_s(u)$ denotes the neighbors of node u in view s.

In the case of inter-view coupling, the same node in different views may have different neighborhood. Similarly, we initialize the inter-view coupling strength as follows.

Definition 3 (Initialization of the inter-view coupling strength in unweighted graph)**.** In the initial state, the inter-view coupling strength is assigned using the unweighted Jaccard similarity

$$J_{sr}(u) = \frac{|\Gamma_s(u) \cap \Gamma_r(u)|}{|\Gamma_s(u) \cup \Gamma_r(u)|} \tag{3}$$

Fig. 2 illustrates an example to compute $J_{sr}(u)$. As shown in the figure, $\Gamma_s(u) = \{u, a, b, c, e, v\}$, $\Gamma_r(u) = \{u, a, b, c, v\}$. Thus, $J_{sr}(u) = \frac{5}{6}$.

3.2.2. Weighted graph

For the multi-view graph with the weighted intra-view connection, the information of weights should be considered. To this end, we make the definitions as follows.

Definition 4 (Initialization of the intra-view connection strength in weighted graph)**.** In the initial state, the intra-view connection strength is initialized by a variant of the weighted Jaccard similarity

$$J_s(u,v) = \frac{\sum_{x \in \Gamma_s(u) \cap \Gamma_s(v)} w_{uxs} + w_{vxs}}{\sum_{\{x,y\} \in \mathcal{E}; x,y \in \Gamma_s(u) \cup \Gamma_s(v)} w_{xys}}. \tag{4}$$

Definition 5 (Initialization of the inter-view coupling strength in weighted graph)**.** In the initial state, the inter-view coupling strength is initialized as

$$J_{sr}(u) = \frac{\sum_{x \in \Gamma_s(u) \cap \Gamma_r(u) \setminus \{u\}} w_{uxs} + w_{uxr}}{\sum_{\{x,y\} \in \mathcal{E}; x,y \in \Gamma_s(u) \cup \Gamma_r(u)} w_{xys} + w_{xyr}}. \tag{5}$$

Fig. 3 illustrates an example to obtain $J_s(u, v)$ and $J_{sr}(u)$. In view s, $\Gamma_s(u) \cap \Gamma_s(v) = \{c, e, u, v\}$, and the weight sum for all the edges in the neighborhood of node u and node v is 5.845, thus $J_s(u, v) = \frac{0.43+0.375+0.5+0.43+0.44+0.44}{2*5.845} = 0.22$. In view s and view r, except for node u, the common neighbors for node u, i.e. $\Gamma_s(u) \cap \Gamma_r(u) \setminus \{u\} = \{a, b, c, v\}$. Therefore, the initial value of the inter-view coupling for node u between view s and view r is $J_{sr}(u) = \frac{1.49+2.075}{2*2.795+2*3.505} = 0.28$.

3.3. Update procedure

Due to the mutual interaction among nodes, the relationship will be affected gradually. That is, the strength of the intra-view connection and the inter-view coupling should be updated.

3.3.1. Intra-view connection

The update of the intra-view connection integrates the information from both the intra-view connection and the inter-view coupling. In order to discover the metric space with stronger discriminability of cluster structure, we focus on the interplay in the local way instead of the collective information over the entire network. It's natural for a node to interact only with its neighbors. For both the intra-view connection and the inter-view coupling associated with a single node u, any change of u or its adjacent nodes can result in the update of its associated edge weights. According to the local topological structure, inspired by [13], we can divide the influence type into three scenarios, i.e., inter-view coupling adjusted influence from directly linked nodes, common neighbors, and exclusive neighbors. We will make a detailed description of these three influences below.

Inter-view Coupling Adjusted Influence from Directly Linked Nodes. It's obvious that the connection is influenced by the two directly linked nodes (see Fig. 4(b)). Each node's behavior can affect its neighbors, leading to the change of their cohesiveness. For a multi-view graph, each node has couplings with itself among distinct views, which means the behavior in one view can affect its behavior in other views. Apart from the influence from the directly linked nodes, the inter-view coupling also makes some influence. Thus, when considering the influence from the directly linked nodes, we should take the inter-view coupling into account. We use notation DI_{uvs} to represent the inter-view coupling adjusted influence from directly linked nodes on the connection between node u and node v in view s as follows:

$$DI_{uvs} = \frac{f(J_s(u,v))c_{us}}{k_{us} + c_{us}} + \frac{f(J_s(u,v))c_{vs}}{k_{vs} + c_{vs}} \tag{6}$$

where $f(x) = \sin(x)$ is a coupling function, which is proven to be effective compared with the other functions [13]. The total inter-view coupling strength c_{us} (c_{vs}) in the numerator is for the consideration of cross-view behaviours of two directly linked nodes. That is, a node with stronger inter-view coupling is more consistent in the corresponding views, which is more helpful for the interplay to its neighbors [16]. While for the total intra-view interaction strength k_{us} (k_{vs}) and total inter-view coupling strength c_{us} (c_{vs}) in the denominator, they are for considering the distinct influences of two directly linked nodes. Namely, it's more difficult to influence node with higher strength, while nodes with relatively lower strength can be more easily affected by their neighbors [13].

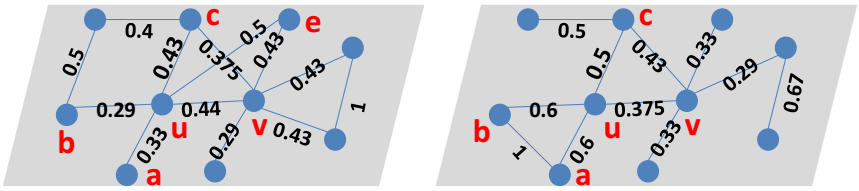
Inter-view Coupling Adjusted Influence from Common Neighbors. In general, besides the influence from directly linked nodes, their relationship can be affected from their common neighbors. Before discussing the influence from the common neighbors, let's define the common neighbors as follows.

Definition 6 (Common neighbors for node u and node v in view s). The common neighbors of node u and node v in view s are defined to be the common nodes in $\Gamma_s(u)$ and $\Gamma_s(v)$ excluding nodes u and node v themselves,

$$\mathcal{CN}_s(u,v) = (\Gamma_s(u) \setminus \{u\}) \cap (\Gamma_s(v) \setminus \{v\}). \tag{7}$$

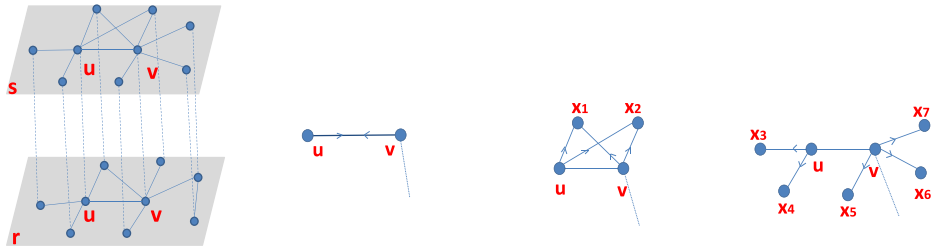
Obviously, good relationship with their common neighbors helps building stronger intra-view connection between them. Take Fig. 4(c) as an example. For node u and node v, when node x_1 is a close friend to both of them, there is a high probability that they can be good friends. Namely, the relationship between node u and node v becomes stronger under the influence from their common neighbor x_1. In the multi-view case, the inter-view coupling of node x_1 affects the intra-view connection between node u and node v indirectly (i.e., through the connection between node u and node x_1 which already contains the coupling information of node x_1). On the other hand, couplings of node u and node v

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012



(a) Neighborhoods of nodes u and v in view s (b) Neighborhoods of nodes u and v in view r

Fig. 3. Example to show how to obtain the initial value for the intra-view connection and inter-view coupling from the original graph with weighted intra-view connection.



(a) (b) (c) (d)

Fig. 4. Example to show the inter-view coupling adjusted influence from directly linked nodes, common neighbors and exclusive neighbors for each intra-view connection. (a) shows the neighborhood for node u and node v in view s and view r. (b) is based on (a) and represents the inter-view coupling adjusted influence from the two directly linked nodes, i.e., node u and node v are on the connection between them. (c) shows the pattern of common neighbors influence, and in view s, node x_1 and node x_2 are common neighbors of node u and node v. (d) represents the inter-view coupling adjusted influence from exclusive neighbors. In view s, the exclusive neighbors of node u w.r.t. node v are node x_3 and x_4, and similarly, the exclusive neighbors of node v w.r.t. node u are node x_5, node x_6 and node x_7.

still have direct influence to their connection. Accordingly, we define the inter-view coupling adjusted influence from the common neighbors as follows,

$$CI_{uvs} = \sum_{x \in \mathcal{CN}_s(u,v)} \left(\frac{f(J_s(u,x))J_s(v,x)}{k_{us} + c_{us}} + \frac{f(J_s(v,x))J_s(u,x)}{k_{vs} + c_{vs}} \right). \quad (8)$$

Inter-view Coupling Adjusted Influence from Exclusive Neighbors. When node u or node v has exclusive neighbors, the exclusive neighbors may have the stretching influence on the relationship between the two nodes [13].

