

Article

# Cancer Subtype Recognition Based on Laplacian Rank Constrained Multiview Clustering

Shuguang Ge, Xuesong Wang , Yuhu Cheng and Jian Liu \*

School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China; gesgcumt17@163.com (S.G.); wangxuesongcumt@163.com (X.W.); chengyuhu@163.com (Y.C.)

\* Correspondence: tbh166@cumt.edu.cn

**Abstract:** Integrating multigenomic data to recognize cancer subtype is an important task in bioinformatics. In recent years, some multiview clustering algorithms have been proposed and applied to identify cancer subtype. However, these clustering algorithms ignore that each data contributes differently to the clustering results during the fusion process, and they require additional clustering steps to generate the final labels. In this paper, a new one-step method for cancer subtype recognition based on graph learning framework is designed, called Laplacian Rank Constrained Multiview Clustering (LRCMC). LRCMC first forms a graph for a single biological data to reveal the relationship between data points and uses affinity matrix to encode the graph structure. Then, it adds weights to measure the contribution of each graph and finally merges these individual graphs into a consensus graph. In addition, LRCMC constructs the adaptive neighbors to adjust the similarity of sample points, and it uses the rank constraint on the Laplacian matrix to ensure that each graph structure has the same connected components. Experiments on several benchmark datasets and The Cancer Genome Atlas (TCGA) datasets have demonstrated the effectiveness of the proposed algorithm comparing to the state-of-the-art methods.



**Citation:** Ge, S.; Wang, X.; Cheng, Y.; Liu, J. Cancer Subtype Recognition Based on Laplacian Rank Constrained Multiview Clustering. *Genes* **2021**, *12*, 526. <https://doi.org/10.3390/genes12040526>

Academic Editor: Martin Vingron

Received: 3 February 2021

Accepted: 31 March 2021

Published: 3 April 2021

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



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** cancer subtype recognition; Laplacian Rank Constrained; multiview clustering; graph learning

## 1. Introduction

Tumor is a malignant heterogeneous disease caused by changes in cellular components at the levels of expression, epigenetics, transcription and proteomics. The heterogeneity will be reflected in that the same cancer will produce the subtypes with different phenotypes, which will affect the clinical treatment and prognosis [1,2]. With the development and maturity of new generation sequencing technologies, large amounts of biological data are collected in public databases that are easily accessible to researchers [3]. For example, The Cancer Genome Atlas (TCGA), a landmark cancer genomics project, stores information on biological processes such as mRNA expression data, DNA methylation data, miRNA expression data and mutation data for more than 30 cancers and thousands of cancer patients [4]. Therefore, in order to solve the problem of cancer subtype recognition, building a multiview clustering model that makes full use of biological information plays a significant role.

In order to implement the task of clustering, scholars initially focus on dimensionality reduction, matrix decomposition and linear regression technologies. They all use different strategies to project high-dimensional data into low-dimensional feature space, and then achieve clustering by k-means [5–10]. For example, an effective classical method, iCluster [5], builds a Gaussian latent variable model and its modified version, iClusterPlus [6], considers different variable types following different linear probabilistic relationships to build a regression model. Both of them achieve a low-dimensional space with the combination of different biological characteristics. The other method, Pattern Fusion Analysis (PFA) [10], first uses an improved Principal Component Analysis (PCA) to find out a

low-dimensional matrix of each sample, and then uses an adaptive alignment algorithm to build a fused low-dimensional feature space. However, these methods may further dilute the already low signal-to-noise ratio and increase the noise pollution to the results. Considering that the sample (patient) size of the biological data is much smaller than the feature (gene) size, some graph-based learning methods for cancer subtype recognition are designed [11–17]. These methods use the sample points to quickly construct the similarity graph, which can be converted into the problem of spectral clustering. For example, a widely mentioned algorithm, Similarity Network Fusion (SNF) [11], constructs the global and local similar networks for each data, and then integrates them into the final similar network based on the strategy of information propagation to dilute low similarity and enhance high similarity. Inspired by SNF, Ma et al. provided Affinity Network Fusion (ANF) [12], which constructs patients'  $k$ -nearest neighbor similar network for each data type, and then fuses these networks based on the random walk method. In addition, Yu et al. proposed Multiview Clustering using Manifold Optimization (MVCMO) [17], and solved the problem of spectral clustering optimization by using the line search method on Stiefel manifold space.

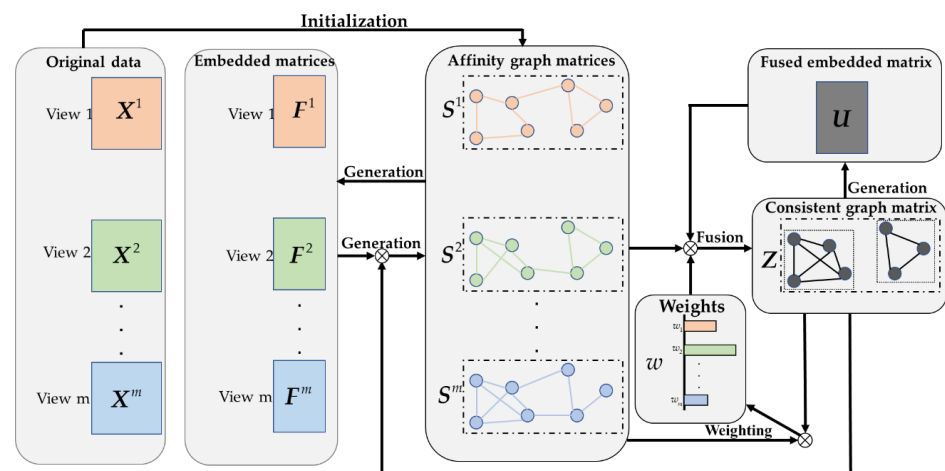
However, most existing graph-based multiview clustering methods separate the data clustering process from graph learning process [18,19]. In some methods, the construction of the graph is independent of the clustering task, resulting in its performance being highly dependent on the predefined graph. Recently, some adaptive graph learning methods using a rank constraint on the Laplacian matrix have been able to directly reveal the clustering structure, which makes the graph construction closely related to the clustering task [20–24]. In addition, the similarity between sample points may commonly behave differently in different views in the process of graph fusion. Some existing algorithms simply take the average of the affinity graph of multiple views to represent the result of the fusion graph [25,26]. Therefore, the rich heterogeneous information is not fully utilized.

To sum up, we designed a graph-based multiview clustering algorithm, called Laplacian Rank Constrained Multiview Clustering (LRCMC). Firstly, the Laplacian Rank Constraint (LRC) algorithm [27] is used to simultaneously find the affinity graph and the embedding matrix in each view to ensure that the graph structure is on the same connected components. Then, based on the method of Nie et al. [24], we use LRC method to obtain the consistent graph, whose connected components are the same as the affinity graph of each view. Finally, the clustering structure is obtained. In the process of graph fusion, the inverse distance weighting scheme is employed to design different weights for each view's affinity graph [24], so as to adjust the structure of the consistent graph more effectively. Moreover, the processes of graph learning, graph fusion and clustering are coupled into an optimization problem to update the more accurate consistent graph and improve the results of the clustering. In order to evaluate the effectiveness of the proposed method, experiments were carried out on four benchmark datasets and four TCGA datasets. Four state-of-the-art methods were used for comparison. The values of Accuracy (ACC), Normalized Mutual Information (NMI) and Purity on benchmark datasets, which are commonly used metrics in clustering analysis, and the  $p$  value obtained from survival analysis on the TCGA dataset can all show that the proposed LRCMC approach achieves considerable improvement over the state-of-the-art baseline methods. In the analysis of the Glioblastoma Multiforme (GBM) subtypes, we found these clusters have biological significance, e.g., the Proneural subtype granted by G-CIMP phenotype has a better survival advantage. The source code and datasets can be found in the Supplementary File 1.