Definition 7 (Exclusive neighbors of node u w.r.t. node v in view s). To obtain the exclusive neighbors, we can remove the common neighbors of node u and node v from $\Gamma_s(u)$

$$\mathcal{EN}_s(u \backslash v) = \Gamma_s(u) \backslash \{\Gamma_s(u) \cap \Gamma_s(v)\}. \quad (9)$$

In Fig. 4(d), each exclusive node tries to attract node u to get closer, and if the exclusive neighbor is similar to node v, the mutual interaction of node u and node v is strengthened, i.e. a higher weight value of their connection. And similarly, if the exclusive neighbor is not similar to node v, the influence from exclusive neighbors should be negative, resulting in the decrease of the strength of their interaction. In this case, the weight value for their connection should be decreased [13]. While in the multi-view graph, similar to the inter-view coupling adjusted influence from common neighbors, we focus on the intra-view connection between $\mathcal{EN}_s(u \backslash v)$ and u. To determine whether the influence from exclusive neighbor is positive or negative, a cohesion parameter λ is introduced [13], which will be further discussed in Section 4.3,

$$\rho_s(x,u) = \begin{cases} J_s(x,v) & J_s(x,v) \geq \lambda \\ J_s(x,v) - \lambda & \text{otherwise} \end{cases} \quad (10)$$

where $\rho_s(x, u)$ characterizes the degree of the positive or negative influence on the interaction between node u and node v. Accordingly, we define the measure for the inter-view coupling adjusted influence from the exclusive neighbors as follows,

$$EI_{uvs} = \sum_{x \in \mathcal{EN}_s(u \backslash v)} \left(\frac{1}{k_{us} + c_{us}} \times f(J_s(x,u)) \times \rho_s(x,u) \right) + \sum_{y \in \mathcal{EN}_s(v \backslash u)} \left(\frac{1}{k_{vs} + c_{vs}} \times f(J_s(y,v)) \times \rho_s(y,v) \right). \quad (11)$$

Combining all these three patterns of inter-view coupling adjusted influence, the intra-view connection is updated over time as follows,

$$J_s(u,v,t+1) \leftarrow J_s(u,v,t) + DI_{uvs}(t) + CI_{uvs}(t) + EI_{uvs}(t) \quad (12)$$

where $J_s(u, v, t+1)$ is the updated intra-view connection between node u and node v in view s in the $(t+1)$th iteration, and $DI_{uvs}(t)$, $CI_{uvs}(t)$, $EI_{uvs}(t)$ indicate the inter-view coupling adjusted influence from the directly linked nodes, common neighbors and exclusive neighbors at time t respectively.

3.3.2. Inter-view coupling

To our knowledge, the weighted inter-view coupling has rarely been studied. Most of the existing work just focuses on the unweighted case [14], or assigns each inter-view coupling with the same constant weighting value [17]. In fact, it's possible to have different mutual coupling among views, and therefore assigning different inter-view couplings with various weighting values is more reasonable. Due to the update of the intra-view connections, the inter-view coupling should be updated accordingly. The update of the inter-view coupling also integrates the information from both the intra-view connection and the inter-view coupling.

For a node u in two views s and r, its neighbors in these two views may be different, resulting in two different topological structures. We consider that the more similar the two structures, the stronger the mutual interaction between these two views, i.e. higher value of inter-view coupling [16]. Besides the link structure, higher degree of numeric similarity for two connections also results in increment of the inter-view coupling. To estimate the similarity of node u's behavior in two different views, we use the

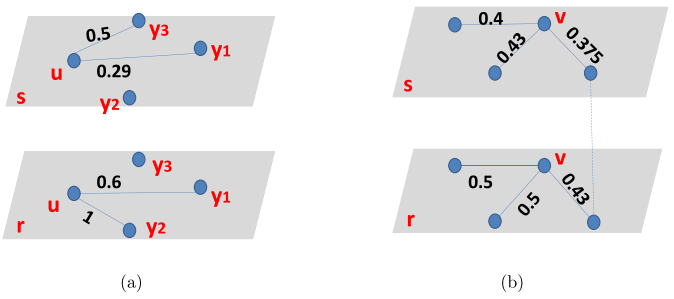


Fig. 5. Example to show the similarity of node behavior in two different views.

weighted Jaccard similarity,

$$DC_{sr}(u) = \frac{\sum_{x\in\Gamma_s(u)\cup\Gamma_r(u)} \min(w_{uxs}, w_{uxr})}{\sum_{x\in\Gamma_s(u)\cup\Gamma_r(u)} \max(w_{uxs}, w_{uxr})}. \tag{13}$$

For better understanding, let's take Fig. 5 as an example. In Fig. 5(a), the neighbors for node u in view s are node y_1 and node y_3, while the neighbors in view r are node y_1 and node y_2. On the other hand, node v in Fig. 5(b) has the same neighborhood in both views. Apparently, the topological structures between two views in Fig. 5(b) has higher degree of similarity than that in Fig. 5(a). Apart from the topological structure, we can see that the common connections in Fig. 5(b) show significant numeric similarity between two views: the maximum difference is $0.5 - 0.4 = 0.1$. While for Fig. 5(a), the maximum difference is $0.6 - 0.29 = 0.31$. A conclusion can be drawn that node v should has stronger inter-view coupling than u between view s and view r. The metric defined in equation Eq. (13) coincides with the above analysis, i.e. $DC_{sr}(u) = \frac{0.29+0+0}{0.6+0.5+1} = 0.138$, and $DC_{sr}(v) = \frac{0.4+0.43+0.375}{0.5+0.5+0.43} = 0.843$.

Besides the influence from the neighboring connections, the couplings of the common neighbors across two views also affect a node's behaviors in different views.

Definition 8 (Common neighbors of node u across view s and view r). The common neighbors of node u across view s and view r are defined to be the common nodes in $\Gamma_s(u)$ and $\Gamma_r(u)$ excluding node u itself,

$$\mathcal{CN}_{sr}(u) = \{\Gamma_s(u) \cap \Gamma_r(u)\} \backslash \{u\}. \tag{14}$$

Naturally, if the friends of one user are extremely active in several social platforms and have strong inter-view coupling among them, there is a large probability for the user to behave more active resulting in increment for the coupling strength. Thus, for node u, the inter-view coupling between view s and r in the $(t+1)$th iteration is updated as

$$J_{sr}(u, t+1) \leftarrow \frac{\sum_{x\in\mathcal{CN}_{sr}(u)} \big(DC_{sr}(u,t) + DC_{sr}(x,t)\big)\big(J_s(u,x,t) + J_r(u,x,t)\big)}{2\big(k_{us}(t) + k_{ur}(t)\big)} \tag{15}$$

where k_{us} is used to indicate that the percentage of common neighbors between two views is used rather than the number of common neighbors. Namely, two nodes may have the same number of common neighbors but different degrees across two views, and it is believed that the node with lower degree should have stronger coupling.

From the above description, different strategies are used to update the two types of strengths of the multi-view graph. The intra-view connection updates in each iteration by considering the inter-view coupling adjusted influence from directly nodes, common neighbors and exclusive neighbors, as well as the intra-view connection in the previous iteration. To update the inter-view coupling, we take into account the behavior similarity of nodes in different views and the influence from their neighbors. The iteration proceeds until convergence.

3.4. Algorithm summary and demonstration

The proposed *Multi-view DML* algorithm is summarized in Algorithm 1. In the proposed algorithm, the iteration stops after convergence, i.e. the increment (decrement) difference between two consecutive steps is smaller than ϵ or the iteration number has reached the pre-defined maximum value. In this case, the steady state of the multi-view graph is obtained, resulting in a more discriminative cluster structure.

Algorithm 1 Multi-view DML

Input: The intra-view connection matrix $A = \mathbb{R}^{N\times N\times S}$ and the inter-view coupling matrix $C = \mathbb{R}^{N\times S\times S}$ of the multi-view graph $G(\mathcal{V}, \mathcal{E}, \mathcal{C}, \mathcal{W})$; cohesion parameter λ; ϵ.

Output: The steady state of the multi-view graph, i.e. the learned metric of the multi-view graph with more discriminative cluster structure.

1: Initialize the intra-view connection strength $J_s(u, v)$ by Eq. (2) or Eq. (4).
2: Initialize the inter-view coupling strength $J_{sr}(u)$ by Eq. (3), or Eq. (5).
3: **repeat**
4: Update the intra-view connection strength $J_s(u, v)$ by Eq. (12).
5: Update the inter-view coupling strength $J_{sr}(u)$ by Eq. (15).
6: **until** Convergence

For illustration, Fig. 6 shows the updating procedure of *Multi-view DML*. The original state is presented in the left of Fig. 1, where only the link structure is obtained. In Fig. 6, each node interacts with its neighbors and the corresponding connections and couplings are iteratively updated based on the designed model (Eqs. (12) and (15)). When the iteration converges, the steady state for multi-view graph is obtained, i.e. the learned metric of the multi-view graph with more discriminative cluster structure. Then the graph partitioning algorithms (e.g., permanence [6] and modularity[17]) are applied to discover the cluster structure (see the last figure with four colors in Fig. 6).

3.5. Complexity analysis

The computational time complexity is $O(|\mathcal{E}| + |\mathcal{C}|)$ in the initial state, as we need to compute the initial value of each intra-view connection strength and the inter-view coupling strength. In the update procedure, we still need to compute the Jaccard similarity for the exclusive neighbors. Thus the computational time complexity is $O(k \cdot |\mathcal{E}|)$ where k is the average number of the exclusive neighbors between two nodes. In each iteration, we need to update the intra-view connection strength and the inter-view coupling strength, thus the time complexity is $O(T \cdot (|\mathcal{E}| + |\mathcal{C}|))$, where T is the number of iterations. The whole time complexity is $O(|\mathcal{E}| + |\mathcal{C}| + k \cdot |\mathcal{E}| + T \cdot (|\mathcal{E}| + |\mathcal{C}|))$.

4. Experiments

In this section, we will evaluate our proposed algorithm by conducting extensive experiments on both unweighted and weighted graphs including synthetic and real-world graphs.

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012

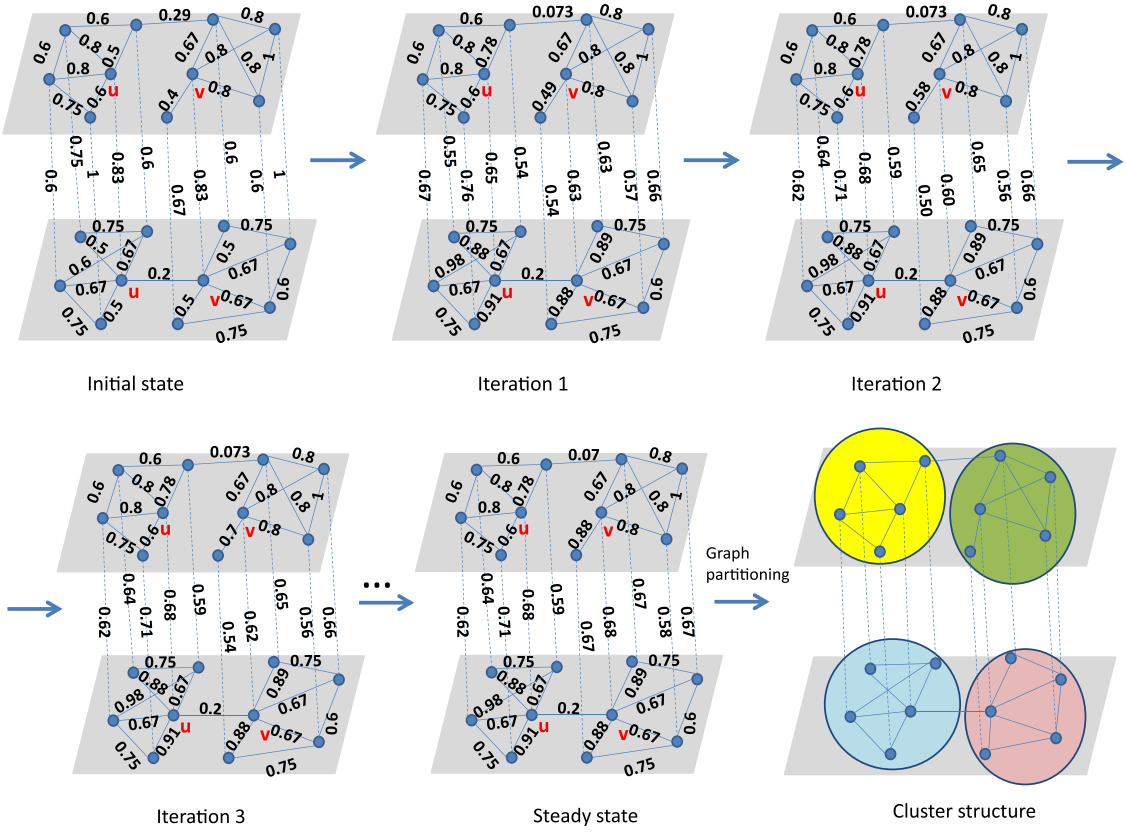


Fig. 6. Demonstration of graph updating from the initial state to the steady state. After applying the graph partitioning algorithms, the partitioning results are generated as shown in different colors.

4.1. Experimental setting and evaluation metrics

In our experiments, two popular multi-view graph partitioning algorithms, namely the *permanence* method [18] and the *modularity* method [17] are used to show that our algorithm is able to learn the more discriminative metric by comparing the graph in the original state and the graph in the steady state from the following two perspectives.

1. **The discriminability**.
   The *permanence* score and the *modularity* score are used as evaluators to compare the discriminability of the graph in the original state and in the steady state. Both of them are used to measure how well the graph can be partitioned without the ground-truth. The permanence method detects the graph cluster structure from the viewpoint of individual nodes and the corresponding permanence score measures the discriminability of graph considering the influence from internal and external connections. On the other hand, the modularity method partitions the graph based on the random null model assumption that each node has the equal probability to connect other nodes while keeping the degree unchanged, and the modularity score uses the expected cuts to measure the clustering quality. Higher permanence and modularity score values indicate the higher discriminability of the topological structure.
2. **The accuracy of the discovered clusters**.
   From this perspective, with the availability of the ground-truth cluster labels, we compare the accuracy of the discovered clusters (generated by the *permanence* method and the *modularity* method) in the graph in the original state and in the steady state. In particular, we use Normalized Mutual Information (NMI) to measure the accuracy of the graph partitioning results compared with the ground-truth. NMI [34] is used for measuring the similarity of two graph partitioning results or the similarity between the ground-truth cluster labels and the predicted cluster labels. Given a graph $\mathcal{X}$ consisting of n nodes, the predicted cluster labels π of c clusters and the ground-truth cluster labels ζ of $\hat{c}$ clusters, a confusion matrix is formed first, where entry (i, j), $n_i^{(j)}$ gives the number of nodes in the predicted cluster i and the ground-truth cluster j. Then NMI can be computed from the confusion matrix [34]

$$NMI(\zeta, \pi) = \frac{2\sum_{l=1}^{c}\sum_{h=1}^{\hat{c}} \frac{n_l^{(h)}}{n} \log \frac{n_l^{(h)} n}{\sum_{i=1}^{c} n_i^{(h)} \sum_{i=1}^{\hat{c}} n_l^{(i)}}}{H(\pi) + H(\zeta)} \tag{16}$$

where $H(\pi) = -\sum_{i=1}^{c} \frac{n_i}{n} \log \frac{n_i}{n}$ and $H(\zeta) = -\sum_{j=1}^{\hat{c}} \frac{n^{(j)}}{n} \log \frac{n^{(j)}}{n}$ are the Shannon entropy of the predicted cluster labels π and the ground-truth cluster labels ζ respectively, with n_i and $n^{(j)}$ denoting the number of nodes in the predicted cluster i and the ground-truth cluster j. A higher NMI indicates the predicted and ground-truth cluster labels match better.

4.2. Testing graphs

4.2.1. Unweighted graphs

The graphs with unweighted intra-view connections are summarized in Table 2. In these testing graphs, after obtaining a graph, it is taken as the first view, on which some perturbations (i.e., remove some edges randomly) are then applied to generate another graph as the second view. In this way, a two-view graph is generated with the same ground-truth clusters for both views.

1. Football: Girvan and Newman collected this graph from the schedule of Division I games for the 2000 season in the United States college football game [35]. The teams are represented as

Table 2
Properties of unweighted graphs: n is the number of nodes, e is the number of edges, $< k >$ is the average degree, k_{max} is the maximum degree and *#class* is the number of clusters. ULFR stands for unweighted LFR.

Graphs	n	e		$< k >$		k_{max}		*#class*
		View 1	View 2	View 1	View 2	View 1	View 2	
Football	115	613	593	10.661	10.365	12	12	12
Railway	301	1224	1174	8.133	7.801	48	48	22
ULFR($\mu = 0.1$)	3000	14942	13942	9.961	9.295	16	16	267
ULFR($\mu = 0.3$)	3000	10033	9033	6.689	6.022	10	10	427

Table 3
Properties of weighted graphs: n is the number of nodes, e is the number of edges, $< k >$ is the average degree, k_{max} is the maximum degree and *#class* is the number of clusters. WLFR stands for weighted LFR.