## 2. Methods

The overall flow of LRCMC is shown in Figure 1. Specifically, given a set of omics data with  $m$  views  $X^1, \dots, X^m$ , a set of affinity graph matrices  $S^1, \dots, S^m$  are constructed, respectively, according to  $X^1, \dots, X^m$ . It should be emphasized that the process of learning the affinity matrix in LRCMC is different from most multiview clustering algorithms.  $S^1, \dots, S^m$  are not calculated directly from the original matrix, but are constructed, respec-

tively, from a set of embedding matrices  $F^1, \dots, F^m$  by the LRC method. Therefore, each affinity graph matrix is constrained to the same connected components, which ensures that each affinity graph has a similar structure before the fusion process. Then, the proposed fusion method is applied to the affinity graph matrices of all views in order to learn a consistent graph matrix  $Z$ . Simultaneously, each view is automatically assigned a different weight  $w_1, \dots, w_m$  to represent its contribution to  $Z$  during the fusion process. Finally, the learned consistent graph matrix  $Z$  is used to optimize the affinity graph matrix for each view. The LRC method is also imposed to constrain that the number of connected components in the  $Z$  is equal to the required number of clusters  $c$  by constructing the fusion embedded matrix  $U$ . Our LRCMC improves the affinity matrix of each view, builds a fused consistent graph matrix and obtains clustering results simultaneously.



**Figure 1.** The flow chart of Laplacian Rank Constrained Multiview Clustering.

### 2.1. Construction of Affinity Graph Based on LRC

Given a single biological data  $X^v = \{x_1^v, \dots, x_n^v\} \in R^{d_v \times n}$  denotes the  $v$ -th view data with  $d_v$  features, where  $n$  is the number of data points.  $S^v \in R^{n \times n}$  represents the similar relationship between the sample points in the graph learning framework. The smaller the distance between a pair of vertices in the graph is, the greater the similarity between the pair of vertices will be, the greater the corresponding weight will be, and vice versa. Based on the manifold structure of graph, the most traditional way to build  $S^v$  is by generating a  $k$ -nearest neighbor graph for it. A pair of vertices are considered connected if they are near neighbors. There are other effective strategies to design more accurate affinity graph  $S^v$ , e.g., smooth representation [28], Gaussian kernel for similarity learning [29], etc. For the purpose of clustering, if the sample points can be assigned to the  $c$  categories, the obtained  $S^v$  should contain exact  $c$  connected components. Based on the following Theorem 1,  $S^v$  can be realized.

**Theorem 1.** Multiplicity  $c$  of zero eigenvalues of Laplacian matrix  $L^v$  is equal to the number of connected components of its similarity matrix  $S^v$ .

When all the elements in  $S^v$  satisfy the non-negative condition, its Laplacian matrix  $L^v$  has the above property [30,31].

Theorem 1 means if  $\sum_{i=1}^c \lambda_i = 0$ , where  $\lambda_i$  is  $i$ -th smallest eigenvalue of  $L^v$ , the data points on  $S^v$  have been ideally assigned to  $c$  categories [32], Laplacian rank meets the constraint condition  $\text{rank}(L^v) = n - c$ . Therefore, based on the Ky Fan's theorem [33],

we can minimize  $\sum_{i=1}^c \lambda_i$  approximately meeting the requirement of Theorem 1. The objective function is written as:

$$\sum_{i=1}^c \lambda_i = \min_{\mathbf{F}^v \in R^{n \times c}, (\mathbf{F}^v)^T \mathbf{F}^v = \mathbf{I}} \text{Tr} \left( (\mathbf{F}^v)^T \mathbf{L}^v \mathbf{F}^v \right) \quad (1)$$

where  $\mathbf{F}^v \in R^{n \times c}$  is obtained by the  $c$  eigenvectors of  $\mathbf{L}^v$  corresponding to the  $c$  smallest eigenvalues.  $\text{Tr}(\cdot)$  denotes the trace operator,  $\mathbf{L}^v = \mathbf{D}^v - \left[ \left( (\mathbf{S}^v)^T + \mathbf{S}^v \right) / 2 \right]$  is the Laplacian matrix,  $\mathbf{D}^v$  is a diagonal matrix and its elements are column sums of  $\left[ \left( (\mathbf{S}^v)^T + \mathbf{S}^v \right) / 2 \right]$ .

However, the solution to  $\mathbf{F}^v$  in Equation (1) is actually to solve trivial solution to  $\mathbf{S}^v$ . Therefore, a  $\ell_2$ -norm regularization term is employed to obtain smooth  $\mathbf{S}^v$  and each column of  $\mathbf{S}^v$  satisfies  $\mathbf{1}^T \mathbf{s}_j^v = 1$ , where  $\mathbf{s}_j^v$  is the  $j$ th column of  $\mathbf{S}^v$  [21]. Finally, we can obtain the objective function related to  $\mathbf{F}^v$  and  $\mathbf{S}^v$  simultaneously:

$$\begin{aligned} \min_{\mathbf{F}^v, \mathbf{S}^v} & 2\text{Tr} \left( (\mathbf{F}^v)^T \mathbf{L}^v \mathbf{F}^v \right) + \alpha \|\mathbf{S}^v\|_{\text{F}}^2 \\ \text{s.t.} & (\mathbf{F}^v)^T \mathbf{F}^v = \mathbf{I}, \forall j, \mathbf{1}^T \mathbf{s}_j^v = 1, \mathbf{s}_j^v \geq 0, \mathbf{s}_{jj}^v = 0 \end{aligned} \quad (2)$$

where  $\alpha$  is the regularization parameter.

A set of the affinity graph matrices  $\mathbf{S}^1, \dots, \mathbf{S}^m$  and the embedded matrices  $\mathbf{F}^1, \dots, \mathbf{F}^m$  are obtained through Equation (2) without the participation of the original data. However, these affinity matrices are unrelated; if they are simply stacked together for clustering, the graphs will be badly damaged and the algorithm performance will degrade. Therefore, we need to introduce a graph fusion strategy to construct a consistent graph matrix with the unified connected components.

## 2.2. Graph Fusion with LRC

Integrating these basic graphs to form the fused affinity graph  $\mathbf{Z} \in R^{n \times n}$ , two intuitive points should be considered: (1) The designed graph  $\mathbf{S}^v$  for each view can be considered as the consistent graph  $\mathbf{Z}$  with noise representation and outlier interference. (2)  $\mathbf{S}^v$  closer to  $\mathbf{Z}$  should be given greater weight to reduce the perturbation of the low-quality graphs on the fusion graph. In this way,  $\mathbf{Z}$  can accurately capture the true similarity hidden in the multiview data. Therefore, we employed the proposed method of Nie et al. [24] to optimize  $\mathbf{Z}$  as follows:

$$\min_{\mathbf{Z}} \sum_{v=1}^m w_v \|\mathbf{Z} - \mathbf{S}^v\|_{\text{F}}^2 \quad (3)$$

where  $w_v$  is the weight of the single affinity graph  $\mathbf{S}^v$ . The inverse distance weighting scheme is designed to calculate  $w_v$ . The Lagrange function of Equation (3) can be written as:

$$\min_{\mathbf{Z}} \sum_{v=1}^m w_v \|\mathbf{Z} - \mathbf{S}^v\|_{\text{F}}^2 + \zeta(\mathbf{\Lambda}, \mathbf{Z}) \quad (4)$$

where  $\mathbf{\Lambda}$  is the Lagrange multiplier,  $\zeta(\mathbf{\Lambda}, \mathbf{Z})$  is the formal term derived from constraint condition. Taking the derivative of Equation (4) w.r.t  $\mathbf{Z}$  and setting the derivative to zero, we can obtain:

$$\sum_{v=1}^m w_v \frac{\partial \|\mathbf{Z} - \mathbf{S}^v\|_{\text{F}}^2}{\partial \mathbf{Z}} + \frac{\partial \zeta(\mathbf{\Lambda}, \mathbf{Z})}{\partial \mathbf{Z}} = 0 \quad (5)$$

where  $w_v$  is given as follows:

$$w_v = \frac{1}{2\sqrt{\|\mathbf{Z} - \mathbf{S}^v\|_{\text{F}}^2}} \quad (6)$$

Here, a set of weights  $w_1, \dots, w_m$  and a consistent graph matrix  $\mathbf{Z}$  are obtained from Equation (3). In order to make the learned  $\mathbf{Z}$  also have  $c$  connected components for

clustering, the LRC term is added to Equation (3) according to Theorem 1 and Ky Fan's Theorem. The objective function is as follows:

$$\begin{aligned} \min_{\mathbf{Z}} \sum_{v=1}^m w_v \|\mathbf{Z} - \mathbf{S}^v\|_F^2 + 2\beta \text{Tr}(\mathbf{U}^T \mathbf{L}_Z \mathbf{U}) \\ \text{s.t. } \mathbf{U}^T \mathbf{U} = \mathbf{I}, \forall j, \mathbf{1}^T \mathbf{z}_j = 1, \mathbf{z}_j \geq 0, z_{jj} = 0 \end{aligned} \quad (7)$$

where  $\mathbf{U} \in R^{n \times c}$  is obtained by the  $c$  eigenvectors of  $\mathbf{L}_Z$  corresponding to the  $c$  smallest eigenvalues, i.e., the embedded matrix corresponding to  $\mathbf{Z}$ .  $\mathbf{L}_Z = \mathbf{D}_Z - [(\mathbf{Z}^T + \mathbf{Z})/2]$  is the Laplacian matrix,  $\mathbf{D}_Z$  is a diagonal matrix and its elements are column sums of  $[(\mathbf{Z}^T + \mathbf{Z})/2]$ .  $\beta$  is the regularization parameter.