Graphs	n	e		$< k >$		k_{max}		*#class*
		View 1	View 2	View 1	View 2	View 1	View 2	
VIS/NIR	1056	12467	11210	23.612	21.231	87	57	22
WLFR($\mu = 0.1$)	3000	22566	21566	15.044	14.377	20	20	190
WLFR($\mu = 0.3$)	3000	10762	9762	7.175	6.508	12	12	382

nodes and the edges denote the regular-season games between the two corresponding teams.

2. Railway: Ghosh et al. collected this graph derived from the India Railway system [36]. The railway stations are regarded as nodes in the graph, and two nodes are connected if there exists at least one train-route between them. The stations are assigned to the belonging states which are taken as clusters, as the number of the trains within each state is larger than the number of trains between two states, and thereby the ground-truth is already known.
3. ULFR($\mu = 0.1,\ \mu = 0.3$): Lancichinetti-Fortunato-Radicchi (LFR) benchmark is an algorithm that generates synthetic benchmark networks that resemble real-world networks [37]. The LFR benchmark with the ground-truth is generated by specifying the following parameters as input: the number of nodes n, the average degree $< k >$, the maximum degree k_{max} and the mixing parameter μ. The mixing parameter is defined as the fraction of the external connections of each node to its degree, measuring the discriminability. Lower value of the mixing parameter indicates stronger topological structure, i.e. higher degree of discriminability. By using two different values of μ, two different LFR benchmark networks can be generated.

4.2.2. Weighted graphs

The graphs with weighted intra-view connections are summarized in Table 3. Similarly, the second view can be generated by perturbation from the first view, and then a two-view graph can be generated. For the graph that has weight larger than 1, all the weights are mapped into the range of [0,1] by dividing all the weights by the largest weight.

1. VIS/NIR: This real-world multi-view graph is generated from different face images captured in Visible Light and Near-IR illumination [16]. Following [16], the similarity matrix is obtained respectively for each view which is further processed by removing entries with similarity below a threshold (e.g., 0.975 in our experiment).
2. WLFR($\mu = 0.1,\ \mu = 0.3$): Unlike the previous unweighted case, the weighted LFR benchmark networks can be generated by using the method in [38].

4.3. Parameter analysis

To analyze how the value of the cohesion parameter λ affects the performance of the proposed method in discriminative metric learning, we modulate λ from 0 to 1 on both unweighted and weighted LFR benchmarks from the perspectives of both the discriminability and the accuracy of the discovered clusters.

4.3.1. Parameter analysis on unweighted LFR networks

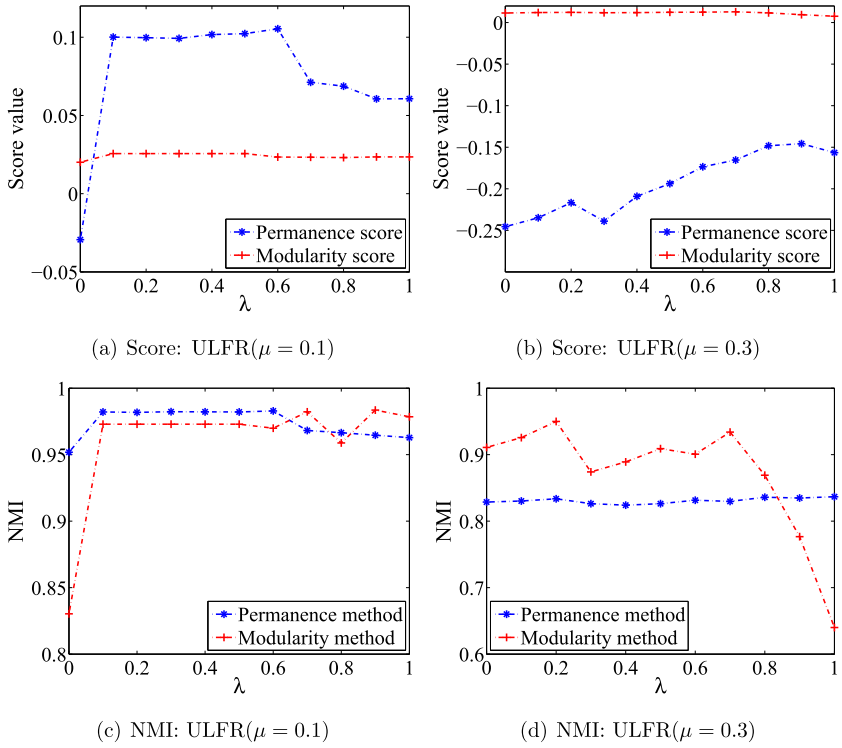
By applying the proposed Multi-view DML algorithm on ULFR($\mu = 0.1$) and ULFR($\mu = 0.3$) with different cohesion parameter λ, different resulting graphs can be generated. To show the discriminability of the different resulting graphs, we apply the graph partitioning methods, namely the *permanence* method [18] and the *modularity* method [17], on the resulting graphs and report the discriminability values in terms of the *permanence* score and the *modularity* score as well as the graph partitioning accuracy in terms of NMI, as plotted in Fig. 7.

From the first row of Fig. 7, we can see that the discriminability in terms of the modularity score is relatively stable on the two graphs when λ changes. And for the permanence score on ULFR($\mu = 0.1$) in Fig. 7(a), the strength of the topological structure has a stable range when λ is between 0.1 and 0.6. The NMI values presented in Fig. 7(c) show similar changing characteristics to the score values in Fig. 7(a). On ULFR($\mu = 0.3$) in Fig. 7(b), there exist more fluctuations in the permanence score than ULFR($\mu = 0.1$) when λ is within the range of [0.1,0.6] due to the weaker topological structure. While for the accuracy of the resulting clusters in terms of NMI, as shown in Fig. 7(d), the modularity method degenerates when $\lambda > 0.7$ and the permanence method behaves relatively stable.

4.3.2. Parameter analysis on weighted LFR networks

Similarly, Fig. 8 shows the parameter analysis results on the two weighted LFR networks. Compared with the unweighted LFR network case, there exist more fluctuations. When measuring the discriminability of topological structure in terms of the *permanence* score and the *modularity* score, the latter behaves relatively stable like the unweighted graph case, while the permanence score degenerates seriously on WLFR($\mu = 0.1$) when $\lambda > 0.5$ as shown in Fig. 8(a). On WLFR($\mu = 0.3$) in Fig. 8(b), the discriminability values start to decrease when $\lambda > 0.5$. When estimating the accuracy of the discovered clusters on WLFR($\mu = 0.1$), the modularity method starts to degenerate when $\lambda > 0.5$ as shown in Fig. 8(c). And on WLFR($\mu = 0.3$) in Fig. 8(d), the modularity method degenerates dramatically when $\lambda > 0.6$, and the worst case is obtained when $\lambda = 0.8$.

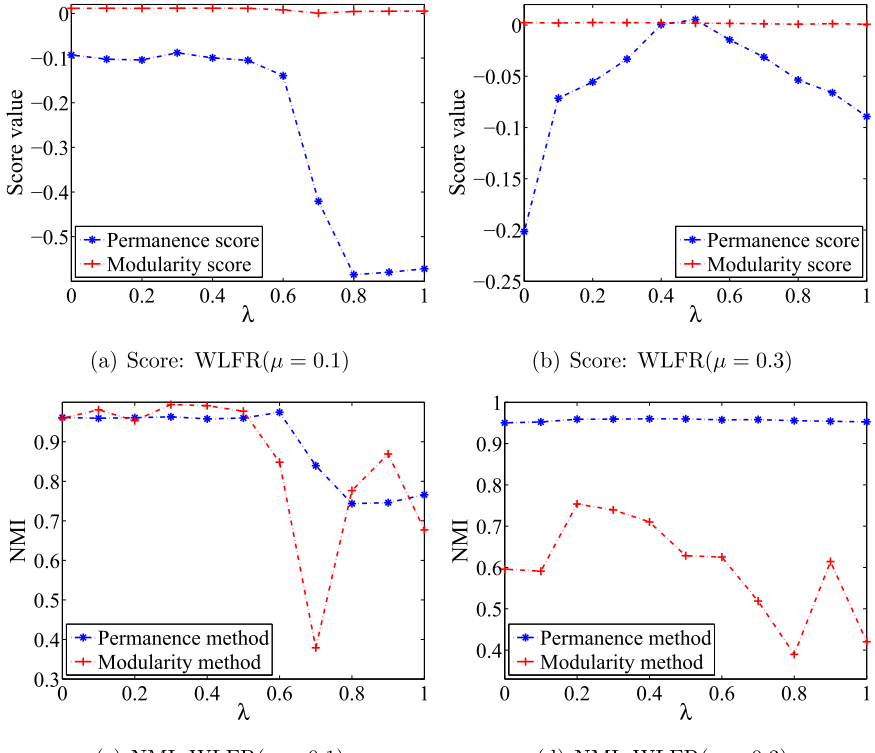
Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012



(a) Score: ULFR($\mu = 0.1$) (b) Score: ULFR($\mu = 0.3$)

(c) NMI: ULFR($\mu = 0.1$) (d) NMI: ULFR($\mu = 0.3$)

Fig. 7. Parameter analysis on the *unweighted* LFR networks: the first row shows the values of the permanence score and the modularity score as a function of λ; the second row shows the NMI values obtained by the permanence method and the modularity method as a function of λ.