### 2.3. LRCMC Algorithm

As described in Sections 2.1 and 2.2, the LRC operation is used to guarantee the structures of  $\mathbf{S}^1, \dots, \mathbf{S}^m$  and  $\mathbf{Z}$ . Therefore, we can combine Equations (2) and (7) into a final objective function, i.e., the proposed Laplacian Rank Constrained Multiview Clustering (LRCMC). It is represented as:

$$\begin{aligned} \min_{\mathbf{F}^v, \mathbf{S}^v, w_v, \mathbf{Z}, \mathbf{U}_{v=1}^m} \sum_{v=1}^m 2\text{Tr}((\mathbf{F}^v)^T \mathbf{L}^v \mathbf{F}^v) + \alpha \|\mathbf{S}^v\|_F^2 + w_v \|\mathbf{Z} - \mathbf{S}^v\|_F^2 + 2\beta \text{Tr}(\mathbf{U}^T \mathbf{L}_Z \mathbf{U}) \\ \text{s.t. } (\mathbf{F}^v)^T \mathbf{F}^v = \mathbf{I}, \mathbf{U}^T \mathbf{U} = \mathbf{I}, \\ \forall j, \mathbf{1}^T \mathbf{s}_j^v = 1, \mathbf{s}_j^v \geq 0, s_{jj}^v = 0, \mathbf{1}^T \mathbf{z}_j = 1, \mathbf{z}_j \geq 0, z_{jj} = 0 \end{aligned} \quad (8)$$

Here, we complete the tasks of graph construction, graph fusion and clustering in one step through the integrated model. In this way, the learning of  $\mathbf{S}^1, \dots, \mathbf{S}^m$  and  $\mathbf{Z}$  can help each other embedded in a joint coupling problem. The objective function Equation (8) enjoys the following properties:

- Our method can effectively learn a set of affinity graph matrices with  $c$  connected components, instead of most multiview clustering methods requiring predefined graphs;
- In the graph fusion process, we assign the weight to each view to represent their contribution to the consistent graph  $\mathbf{Z}$ , rather than simply superimposing them together;
- We use LRC to constantly adjust the structures of  $\mathbf{S}^1, \dots, \mathbf{S}^m$  and  $\mathbf{Z}$ , and at the same time complete the task of clustering.

### 2.4. Optimization Algorithm of LRCMC

Obviously, since the variables in Equation (8) are coupled to each other, we use alternating iterative method and Augmented Lagrange Multiplier (ALM) scheme to solve  $\mathbf{S}^1, \dots, \mathbf{S}^m, \mathbf{F}^1, \dots, \mathbf{F}^m, w_1, \dots, w_m, \mathbf{Z}, \mathbf{U}$ . The specific solution process is as follows:

1. Fix  $\mathbf{F}^1, \dots, \mathbf{F}^m, w_1, \dots, w_m, \mathbf{Z}$  and  $\mathbf{U}$ , solve  $\mathbf{S}^1, \dots, \mathbf{S}^m$ ;

The Equation (8) becomes:

$$\begin{aligned} \min_{\mathbf{S}^v} \sum_{v=1}^m 2\text{Tr}((\mathbf{F}^v)^T \mathbf{L}^v \mathbf{F}^v) + \alpha \|\mathbf{S}^v\|_F^2 + w_v \|\mathbf{Z} - \mathbf{S}^v\|_F^2 \\ \text{s.t. } \forall j, \mathbf{1}^T \mathbf{s}_j^v = 1, \mathbf{s}_j^v \geq 0, s_{jj}^v = 0 \end{aligned} \quad (9)$$

Due to  $\text{Tr}((\mathbf{F}^v)^T \mathbf{L}^v \mathbf{F}^v) = \frac{1}{2} \sum_{i,j} \|\mathbf{f}_i^v - \mathbf{f}_j^v\|_2^2 s_{ij}^v$ , where  $\mathbf{f}_i^v$  and  $\mathbf{f}_j^v$  denote the  $i$ -th and  $j$ -th column of  $\mathbf{F}^v$ , respectively,  $s_{ij}^v$  denotes the  $(i, j)$ th element of  $\mathbf{S}^v$ , the Equation (9) can be written in vector form as:

$$\begin{aligned} \min_{\mathbf{s}_j^v} \sum_{i=1}^n \|\mathbf{f}_i^v - \mathbf{f}_j^v\|_2^2 s_{ij}^v + \alpha \|\mathbf{s}_j^v\|_2^2 + w_v \|\mathbf{z}_j - \mathbf{s}_j^v\|_2^2 \\ \text{s.t. } s_{jj}^v = 0, \mathbf{s}_j^v \geq 0, \mathbf{1}^T \mathbf{s}_j^v = 1 \end{aligned} \quad (10)$$

Denote  $p_{ij}^v = \|f_i^v - f_j^v\|_2^2$ , Equation (10) is obviously written as:

$$\min_{s_j^v} \sum_{i=1}^n p_{ij}^v s_{ij}^v + \alpha \|s_{ij}^v\|_2^2 + w_v \|z_{ij} - s_{ij}^v\|_2^2 \quad (11)$$

$$s.t. s_{jj}^v = 0, s_j^v \geq 0, 1^T s_j^v = 1$$

Then, the Equation (11) can be written as:

$$\min_{s_j^v} \left\| s_j^v + \frac{p_j^v/2 - z_j}{\alpha + w_v} \right\|_2^2 \quad (12)$$

Therefore, the Lagrangian function of Equation (12) combined with its constraints can be defined as:

$$\ell(s_j^v, \eta, \varphi) = \frac{1}{2} \left\| s_j^v + \frac{p_j^v/2 - z_j}{\alpha + w_v} \right\|_2^2 + \eta (1^T s_j^v - 1) + \varphi^T s_j^v \quad (13)$$

where  $\eta$  is the Lagrangian coefficient scalar and  $\varphi$  is the Lagrangian coefficient vector. Based on the Karush-Kuhn-Tucker (KKT) condition [34], the optimal solution of  $s_j^v$  can be estimated as:

$$s_j^v = \left( \frac{p_j^v/2 - z_j}{\alpha + w_v} + \eta \right)_+ \quad (14)$$

The study in [35] found that sparse representation is robust to noise and outliers. In order to obtain the sparse affinity graph  $S^v$ , we can find the  $k$  nonzero adaptive neighbors for  $s_j^v$  to satisfy  $s_{jk}^v > 0$  and  $s_{j,k+1}^v = 0$ . Denote  $\alpha + w_v = \delta$ , then, we arrive at:

$$-\frac{p_{jk}^v}{2} + w_v z_{jk} + \delta \eta > 0, -\frac{p_{j,k+1}^v}{2} + w_v z_{j,k+1} + \delta \eta \leq 0 \quad (15)$$

Moreover, according to Equation (15) and the constraint condition  $1^T s_j^v - 1 = 0$ ,  $\eta$  is given:

$$\eta = \frac{1}{k} \left( 1 + \frac{w_v}{\delta} + \frac{\sum_{l=1}^k p_{jl}^v}{2\delta} \right) \quad (16)$$

Therefore, according to Equations (15) and (16), the range of  $\delta$  is obtained as follows:

$$\begin{cases} \delta > \frac{kp_{jk}^v - 2kw_v z_{jk} - 2w_v - \sum_{l=1}^k p_{jl}^v}{2} \\ \delta \leq \frac{kp_{j,k+1}^v - 2kw_v z_{j,k+1} - 2w_v - \sum_{l=1}^k p_{jl}^v}{2} \end{cases} \quad (17)$$

Then, the parameter  $\delta$  can be set as:

$$\delta = \frac{kp_{j,k+1}^v - 2kw_v z_{j,k+1} - 2w_v - \sum_{l=1}^k p_{jl}^v}{2} \quad (18)$$

Finally, according to Equations (15), (16) and (18), the optimal solution of  $s_j^v$  in  $s_{ij}^v$  is represented as:

$$s_{ij}^v = \begin{cases} \frac{p_{j,k+1}^v - p_{ij}^v + 2w_v z_{ij} - 2w_v z_{j,k+1}}{kp_{j,k+1}^v - \sum_{l=1}^k p_{jl}^v - 2kw_v z_{j,k+1} + 2\sum_{l=1}^k w_v z_{jl}} & j \leq k \\ 0 & j > k \end{cases} \quad (19)$$



2. Fix  $S^1, \dots, S^m, w_1, \dots, w_m, Z$  and  $U$ , solve  $F^1, \dots, F^m$ ;

Updating  $F^v$  in Equation (8) is converted to Equation (2). Therefore,  $F^v$  is updated from Equation (1) in Section 2.1.

3. Fix  $S^1, \dots, S^m, F^1, \dots, F^m, Z$  and  $U$ , solve  $w_1, \dots, w_m$ ;

Updating  $w_v$  in Equation (8) is equivalent to Equation (3). Therefore,  $w_v$  is updated from Equation (6) in Section 2.2.

4. Fix  $S^1, \dots, S^m, F^1, \dots, F^m, w_1, \dots, w_m$  and  $U$ , solve  $Z$ ;

Updating  $Z$  in Equation (8) is converted to Equation (7). Due to  $\text{Tr}(U^T L_Z U) = \frac{1}{2} \sum_{i,j} \|u_i - u_j\|_2^2 z_{ij}$ , where  $u_i$  and  $u_j$  denote the  $i$ -th and  $j$ -th column of  $U$ ,  $z_{ij}$  denotes the  $(i, j)$ th element of  $Z$ , Equation (7) yields:

$$\min_{z_j} \sum_{v=1}^m \sum_i^n \left\{ w_v \|z_j - s_j^v\|_2^2 + \beta \|u_i - u_j\|_2^2 z_{ij} \right\} \quad (20)$$

$$\text{s.t. } \mathbf{1}^T z_j = 1, z_j \geq 0, z_{jj} = 0$$

Denote  $q_{ij} = \|u_i - u_j\|_2^2$ , we have:

$$\min_{z_j} \sum_{v=1}^m \|z_j - s_j^v + \frac{\beta}{2mw_v} q_j\|_2^2 \quad (21)$$

$$\text{s.t. } \mathbf{1}^T z_j = 1, z_j \geq 0, z_{jj} = 0$$

Based on the Karush–Kuhn–Tucker (KKT) condition [34], the closed form solution of  $z_j$  can be estimated as:

$$z_j = \left( s_j^v - \frac{\beta}{2mw_v} q_j + \eta \right)_+ \quad (22)$$

Equation (22) can be solved by an efficient optimization method proposed in [35].

5. Fix  $S^1, \dots, S^m, F^1, \dots, F^m, w_1, \dots, w_m$  and  $Z$ , solve  $U$ .

According to the method of finding  $F^v$ ,  $U$  is obtained as follows:

$$\min_{U \in R^{n \times c}, U^T U = I} \text{Tr}(U^T L_Z U) \quad (23)$$

The final solution of  $U$  is the  $c$  eigenvectors of  $L_Z$  corresponding to the  $c$  smallest eigenvalues.

### 3. Experiments' Results

In order to verify effectiveness of LRCMC in cancer subtype recognition, LRCMC was compared with four state-of-the-art clustering algorithms, i.e., ANF [12], SNF [11], PFA [10] and MVCMO [17]. Since biological omics data are not labeled, we first downloaded four widely used benchmark datasets containing real labels, i.e., 3-source, Calt-7, MSRC, WebKB, to verify that proposed LRCMC can achieve good clustering effect. Furthermore, we applied LRCMC to the datasets downloaded and preprocessed by Wang et al. [11] from TCGA. The datasets contain four types of cancer, i.e., GBM, Breast Invasive Carcinoma (BIC), Lung Squamous Cell Carcinoma (LSCC) and Colon Adenocarcinoma (COAD).

#### 3.1. Comparison Experiments on Benchmark Datasets

The benchmark datasets are described as follows:

- 3-source [20]: It contains 169 news that were reported by three news magazines, i.e., BBC, Reuters, and The Guardian. There are six different thematic labels for each news;
- Calt-7 [36]: The object recognition dataset is drawn from the Caltech101 dataset to screen 7 widely used classes, i.e., faces, motorbikes, dollar bill, Garfield, stop sign, and

- Windsor chair. Each class has 1474 images. Each image is described by 6 features, i.e., GABOR, wavelet moment (WM), CENT, HOG, GIST and LBP;
- MSRC [37]: The scene recognition dataset contains 7 classes of aircraft, car, bicycle, cow, faces, tree, and building. Each image is described by 5 features, i.e., color moment (CMT), HOG, LBP, CENT, GIST;
  - WebKB [20]: It collects 203 web pages in 4 classes from the University's Computer science department. Each page has 3 features, i.e., the content of the page, the anchor text of the hyperlink, and the text description in the title.

Table 1 is an overview of these datasets, where  $n$ ,  $m$ , and  $c$  describe the number of samples, views, and classes for each dataset, respectively,  $d_v$  denotes the  $i$ -th feature of these datasets.

**Table 1.** Overview of four benchmark datasets.

Dataset	$n$	$m$	$c$	$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$d_6$
3-source	169	3	6	3560	3631	3638	-	-	-
Calt-7	1474	6	7	48	40	254	1984	512	928
MSRC	210	5	7	48	100	256	1302	512	-
WebKB	203	3	4	1703	230	230	-	-	-

Three commonly used evaluation metrics, i.e., Accuracy (ACC), Normalized Mutual Information (NMI) and Purity, are used to quantitatively measure the clustering performance of the algorithms. The metrics compare the resulting labels with the real labels provided by the dataset. The larger the value obtained, the better the clustering results. To ensure the fairness of the comparison experiments, each algorithm was run 10 times to reduce the impact of randomness. The mean and standard deviation of the obtained metrics were calculated. In addition, the neighbor  $k$  required by ANF, SNF, MVCMO and LRCMC was set within the range of [5, 50], and other parameters were specified as the default values provided by the authors. Only one parameter  $\beta$  needs to be set in our LRCMC algorithm, which is caused by the introduction of LRC. In order to achieve rapid convergence of Algorithm 1, we adopt a dynamic parameter updating method proposed by Nie et al. [23].  $\beta$  is set in the range of [1, 30]. If the number of connected components of  $\mathbf{Z}$  is greater than  $c$ , we will shrink  $\beta$  ( $\beta = \beta/2$ ). On the contrary, if less than  $c$ , we will increase  $\beta$  ( $\beta = 2 \times \beta$ ) until finding the right components for  $\mathbf{Z}$ . Table 2 shows the final evaluation metrics obtained by these algorithms in the four datasets. It is obvious that LRCMC achieves better clustering performance in the multiview clustering task than the other methods.

**Algorithm 1.** LRCMC algorithm

**Input:** Original data  $X^1, \dots, X^m$  with  $m$  views, the number of clusters  $c$ , the number of neighbors  $k$ , the regularization parameter  $\beta$ .

**Output:** The learned consensus matrix  $\mathbf{Z}$ .

Initialize the affinity matrices  $S^1, \dots, S^m$  for each view by solving the following problem:

$$\min_{S_j^v} \sum_{i=1}^n \|x_i - x_j\|_2^2 s_{ij}^v + \alpha \|s_j^v\|_2^2;$$

Initialize the embedded matrices  $F^1, \dots, F^m$  for each view by using Equation (1);

Initialize the weights  $w_1, \dots, w_m$  for each view by  $w_v = 1/m$ ;

Initialize  $\mathbf{Z}$  by connecting  $S^1, \dots, S^m$  with  $w_1, \dots, w_m$ ;

Initialize the fused embedded matrix  $\mathbf{U}$  by using Equation (23);

**Repeat**

Fix  $F^1, \dots, F^m, w_1, \dots, w_m, \mathbf{Z}$  and  $\mathbf{U}$ , update  $S^1, \dots, S^m$  by using Equation (19);

Fix  $S^1, \dots, S^m, w_1, \dots, w_m, \mathbf{Z}$  and  $\mathbf{U}$ , update  $F^1, \dots, F^m$  by using Equation (1);

Fix  $S^1, \dots, S^m, F^1, \dots, F^m, \mathbf{Z}$  and  $\mathbf{U}$ , update  $w_1, \dots, w_m$  by using Equation (6);

Fix  $S^1, \dots, S^m, F^1, \dots, F^m, w_1, \dots, w_m$  and  $\mathbf{U}$ , update  $\mathbf{Z}$  by using Equation (22);

Fix  $S^1, \dots, S^m, F^1, \dots, F^m, w_1, \dots, w_m$  and  $\mathbf{Z}$ , update  $\mathbf{U}$  by using Equation (23);

**Until** Satisfy Theorem 1 or the maximum iteration reached.

The learned consensus matrix  $\mathbf{Z}$  with exact  $c$  connected components, which are the final clusters.