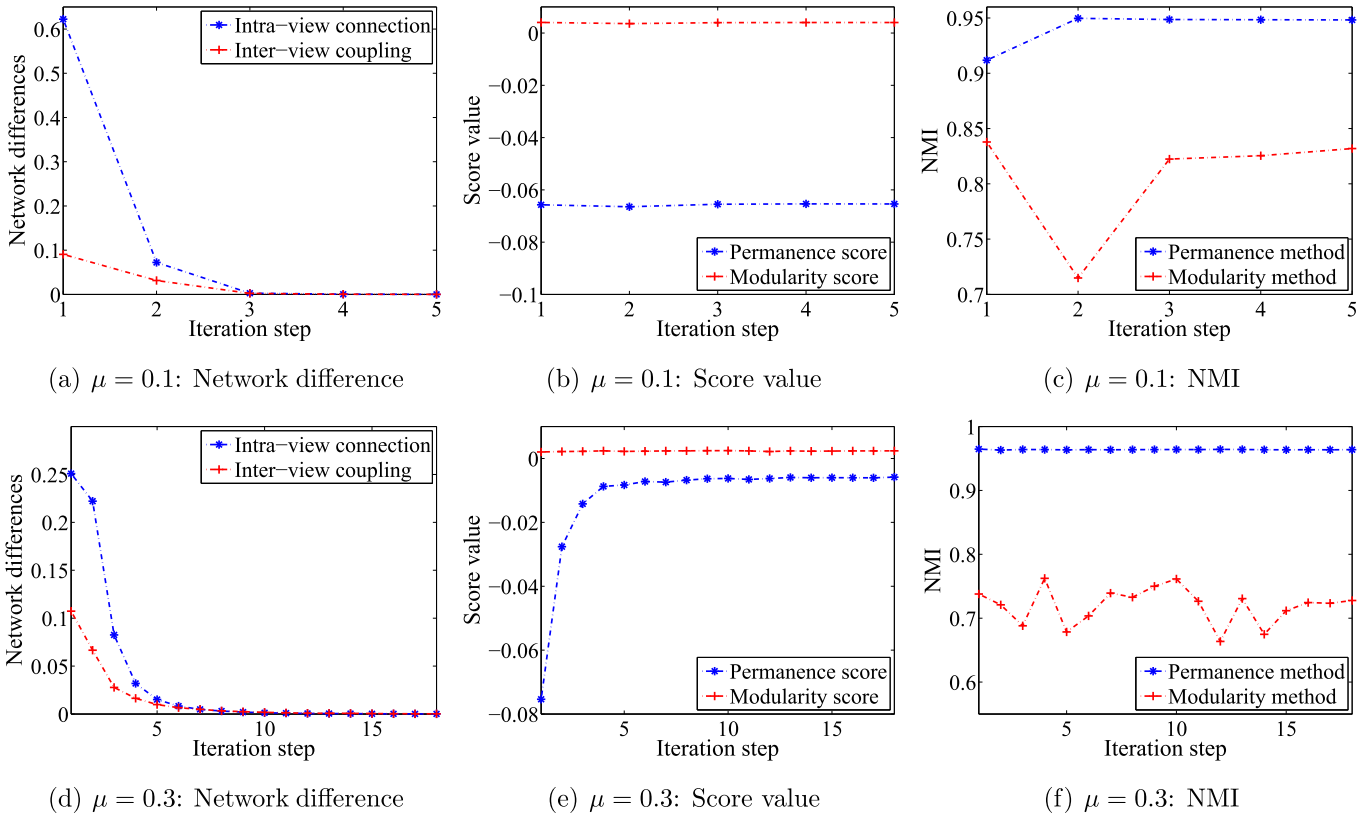


(a) Score: WLFR($\mu = 0.1$) (b) Score: WLFR($\mu = 0.3$)

(c) NMI: WLFR($\mu = 0.1$) (d) NMI: WLFR($\mu = 0.3$)

Fig. 8. Parameter analysis on the *weighted* LFR networks.

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012



(a) $\mu = 0.1$: Network difference (b) $\mu = 0.1$: Score value (c) $\mu = 0.1$: NMI

(d) $\mu = 0.3$: Network difference (e) $\mu = 0.3$: Score value (f) $\mu = 0.3$: NMI

Fig. 9. Convergence analysis on *unweighted* LFR networks with different $\mu = 0.1, 0.3$: (a) (d) The network difference in terms of average intra-view connections and average inter-view coupling computed between two consecutive iterations; (b) (e) The discriminability in terms of the permanence score and the modularity score in each iteration; (c) (f) The NMI value obtained by the permanence method and the modularity method in each iteration.

As discussed above, it's safe to conclude that there exists a tendency of improvement when $\lambda \leq 0.2$ on almost all the LFR networks, and when λ ranges from 0.2 to 0.5, the performance of the permanence method and the modularity method is relatively better than that in the range of [0.5,1]. The influence of parameter λ shown here coincides its influence to the intra-view connection, i.e. higher or lower value of λ yields lower degree of discriminability of topological structure. According to the above parameter analysis, we set the value of λ as 0.5.

4.4. Convergence analysis

In order to show the convergence property and determine the value of parameter ϵ, some convergence analysis is conducted to evaluate how the network difference, network discriminability and the accuracy of the discovered clusters change in each iteration on the LFR networks.

4.4.1. Convergence analysis on unweighted LFR networks

The ULFR($\mu = 0.1$) and ULFR($\mu = 0.3$) networks are used to analyze the convergence property of the proposed method. In Fig. 9, the score values in terms of the permanence score and the modularity score start to converge in the earlier iteration step. For the NMI values obtained by the permanence method and the modularity method, the curves are relatively stable in the case of the permanence method while in the case of the modularity method, there exist more fluctuations and the NMI values start to converge when the network differences reach a very small value. That is, in Fig. 9(c), the NMI values obtained by the modularity method converge after the third iteration while the network differences are very closed to zero.

4.4.2. Convergence analysis on weighted LFR networks

Similarly, we conduct some convergence analysis experiments on WLFR($\mu = 0.1$) and WLFR($\mu = 0.3$), and the results are plotted in Fig. 10. The score values in terms of the permanence score and the modularity score exhibit strong convergence property as shown in Fig. 10(b) and (e). Higher value of μ yields weaker graph structure and we can see that in Fig. 10(c) and (f), more fluctuations appear as μ increases. Similar to ULFR, the NMI values start to converge when the network differences reach a pretty small value.

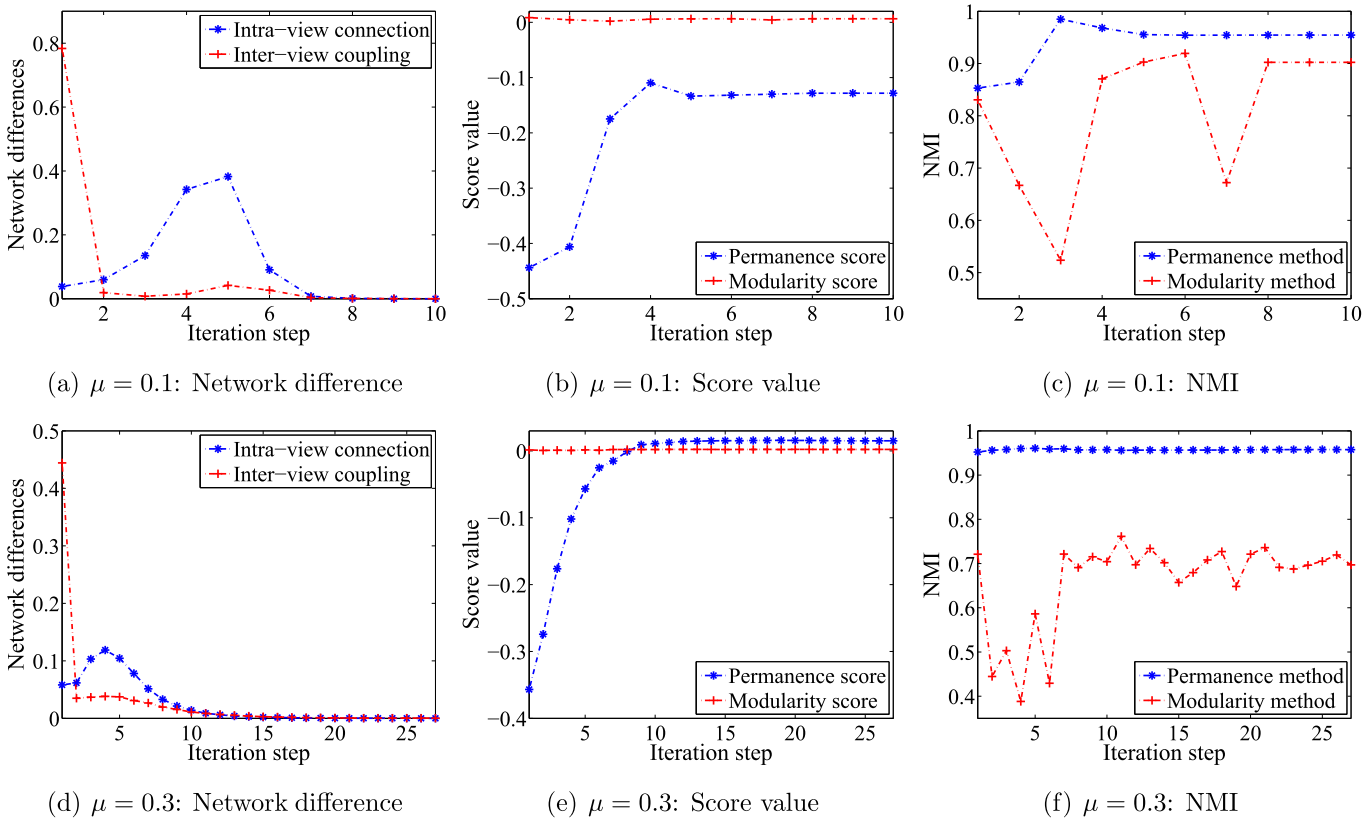
As discussed above, the networks exhibit strong convergence property as shown in Figs. 9 and 10 when we assign $\epsilon = 10^{-4}$. Therefore, we use this value in the following experiments.

4.5. Evaluation on unweighted graphs

In this section, we evaluate the performance of our algorithm on graphs with unweighted intra-view connections.

To compare the discriminability of the topological structure in the original state with that in the steady state, we transform the original unweighted graph into the corresponding weighted graph using Jaccard similarity. Comparison results in terms of NMI against the ground-truth by applying two graph partitioning algorithms (i.e. permanence and modularity) in the two states (i.e. the original state and the steady state) of the same graph are reported in Table 4. We can see that the NMI value obtained by the permanence method and the modularity method from the graph in the steady state is higher than that from the original graph. In order to exhibit the discriminability enhancement process to the topological structure by our algorithm, we use the permanence score and the modularity score to measure the discriminability of the topological structure in the original state as well as in each iteration

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012



(a) $\mu = 0.1$: Network difference (b) $\mu = 0.1$: Score value (c) $\mu = 0.1$: NMI

(d) $\mu = 0.3$: Network difference (e) $\mu = 0.3$: Score value (f) $\mu = 0.3$: NMI

Fig. 10. Convergence analysis on *weighted* LFR networks with different $\mu = 0.1, 0.3$.

Table 4
Comparison results in terms of NMI obtained by the permanence method and the modularity method in the original unweighted graphs (Original) and the graphs in the steady state (Steady) where improvement is defined as Improvement = (Steady−Original)/Original × 100%.

Graphs	Algorithms	Original	Steady	Improvement
Football	Permanence	0.8361	**0.8782**	5%
	Modularity	0.7198	**0.8784**	22%
Railway	Permanence	0.6913	**0.7200**	4.2%
	Modularity	0.6541	**0.6659**	1.8%
ULFR($\mu = 0.1$)	Permanence	0.9118	**0.9482**	4.0%
	Modularity	0.5689	**0.8319**	46.2%
ULFR($\mu = 0.3$)	Permanence	0.9616	**0.9638**	0.2%
	Modularity	0.4462	**0.7277**	63.1%

of the updating process. Relative values of the permanence score and the modularity score computed in each iteration against that in the original state are reported in Fig. 11. The relative value is computed as

$$\text{Relative value} = \frac{\text{Value in the current iteration} - \text{Value in the original state}}{|\text{Value in the original state}|} \tag{17}$$

which measures the increment rate obtained in each iteration compared with that in the original state. When the index of iteration is zero, the graph is in the original state and thereby the relative values are all zero. From the results plotted in Fig. 11 and the higher graph partitioning accuracy presented in Table 4, it is safe to conclude that the proposed method is able to learn a more discriminative metric for the unweighted multi-view graphs with detailed analysis given below.

4.5.1. On the two real-world unweighted networks

From Fig. 11(a), we can see the discriminability of the Football topological structure is enhanced significantly in terms of the permanence score and the modularity score after the first updating iteration, and then converges gradually. After the sixth iteration, the relative values of the permanence score and the modularity score nearly keep unchanged. Overall, the discriminability of the topological structure has increased by 111.8% in terms of the permanence score and 59.9% in terms of the modularity score.

As shown in Fig. 11(b), the discriminability of the Railway topological structure in terms of the modularity score has been strengthened continually in the updating process. The discriminability of the topological structure has increased by 34.1% in the steady state. While under the estimation of the permanence score, the discriminability of Railway is enhanced after the first iteration and then converges gradually with a little fluctuation. And in this case, the discriminability of the topological structure has increased by 104.8% after the last iteration.

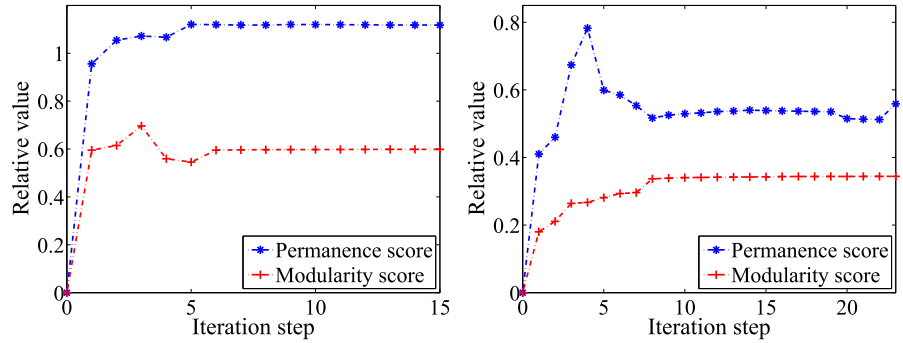
In Table 4, we can see that the two graph partitioning methods discover the clusters with higher accuracy on the two real-world unweighted graphs in the steady state than in the original state, in particular on the Football graph where the improvement as high as 22% has been achieved.

4.5.2. On the Two Unweighted LFR Networks

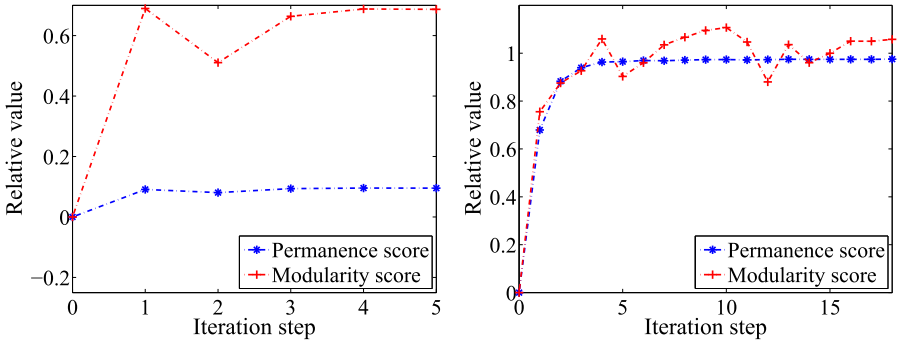
On the ULFR($\mu = 0.1$) network, as shown in Fig. 11(c), the discriminability of this LFR topological structure has increased by 9.5% in terms of the permanence score and 68.7% in terms of the modularity score. For the ULFR($\mu = 0.3$) network (see Fig. 11(d)), the discriminability of this network has increased by 97.5% in terms of the permanence score and 105.8% in terms of the modularity score.

From Table 4, we can see that, in terms of the accuracy, both the permanence method and the modularity method per-

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012

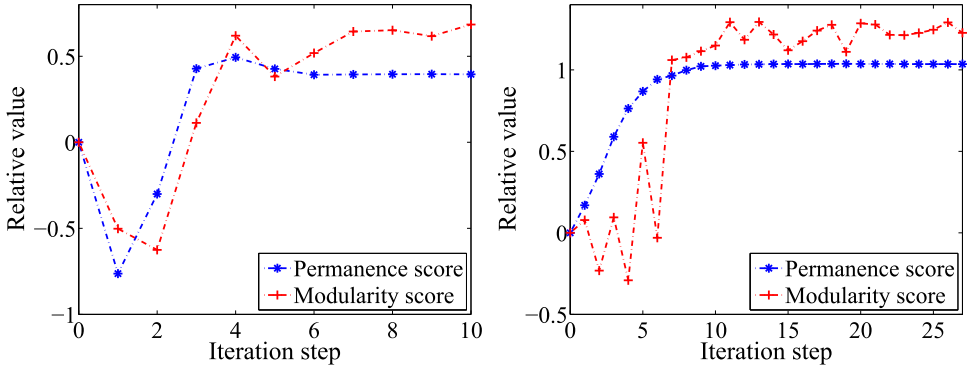


(a) Football (b) Railway



(c) ULFR($\mu = 0.1$) (d) ULFR($\mu = 0.3$)

Fig. 11. Evaluation on the four unweighted graphs: The relative values of the permanence score and the modularity score as a function of the iteration step.