**Table 2.** The clustering performance comparison in terms of ACC, NMI and Purity on the four real datasets.

Datasets	Methods	ACC	NMI	Purity
3-source	ANF	0.4970 (0.0000)	0.2804 (0.0000)	0.5325 (0.0000)
	SNF	0.7811 (0.0000)	0.6942 (0.0000)	0.8166 (0.0000)
	PFA	0.4562 (0.0761)	0.2247 (0.0713)	0.7160 (0.0578)
	MVCMO	0.4221 (0.0123)	0.3035 (0.0128)	0.5266 (0.0118)
	LRCMC	<b>0.8107 (0.0000)</b>	<b>0.7218 (0.0000)</b>	<b>0.8462 (0.0000)</b>
Calt-7	ANF	0.6696 (0.0000)	0.6203 (0.0000)	0.8684 (0.0000)
	SNF	0.6601 (0.0000)	0.5637 (0.0000)	0.8562 (0.0000)
	PFA	-	-	-
	MVCMO	0.6654 (0.0100)	0.5179 (0.0355)	0.8464 (0.0083)
	LRCMC	<b>0.8548 (0.0000)</b>	<b>0.7694 (0.0000)</b>	<b>0.8921 (0.0000)</b>
MSRC	ANF	0.8048 (0.0000)	0.7297 (0.0000)	0.8143 (0.0000)
	SNF	0.8429 (0.0000)	0.7514 (0.0000)	0.8429 (0.0000)
	PFA	-	-	-
	MVSCO	0.7800 (0.0544)	0.6711 (0.0628)	0.7838 (0.0462)
	LRCMC	<b>0.8905 (0.0000)</b>	<b>0.7922 (0.0000)</b>	<b>0.8905 (0.0000)</b>
WebKB	ANF	0.6798 (0.0000)	0.1718 (0.0000)	0.6946 (0.0000)
	SNF	0.7044 (0.0000)	0.2407 (0.0000)	0.7192 (0.0000)
	PFA	0.7143 (0.0000)	0.3191 (0.0000)	0.8128 (0.0000)
	MVCMO	0.7652 (0.0346)	0.3548 (0.0448)	0.7833 (0.0323)
	LRCMC	<b>0.8079 (0.0000)</b>	<b>0.5081 (0.0000)</b>	<b>0.8424 (0.0000)</b>

- means that the metrics cannot be calculated, the best results have been highlighted in bold.

### 3.2. Comparison Experiments on TCGA Datasets

To demonstrate the effectiveness of LRCMC in identifying cancer subtype, the designed LRCMC was applied to four cancer omics datasets, i.e., GBM, BIC, LSCC, and COAD. Each cancer subtype contains three types of expression data from different platforms, i.e., mRNA expression data, DNA methylation data and miRNA expression data. Table 3 shows the number of samples (patients) and features (genes) held by each cancer subtype.

**Table 3.** Overview of the TCGA datasets.

Datasets	N	mRNA Expression	DNA Methylation	miRNA Expression
GBM	213	12,042	1305	534
BIC	105	17,814	23,094	354
LSCC	106	12,042	23,074	352
COAD	92	17,814	23,088	312

To ensure that the identified cancer labels conform to the true clinical diagnosis, we specified that the number of samples in each cluster should be at least 3. We used the number of subtypes of GBM, BIC, LSCC, and COAD specified by Wang et al., which were 3, 5, 4 and 3, respectively. Then, the  $p$  values based on Cox log-rank model were used to evaluate the clustering results of these algorithms in survival analysis [38]. If the  $p$  value is smaller, the survival rate between different groups is more significant and the difference is greater, which means the cluster is considered to have different characteristics of the underlying cancer subtypes. Cancer survival curves can also represent heterogeneity between different subtypes. As shown in Table 4, LRCMC obtained the best  $p$  value in BIC, GBM, KRCCC and COAD. Other algorithms also had good results in specific datasets, but they were all lower than our algorithm. Therefore, we believe that LRCMC is significantly advantageous in the topic of cancer subtype recognition. Figure 2 shows the Kaplan–Meier survival analysis curves of the four cancers. Each curve depicts trends in

the survival time of each cancer cluster and the number of samples for each cluster is also shown in the figure.

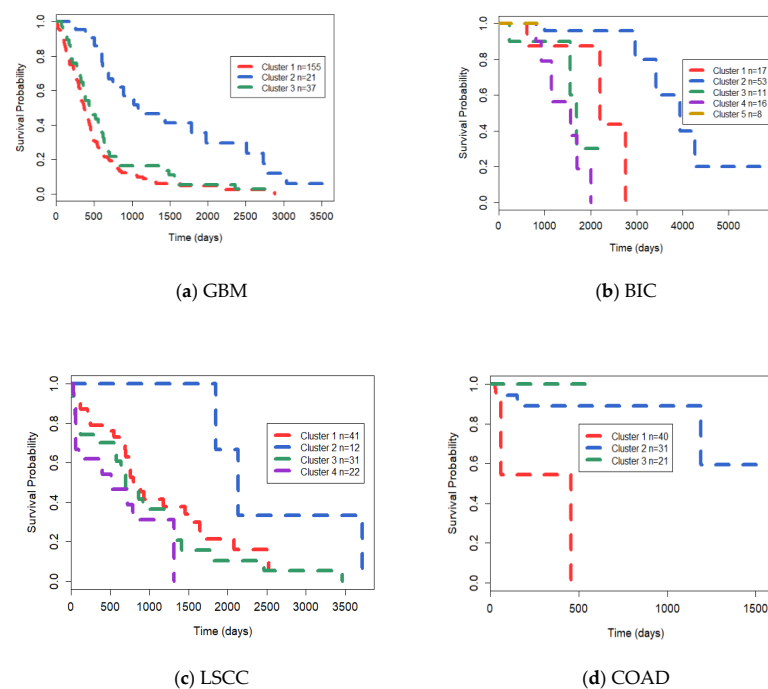
**Table 4.** *p* values of survival analysis in Cox log-rank model for different clustering methods of four cancers on The Cancer Genome Atlas (TCGA) datasets.

Methods	GBM	BIC	LSCC	COAD
ANF	$5.8 \times 10^{-4}$	$3.6 \times 10^{-4}$	$8.9 \times 10^{-3}$	$9.0 \times 10^{-3}$
SNF	$5.0 \times 10^{-5}$	$6.9 \times 10^{-4}$	$7.8 \times 10^{-3}$	$1.6 \times 10^{-3}$
PFA	$1.8 \times 10^{-4}$	$3.1 \times 10^{-4}$	$1.1 \times 10^{-2}$	$2.4 \times 10^{-2}$
MVCMO	$1.4 \times 10^{-3}$	$3.5 \times 10^{-4}$	$9.1 \times 10^{-3}$	$8.5 \times 10^{-3}$
LRCMC	<b><math>1.3 \times 10^{-5}</math></b>	<b><math>3.7 \times 10^{-5}</math></b>	<b><math>3.8 \times 10^{-3}</math></b>	<b><math>1.2 \times 10^{-3}</math></b>

The best results have been highlighted in bold.

### 3.3. Analysis on GBM Dataset

GBM is the most malignant glioma among astrocytomas. It has been studied and analyzed at the genetic level by many scholars, and specific subtypes and treatment protocols have been proposed. For example, according to the mRNA expression data, Verhaak et al. [39] reported that GBM is divided into Mesenchymal, Classical, Neural and Proneural subtypes, and the heterogeneous subtypes were also verified in somatic mutations and copy number variations (CNVs). Another study divided GBM patients into two subtypes, i.e., G-CLMP and non-G-CLMP, based on the difference of CpG Island methylator phenotype (CLMP) [40]. Table 5 shows the distribution of the cluster results obtained by LRCMC on the subtype identified by these two studies. From Table 5, there are more patients in cluster 1 than in cluster 3, and all of them are assigned to non-G-CLMP subtype—also they have four subtypes identified based on mRNA expression. The point is that the Proneural subtype in these two clusters belong to non-G-CLMP subtype. However, cluster 2, with a smaller number of patients, is almost the Proneural subtype, and also belongs to G-CLMP subtype.



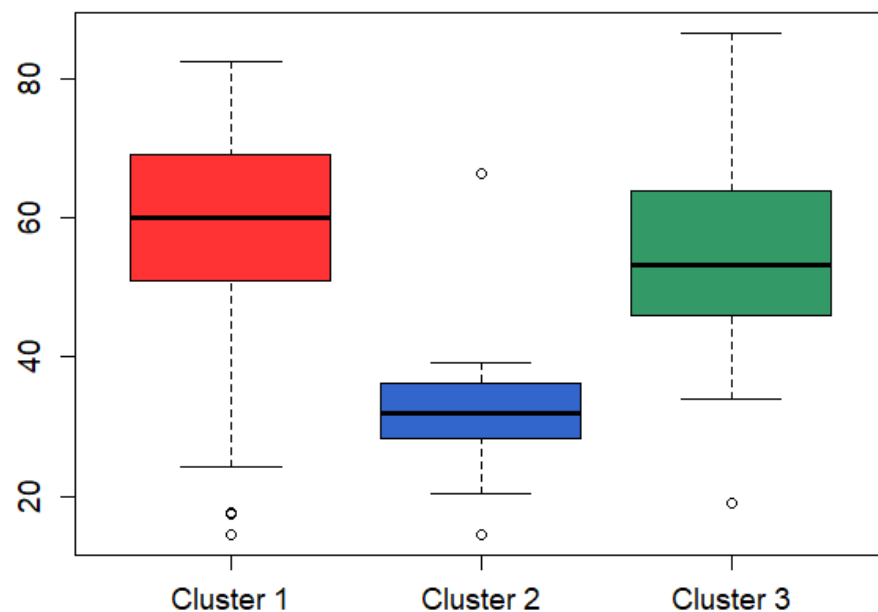
**Figure 2.** The Kaplan–Meier survival curves of (a): Glioblastoma Multiforme (GBM), (b): Breast Invasive Carcinoma (BIC), (c): Lung Squamous Cell Carcinoma (LSCC) and (d): Colon Adenocarcinoma (COAD), respectively.

**Table 5.** The identified clusters are compared with mRNA-expression-based subtypes and methylation-based subtypes.

Our Cluster	mRNA-Expression-Based Subtypes			Methylation-Based Subtypes	
	Mesenchymal Classical	Neural	Proneural	G-CLMP	Non-G-CLMP
cluster 1	46	54	27	0	155
cluster 2	1	0	1	19	20
cluster 3	12	11	7	7	0

The values represent the number of patients counted.

To further analyze the identified clusters, we downloaded clinical data, somatic mutation data and CNV data for all patients from the cBio Cancer Genomics Portal database (<http://www.cbioportal.org/> accessed on 15 December 2020). The age profiles of the three clusters (Figure 3), differential gene statistics of CNVs and mutations (Table 6), and Kaplan–Meier survival curves of Temozolomide (TMZ) (Figure 4) in GBM patients were obtained. Figure 3 shows that the diagnosis age of patients in cluster 2 with the best survival advantage is also lower than that of patients in cluster 1 and cluster 3. The genetic variant signatures associated with GBM in terms of mutation (*IDH1*) and CNVs (*CDKN2A*, *CDKN2B*, *C9orf53*, *MTAP*, *EGFR*) are significantly different in the three identified clusters. In particular, *IDH1* mutation only occurs in cluster 2, while *EGFR* amplification is 0. Then, we divided the patients within the three clusters into two groups: patients treated with TMZ and those not treated with TMZ, then we compared the drug response. TMZ is a drug that is commonly used to treat GBM, but only responds well to a subset of patients. The *p* values of survival analysis in Cox log-rank model of the three cluster comparison experiments are  $2.0 \times 10^{-6}$ , 0.76 and 0.01, respectively, which indicates that TMZ treatment has no effect on the patients in cluster 2. Therefore, in summary, we can infer that the subtype belonging to G-CLMP subtype and Proneural subtype might be a potentially new subtype. This also verified by the fact that that the Proneural subtype granted by the G-CIMP phenotype proposed by Canmenron et al. has unique properties [41].



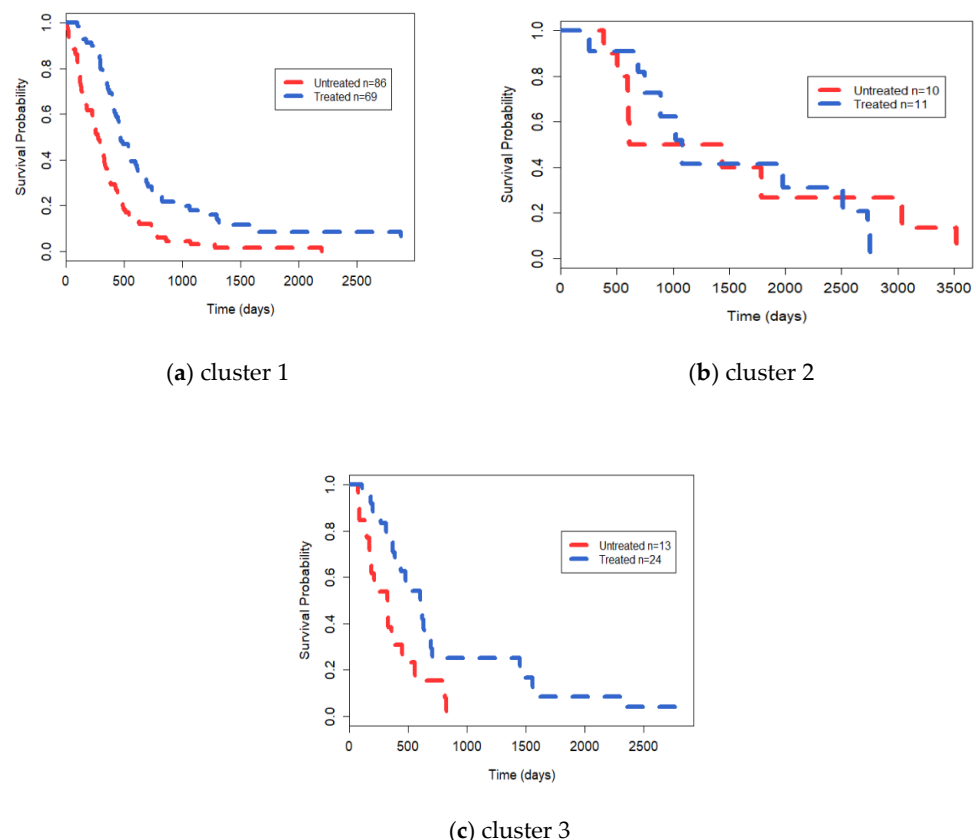
**Figure 3.** Boxplot of diagnosis age for the identified clusters. It reflects the distribution of diagnosis age in each cluster. Black bar represents the median of each cluster.

**Table 6.** Distribution of genetic variant signatures for the identified clusters.

Our Cluster	CDKN2A.del.	CDKN2B.del.	C9orf53.del.	MTAP.del.	EGFR.ampl.	IDH1
cluster 1	84 (56.4%)	84 (56.4%)	80 (53.7%)	57 (38.3%)	70 (47.0%)	0 (0%)
cluster 2	6 (28.6%)	6 (28.6%)	6 (28.6%)	5 (23.8%)	0 (0%)	10 (66.7%)
cluster 3	24 (68.9%)	23 (62.2%)	24 (68.9%)	21 (56.8%)	19 (51.4%)	0 (0%)

The values indicate the number of variations, and the values in parentheses indicate the frequencies of variations after removing statistical missing. ‘ampl.’: amplification, ‘del.’: deletion.