(a) WLFR($\mu = 0.1$) (b) WLFR($\mu = 0.3$)

Fig. 12. Evaluation on the two weighted graphs: The relative values of the permanence score and the modularity score as a function of the iteration step.

form better for discovering clusters in the steady state on the two ULFR networks. For instance, significant improvements have been achieved by the modularity method, where the NMI value has increased from 0.5689 to 0.8319 with improvement of 46.2% on ULFR($\mu = 0.1$) and from 0.4462 to 0.7277 with improvement of 63.1% on ULFR($\mu = 0.3$).

4.6. Evaluation on weighted graphs

In this section, we evaluate the performance of our algorithm on the three weighted graphs. To investigate how the discriminability of the topological structure is enhanced on WLFR graphs, we also plot the relative values of the permanence score and the modularity score in Fig. 12. From the figure, it can be seen that our algorithm has gained significant improvement of the discriminability in terms of both two scores. Furthermore, by applying the permanence method and the modularity method to cluster the nodes on the three weighted graphs, the comparison results against the ground-truth have shown significant improvements achieved by the proposed method in terms of graph partitioning accuracy (see Table 5). A conclusion can be drawn that a more discriminative metric has been learned for the weighted multi-view graph with the detailed analysis given below.

For the WLFR($\mu = 0.1$) network, eventually, the discriminability of the topological structure has been increased by 39.5% in terms of the permanence score and 68.5% in terms of the modularity score. For the WLFR($\mu = 0.3$) network, the discriminability of the topological structure has been increased by 103.5% in terms of the permanence score and 122.7% in terms of the modularity score.

As shown in Table 5, on the VIS/NIR network, Multi-view DML achieves the NMI improvements of 9.2% and 60.02% by applying the permanence method and the modularity method respectively. Especially in terms of the modularity method, the sparse original

Table 5
Comparison results in terms of NMI against the ground-truth using the permanence method and the modularity method in the original weighted graphs (Original) and the graphs in the steady state (Steady) where improvement is defined as Improvement = (Steady−Original)/Original × 100%.

VIS/NIR	Algorithms	Original	Steady	Improvement
	Permanence	0.6864	**0.7496**	9.2%
	Modularity	0.5023	**0.8038**	60.02%
WLFR($\mu = 0.1$)	Permanence	0.9131	**0.9545**	4.5%
	Modularity	0.7359	**0.9022**	22.6%
WLFR($\mu = 0.3$)	Permanence	0.9522	**0.9576**	0.6%
	Modularity	0.5846	**0.6971**	19.2%

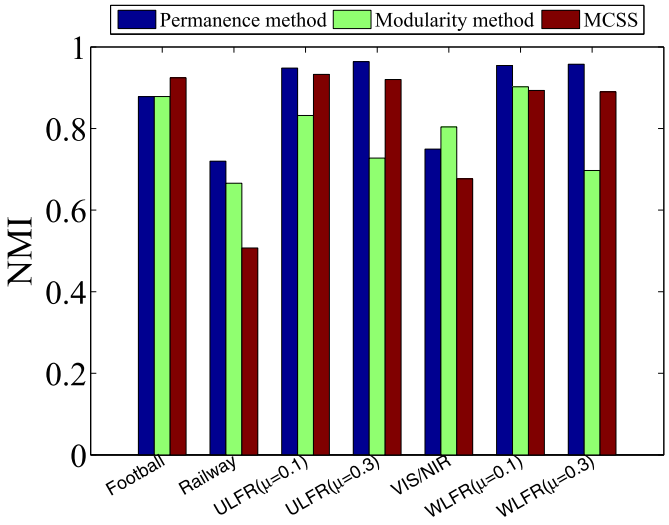


Fig. 13. Comparison results: the NMI values obtained by the MCSS method in the original graphs and by applying the permanence method and the modularity method in the graphs generated through the Multi-view DML.

Table 6
Comparison results in terms of NMI by applying the permanence method and the modularity method on the unweighted graphs generated through the three metric learning methods.

	View	Partition algorithms	SDD	CTD	Multi-view DML
Football	View 1	Permanence	**0.8522**	0.7577	0.8473
		Modularity	0.8262	0.6988	**0.8323**
	View 2	Permanence	0.8208	0.7905	**0.8258**
		Modularity	0.6510	0.7206	**0.7621**
Railway	View 1	Permanence	0.6451	0.6528	**0.6530**
		Modularity	0.5488	0.4703	**0.5793**
	View 2	Permanence	**0.6469**	0.6138	0.6428
		Modularity	0.5346	0.4335	**0.5646**
ULFR($\mu = 0.1$)	View 1	Permanence	0.9615	**0.9698**	0.9615
		Modularity	0.7945	0.7724	**0.7997**
	View 2	Permanence	0.9504	**0.9658**	0.9526
		Modularity	0.7887	0.7256	**0.7901**
ULFR($\mu = 0.3$)	View 1	Permanence	0.9407	0.9259	**0.9416**
		Modularity	0.6835	0.6154	**0.7145**
	View 2	Permanence	**0.9348**	0.9316	**0.9348**
		Modularity	0.7235	0.5588	**0.7745**

Table 7
Comparison results in terms of NMI by applying the permanence method and the modularity method on the weighted graphs generated through the three metric learning methods.

	View	Partition algorithms	SDD	CTD	Multi-view DML
VIS/NIR	View 1	Permanence	0.7029	0.5519	**0.7040**
		Modularity	0.4657	0.423	**0.6325**
	View 2	Permanence	0.7098	0.7070	**0.7185**
		Modularity	**0.9326**	0.9230	0.8822
WLFR($\mu = 0.1$)	View 1	Permanence	0.9651	**0.9856**	0.9654
		Modularity	0.8151	0.8062	**0.8752**
	View 2	Permanence	0.9639	**0.9723**	0.9656
		Modularity	0.8169	0.7952	**0.9013**
WLFR($\mu = 0.3$)	View 1	Permanence	0.9764	0.9139	**0.9791**
		Modularity	0.6145	0.5552	**0.6853**
	View 2	Permanence	0.9696	0.9086	**0.9704**
		Modularity	0.6508	0.5077	**0.6744**

graph leads to lower clustering accuracy as 0.5023. While Multi-view DML updates the intra-view connection and inter-view coupling in an interplay manner and learns more discriminate cluster structure, whereby the NMI value as high as 0.8038 is obtained.

Overall, the experiments on the unweighted and weighted graphs demonstrate that our algorithm not only strengthens the discriminability of the topological structure after the updating procedure, but also allows to extract meaningful clusters in graphs with higher accuracy, i.e. a more discriminative metric has been learned for the multi-view graph partitioning.

4.7. Comparison with the existing methods

In this section, we will compare our method with not only the metric learning method in multi-view clustering, namely Multi-view Clustering via Structured Sparsity (abbr. MCSS) [39], but also the methods designed for single-view graphs, namely Single-view dynamic distance (abbr. SDD) [13] and Commute Time Distance (CTD) [32]. To this end, the existing single-view metric learning methods will be applied to each view separately and compare the discriminability of the learned metric with that learned by our method.

The comparison results with MCSS are shown in Fig. 13. We can see that except on the Football dataset, the NMI values obtained by applying the permanence method in the graphs generated by the proposed multi-view DML are higher than those obtained by the MCSS method in the original graphs. While by applying the modularity method in the graphs generated by the proposed multi-view DML, comparable clustering performance can be achieved.

The permanence method and the modularity method are applied to partition the graphs with the metric learned by SDD, CTD and Multi-view DML. The corresponding clustering accuracy in each view is reported in Table 6 (unweighted graphs) and Table 7 (weighted graphs). We can see that Multi-view DML can provide higher NMI value in most cases. For instance, on the second view of Football, compared with the SDD, our proposed method achieves significant improvement as high as 17.07% in terms of the modularity method.

From the analysis above, we can draw a conclusion that Multi-view DML integrates the information from different views and can provide higher clustering accuracy than the existing multi-view metric learning method and single-view metric learning methods in most cases.

5. Conclusion

In this paper, we have developed an algorithm to learn more discriminative metric for multi-view graph partitioning by considering the interplay of the intra-view connections and the inter-view couplings. We regard the whole graph as a dynamic system with updating the intra-view connection and the inter-view coupling gradually. To update the intra-view connection, inter-view coupling adjusted influence from the directly linked nodes, common neighbors and exclusive neighbors are considered. And for a node in two views, we take into account the intra-view connections associated with this node and the inter-view couplings associated with its neighbors to update the strength of the inter-view coupling. Extensive experiments demonstrate that our algorithm

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012

allows extracting graph of higher discriminability (i.e. higher degree of topological structure discriminability) and also allows finding clusters with higher accuracy.