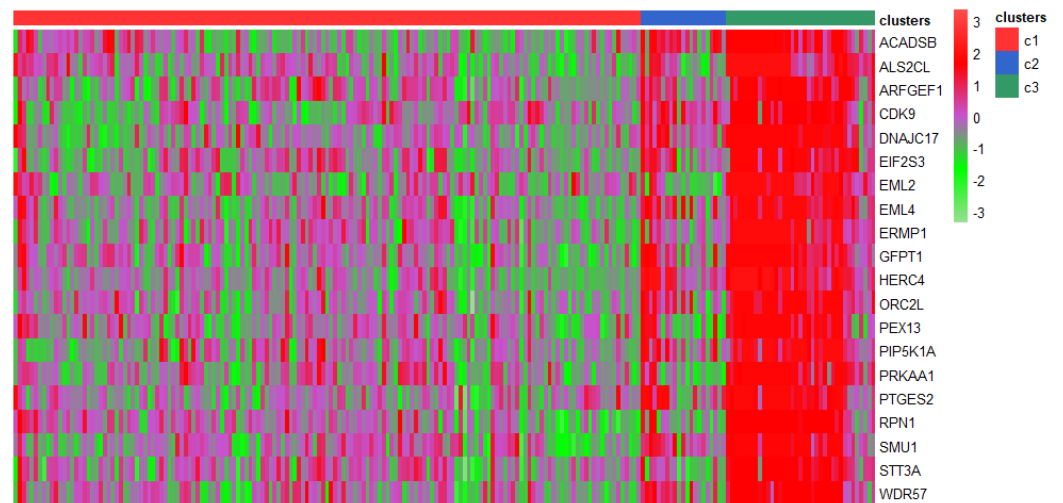
In addition, mRNA expression data and DNA methylation data were used to compare the differentially expressed genes in cluster 1 and 3 to look for the heterogeneity between them. We compared the genes in the two clusters using ANOVA (the lower the  $p$ -value, the higher the ranking). The gene differences in miRNA expression data were not significant enough ( $p$  values were all greater than 0.1) and were omitted from consideration. Figure 5 shows the heatmaps of the top 20 differentially expressed genes in the mRNA expression data and the DNA methylation data, respectively. It is obvious that cluster 1 and cluster 3 are different in gene expression level, and some of the genes on the heatmaps have been shown to be linked to GBM, e.g., *PRKAA1* overexpressed in Cluster 3, also known as *AMPK*, induces antitumor activity in GBM cells and has become a possible tumor control target [42]. *MUC1* overexpressed in cluster 1 is a pathogenic gene that induces GBM and can be used as a target for cellular immunotherapy [43].



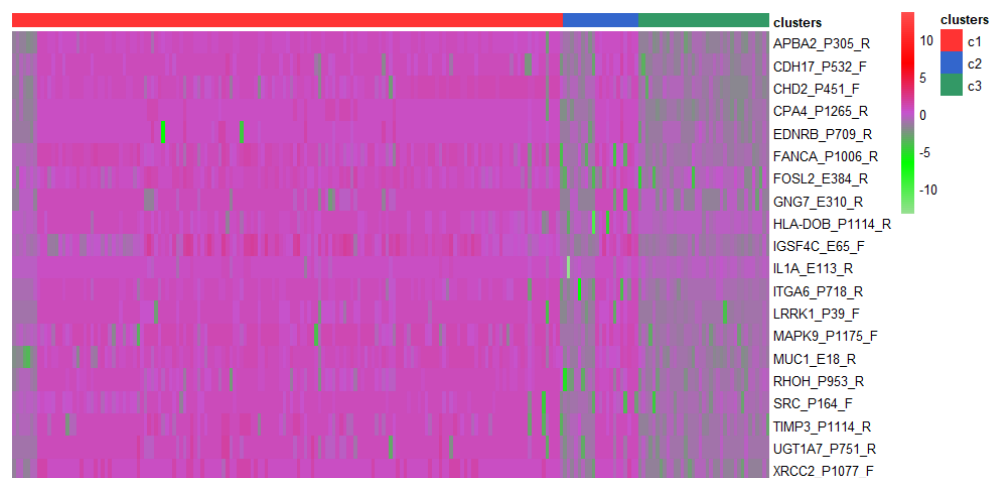
**Figure 4.** The Kaplan–Meier survival curves of the identified clusters (a): cluster 1, (b): cluster 2 and (c): cluster 3) of Temozolomide (TMZ) response. “Untreated” expresses the group which did not receive TMZ treatment and “Treated” expresses the group which received TMZ treatment.

Finally, we compared the three clusters with normal samples and screened for differentially expressed genes using ANOVA. We did Gene Ontology (GO: BP), KEGG pathway and Disease Ontology (DO) enrichment analysis using the top 100 differential genes in

ToppGene Suite database (<https://toppgene.cchmc.org/enrichment.jsp> accessed on 20 December 2020). From Table 7, it is clear that the biological processes of cluster 1 are related to “epithelium development” and “cell adhesion”, while the biological processes of cluster 2 and 3 are mostly related to “protein targeting” and “protein localization”. Moreover, it is interesting to note that all three clusters are associated with anemia in DO enrichment analysis. It is possible that GBM patients treated with TMZ will develop aplastic anemia [44].



(a) mRNA expression data



(b) DNA methylation data

**Figure 5.** Heatmaps of differentially expressed genes in (a): mRNA expression data and (b): DNA methylation data for the identified clusters.

**Table 7.** GO: BP, KEGG pathway, DO enriched terms for the identified cluster.

ENRICHMENT Analysis	Cluster 1	Cluster 2	Cluster 3
GO:BP enriched terms	<ol style="list-style-type: none"> <li>1. Epithelial cell differentiation</li> <li>2. Epithelium development</li> <li>3. Cell adhesion</li> <li>4. Biological adhesion</li> <li>5. cell–cell adhesion</li> </ol>	<ol style="list-style-type: none"> <li>1. Protein targeting to ER</li> <li>2. Establishment of protein localization to endoplasmic reticulum</li> <li>3. Protein localization to endoplasmic reticulum</li> <li>4. Peptide metabolic process</li> <li>5. Protein targeting protein targeting</li> </ol>	<ol style="list-style-type: none"> <li>1. SRP-dependent cotranslational protein targeting to membrane</li> <li>2. Cotranslational protein targeting to membrane</li> <li>3. Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay</li> <li>4. Protein targeting to ER</li> <li>5. Establishment of protein localization to endoplasmic reticulum</li> </ol>
KEGG enriched pathway terms	<ol style="list-style-type: none"> <li>1. Cell adhesion molecules (CAMs)</li> <li>2. Tight junction</li> <li>3. Pathogenic Escherichia coli infection</li> <li>4. Leukocyte transendothelial migration</li> </ol>	<ol style="list-style-type: none"> <li>1. Ribosome</li> <li>2. Protein export</li> </ol>	<ol style="list-style-type: none"> <li>1. Ribosome</li> </ol>
DO enriched terms	<ol style="list-style-type: none"> <li>1. alphaThalassemia</li> <li>2. Dysfibrinogenemia, congenital</li> <li>3. Afibrinogenemia, congenital</li> <li>4. Heinz body anemia</li> </ol>	<ol style="list-style-type: none"> <li>1. Diamond–Blackfan anemia</li> </ol>	<ol style="list-style-type: none"> <li>1. Diamond–Blackfan anemia</li> </ol>

We put the GO: BP terms with ranking in the top 5, and KEGG pathway and DO terms with *p*-value less than 1.00E-4 in the table.

#### 4. Discussion and Conclusions

Over the past decade, it has been widely recognized that the integration and mining of different types of biological data provides meaningful insights into the causes and complexity systems of cancer [45]. Now, the challenge is still how to capture the underlying structure of the sample/features from the omics data for application to a wide range of bioinformatics topics, e.g., the prediction of drug-target relationships [46], the recognition of cancer driver genes [47], finding out about genotype–epigenetic interactions [48], etc.

In this paper, our proposed LRCMC algorithm has the ability to fuse multigenomic data into the consensus graph of the exact connected components. In fact, the sample size of a given cancer is generally relatively small, so the graph learning method can quickly map the feature space into the structure of the affinity graph without the need for feature prescreening. Based on the framework of graph learning, LRCMC uses the operation of LRC to mine the structure of clustering while maintaining the graph structure. At the same time, the graph obtained by the adaptive neighbors method is sparse, so that the weak similarity relationship is even more sparse at 0, which ensures more accurate clustering results. Compared with other start-of-the-art integration clustering algorithms for cancer subtype recognition, LRCMC has the following two characteristics: (1) instead of simply treating each view equally, a wealth of heterogeneous information is taken into account to



provide appropriate weight for each view; (2) the tasks of constructing the affinity matrix of each view, learning the fused matrix and clustering are completed simultaneously in a system. Furthermore, LRCMC has the following two advantages in algorithm running: (1) there is no need to spend a lot of time choosing the appropriate parameters; (2) the final consensus graph has been assigned to the given categories without adding additional base clustering algorithms. We demonstrated the power of LRCMC using four benchmark datasets and four cancer datasets. The experiments show that LRCMC has a good clustering evaluation. The cancer subtype recognition results on GBM data show that LRCMC can effectively capture cancer subtypes with specific biological characteristics based on omics data.