Acknowledgments

This work was supported by National Key Research and Development Program of China (2016YFB1001003), NSFC (61573387 & 61502543) and Guangdong Natural Science Funds for Distinguished Young Scholar (2016A030306014).

References

[1] J. Scott, Social network analysis, Sociology 22 (1) (1988) 109–127.
[2] S. Wasserman, K. Faust, Social network analysis: methods and applications, 8, Cambridge university press, 1994.
[3] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.
[4] C.-D. Wang, J.-H. Lai, P.S. Yu, NEIWalk: community discovery in dynamic content-based networks, IEEE Trans. Knowl. Data Eng. 26 (7) (2014) 1734–1748, doi:10.1109/TKDE.2013.153.
[5] C.-D. Wang, J.-H. Lai, P.S. Yu, Dynamic community detection in weighted graph streams, in: SDM 2013, 2013, pp. 151–161.
[6] T. Chakraborty, S. Srinivasan, N. Ganguly, A. Mukherjee, S. Bhowmick, On the permanence of vertices in network communities, in: KDD 2014, ACM, 2014, pp. 1396–1405.
[7] Z. Lu, Y. Wen, G. Cao, Community detection in weighted networks: algorithms and applications, in: 2013 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2013, pp. 179–184.
[8] M.E. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (3) (2006a) 036104.
[9] M.E. Newman, Modularity and community structure in networks, Proc. Nat. Acad. Sci. 103 (23) (2006b) 8577–8582.
[10] J. Jin, L. Pan, C. Wang, J. Xie, A center-based community detection method in weighted networks, in: 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, IEEE, 2011, pp. 513–518.
[11] J. Leskovec, K.J. Lang, M. Mahoney, Empirical comparison of algorithms for network community detection, in: WWW 2010, ACM, 2010, pp. 631–640.
[12] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.
[13] J. Shao, Z. Han, Q. Yang, T. Zhou, Community detection based on distance dynamics, in: KDD 2015, ACM, 2015, pp. 1075–1084.
[14] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, Science 328 (5980) (2010) 876–878.
[15] C. Shi, Y. Li, J. Zhang, Y. Sun, P.S. Yu, A survey of heterogeneous information network analysis, IEEE Trans. Knowl. Data Eng. 29 (1) (2017) 17–37.
[16] C.-D. Wang, J.-H. Lai, S.Y. Philip, Multi-view clustering based on belief propagation, IEEE Trans. Knowl. Data Eng. 28 (4) (2016) 1007–1021.
[17] H. Zhang, C.-D. Wang, J.-H. Lai, P.S. Yu, Modularity in complex multilayer networks with multiple aspects: a static perspective, arXiv:1605.06190 arXiv preprint, (2016).
[18] J.-H. Li, P.-Z. Li, C.-D. Wang, J.-H. Lai, Community detection in complicated network based on the multi-viewweighted signed permanence, in: 2016 The 14th IEEE International Symposium on Parallel and Distributed Processing with Applications, 2016, pp. 1589–1596.
[19] X. Zhao, A. Chang, A.D. Sarma, H. Zheng, B.Y. Zhao, On the embeddability of random walk distances, Proc. VLDB Endowment 6 (14) (2013) 1690–1701.
[20] A. Mahmood, M. Small, Subspace based network community detection using sparse linear coding, IEEE Trans. Knowl. Data Eng. 28 (3) (2016) 801–812, doi:10.1109/TKDE.2015.2496345.
[21] A.R. Benson, D.F. Gleich, J. Leskovec, Higher-order organization of complex networks, Science 353 (6295) (2016) 163–166.
[22] Y. van Gennip, H. Hu, B. Hunter, M.A. Porter, Geosocial graph-based community detection, in: 2012 IEEE 12th International Conference on Data Mining Workshops, IEEE, 2012, pp. 754–758.
[23] H. Hu, Y. van Gennip, B. Hunter, A.L. Bertozzi, M.A. Porter, Multislice modularity optimization in community detection and image segmentation, in: 2012 IEEE 12th International Conference on Data Mining Workshops, IEEE, 2012, pp. 934–936.
[24] V. Kawadia, S. Sreenivasan, Sequential detection of temporal communities by estrangement confinement, Scient. Rep. 2 (2012).
[25] M.W. Cole, D.S. Bassett, J.D. Power, T.S. Braver, S.E. Petersen, Intrinsic and task-evoked network architectures of the human brain, Neuron 83 (1) (2014) 238–251.
[26] D. Tao, J. Cheng, M. Song, X. Lin, Manifold ranking-based matrix factorization for saliency detection, IEEE Trans. Neural Netw. Learn. Syst. 27 (6) (2016a) 1122–1134.
[27] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, Y.Y. Tang, Person re-identification by dual-regularized kiss metric learning, IEEE Trans. Image Process. 25 (6) (2016b) 2726–2738.
[28] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S.C. Hoi, M. Satyanarayanan, A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 32 (1) (2010) 30–44.
[29] D.-Y. Yeung, H. Chang, Extending the relevant component analysis algorithm for metric learning using both positive and negative equivalence constraints, Pattern Recognit. 39 (5) (2006) 1007–1010.
[30] F. Yin, C.-L. Liu, Handwritten chinese text line segmentation by clustering with distance metric learning, Pattern Recognit. 42 (12) (2009) 3146–3157.
[31] X. Yin, S. Chen, E. Hu, D. Zhang, Semi-supervised clustering with metric learning: an adaptive kernel method, Pattern Recognit. 43 (4) (2010) 1320–1333.
[32] J.A. Albano, D.W. Messinger, Euclidean commute time distance embedding and its application to spectral anomaly detection, SPIE Defense, Security, and Sensing, 2012, International Society for Optics and Photonics, 2012. 83902G–83902G.
[33] P. Jaccard, Etude comparative de la distribution florale dans une portion des Alpes et du Jura, Impr. Corbaz, 1901.
[34] A. Strehl, J. Ghosh, Cluster ensembles–a knowledge reuse framework for combining multiple partitions, J. Mach. Learn. Res. 3 (Dec) (2002) 583–617.
[35] M. Girvan, M.E. Newman, Community structure in social and biological networks, Proc. Nat. Acad. Sci. 99 (12) (2002) 7821–7826.
[36] S. Ghosh, A. Banerjee, N. Sharma, S. Agarwal, N. Ganguly, S. Bhattacharya, A. Mukherjee, Statistical analysis of the indian railway network: a complex network approach, Acta Physica Polonica B Proc. Suppl. 4 (2) (2011) 123–138.
[37] A. Lancichinetti, S. Fortunato, F. Radicchi, Benchmark graphs for testing community detection algorithms, Phys. Rev. E 78 (4) (2008) 046110.
[38] A. Lancichinetti, S. Fortunato, Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities, Phys. Rev. E 80 (1) (2009) 016118.
[39] H. Wang, F. Nie, H. Huang, Multi-view clustering and feature learning via structured sparsity., in: ICML 2013, 2013, pp. 352–360.

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012

Juan-Hui Li received her bachelor's degree in School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. She is currently a master student at Sun Yat-sen University. Her research interest is data clustering.

Chang-Dong Wang received the PhD degree in computer science in 2013 from the Sun Yat-sen University, China. He is currently an associate professor in the School of Data and Computer Science, Sun Yat-sen University. His current research interests include machine learning and pattern recognition, especially focusing on data clustering and its applications. He has published over 50 scientific papers in international journals and conferences such as IEEE TPAMI, IEEE TKDE, IEEE TSMC-C, Pattern Recognition, Knowledge and Information System, Neurocomputing, ICDM and SDM. His ICDM 2010 paper won the Honorable Mention for Best Research Paper Awards. He won 2012 Microsoft Research Fellowship Nomination Award. He was awarded 2015 Chinese Association for Artificial Intelligence (CAAI) Outstanding Dissertation.

Pei-Zhen Li received her bachelor's degree in School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. She is currently a master student at Sun Yat-sen University. Her research interest is data clustering.

Jian-Huang Lai received his M.Sc. degree in applied mathematics in 1989 and his Ph.D. in mathematics in 1999 from Sun Yat-sen University, China. He joined Sun Yat-sen University in 1989 as an Assistant Professor, where currently, he is a Professor with School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His current research interests are in the areas of digital image processing, pattern recognition, multimedia communication, wavelet and its applications. He has published over 200 scientific papers in the international journals and conferences on image processing and pattern recognition, e.g. IEEE TPAMI, IEEE TKDE, IEEE TNN, IEEE TIP, IEEE TSMC (Part B), Pattern Recognition, ICCV, CVPR and ICDM. Prof. Lai serves as a standing member of the Image and Graphics Association of China and also serves as a standing director of the Image and Graphics Association of Guangdong.

Please cite this article as: J.-H. Li et al., Discriminative metric learning for multi-view graph partitioning, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.06.012