In addition, we must admit that LRCMC also has shortcomings and limitations. It is not suitable for binary data (somatic mutation) or categorical data (copy number states: loss/normal/gain), and has only limited application to continuous data (mRNA expression) to identify cancer subtype. It also does not have the ability to find the gene modules that affect differences in each subtype. Therefore, we will continue our efforts to improve and extend the LRCMC algorithm to explore cancer heterogeneity.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/genes12040526/s1>, File 1: Source code and datasets.

**Author Contributions:** S.G. and J.L. conceived and designed the approach; S.G. and J.L. programmed the algorithm; S.G., X.W. and Y.C. analyzed the results of the experiment; S.G., J.L. and X.W. wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (Grant No. 61906198, 61976215, 61772532) and the Natural Science Foundation of Jiangsu Province (Grant No. BK20190622).

**Institutional Review Board Statement:** Ethical review and approval were waived for this study, due to that the samples were from public TCGA databases.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article and supplementary materials.

**Acknowledgments:** We thank the anonymous reviewers and the editor panel for the efforts generously offered to help improving the manuscript.

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

## References

1. Burrell, R.A.; McGranahan, N.; Bartek, J.; Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **2013**, *501*, 338–345. [[CrossRef](#)]
2. Bedard, P.L.; Hansen, A.R.; Ratain, M.J.; Siu, L.L. Tumour heterogeneity in the clinic. *Nature* **2013**, *501*, 355–364. [[CrossRef](#)]
3. Schuster, S.C. Next-generation sequencing transforms today's biology. *Nat. Methods* **2008**, *5*, 16–18. [[CrossRef](#)]
4. Akbani, R.; Ng, K.S.; Werner, H.M.; Zhang, F.; Ju, Z.; Liu, W.; Liu, W.; Yang, J.Y.; Yoshihara, K.; Li, J.; et al. A pan-cancer proteomic analysis of The Cancer Genome Atlas (TCGA) project. *Cancer Res.* **2014**, *74*, 4262.
5. Shen, R.; Olshen, A.B.; Ladanyi, M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **2009**, *25*, 2906–2912. [[CrossRef](#)]
6. Mo, Q.; Wang, S.; Seshan, V.E.; Olshen, A.B.; Schultz, N.; Sander, C.; Powers, R.S.; Ladanyi, M.; Shen, R. Pattern discovery and cancer gene recognition in integrated cancer genomic data. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 4245–4250. [[CrossRef](#)]
7. Shihua, Z.; Chun-Chi, L.; Wenyuan, L.; Hui, S.; Laird, P.W.; Jasmine, Z.X. Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Res.* **2012**, *19*, 9379–9391.
8. Wu, D.; Wang, D.; Zhang, M.Q.; Gu, J. Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: Application to cancer molecular classification. *BMC Genom.* **2015**, *16*, 1022. [[CrossRef](#)] [[PubMed](#)]
9. Speicher, N.K.; Pfeifer, N. Integrating different data types by regularized unsupervised multiple kernel learning with application to cancer subtype discovery. *Bioinformatics* **2015**, *31*, i268–i275. [[CrossRef](#)] [[PubMed](#)]
10. Shi, Q.; Zhang, C.; Peng, M.; Yu, X.; Zeng, T.; Liu, J.; Chen, L. Pattern fusion analysis by adaptive alignment of multiple heterogeneous omics data. *Bioinformatics* **2017**, *33*, 2706–2714. [[CrossRef](#)] [[PubMed](#)]
11. Wang, B.; Mezlini, A.M.; Demir, F.; Fiume, M.; Tu, Z.; Brudno, M.; Haibe-Kains, B.; Goldenberg, A. Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* **2014**, *11*, 333. [[CrossRef](#)] [[PubMed](#)]

12. Ma, T.; Zhang, A. Integrate multi-omic data using affinity network fusion (anf) for cancer patient clustering. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 13–16 November 2017; pp. 398–403.
13. Guo, Y.; Zheng, J.; Shang, X.; Li, Z. A similarity regression fusion model for integrating multi-omics data to identify cancer subtypes. *Genes* **2018**, *9*, 314. [[CrossRef](#)] [[PubMed](#)]
14. Guo, Y.; Li, H.; Cai, M.; Li, L. Integrative subspace clustering by common and specific decomposition for applications on cancer subtype recognition. *BMC Med Genom.* **2019**, *12*, 1–17. [[CrossRef](#)]
15. Meng, C.; Helm, D.; Frejno, M.; Kuster, B. moCluster: Identifying joint patterns across multiple omics data sets. *J. Proteome Res.* **2016**, *15*, 755–765. [[CrossRef](#)]
16. Shi, Q.; Hu, B.; Zeng, T.; Zhang, C. Multi-view subspace clustering analysis for aggregating multiple heterogeneous omics data. *Front. Genet.* **2019**, *10*, 744. [[CrossRef](#)] [[PubMed](#)]
17. Yu, Y.; Zhang, L.H.; Zhang, S. Simultaneous clustering of multiview biomedical data using manifold optimization. *Bioinformatics* **2019**, *35*, 4029–4037. [[CrossRef](#)]
18. Kumar, A.; Daumé, H. A co-training approach for multi-view spectral clustering. In Proceedings of the 28th international conference on machine learning (ICML-11), Bellevue, WA, USA, 28 June–2 July 2011; pp. 393–400.
19. Hu, Z.; Nie, F.; Chang, W.; Hao, S.; Wang, R.; Li, X. Multi-view spectral clustering via sparse graph learning. *Neurocomputing* **2020**, *384*, 1–10. [[CrossRef](#)]
20. Wang, H.; Yang, Y.; Liu, B. GMC: Graph-based multi-view clustering. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1116–1129. [[CrossRef](#)]
21. Zhan, K.; Zhang, C.; Guan, J.; Wang, J. Graph learning for multiview clustering. *IEEE Trans. Cybern.* **2017**, *48*, 2887–2895. [[CrossRef](#)]
22. Liu, B.Y.; Huang, L.; Wang, C.D.; Lai, J.H.; Yu, P. Multi-view Consensus Proximity Learning for Clustering. *IEEE Trans. Knowl. Data Eng.* **2020**. [[CrossRef](#)]
23. Nie, F.; Cai, G.; Li, X. Multi-view clustering and semi-supervised classification with adaptive neighbours. *Proc. AAAI Conf. Artif. Intell.* **2017**, *31*, 2408–2414.
24. Nie, F.; Li, J.; Li, X. Self-weighted Multiview Clustering with Multiple Graphs. *IJCAI* **2017**, 2564–2570. [[CrossRef](#)]
25. Wang, Y.; Zhang, W.; Wu, L.; Lin, X.; Fang, M.; Pan, S. Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering. *arXiv* **2016**, arXiv:1608.05560.
26. Cao, X.; Zhang, C.; Fu, H.; Liu, S.; Zhang, H. Diversity-induced multi-view subspace clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 586–594.
27. Nie, F.; Wang, X.; Jordan, M.; Huang, H. The constrained laplacian rank algorithm for graph-based clustering. *Proc. AAAI Conf. Artif. Intell.* **2016**, *30*, 1969–1976.
28. Hu, H.; Lin, Z.; Feng, J.; Zhou, J. Smooth representation clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3834–3841.
29. Kang, Z.; Peng, C.; Cheng, Q. Twin learning for similarity and clustering: A unified kernel approach. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
30. Mohar, B.; Alavi, Y.; Chartrand, G.; Oellermann, O.R. The Laplacian spectrum of graphs. *Graph. Theory Comb. Appl.* **1991**, *2*, 871–898.
31. Chung, F.R.; Graham, F.C. *Spectral Graph Theory*; American Mathematical Society: Providence, RI, USA, 1997.
32. Tarjan, R. Depth-first search and linear graph algorithms. *SIAM J. Comput.* **1972**, *1*, 146–160. [[CrossRef](#)]
33. Fan, K. On a theorem of Weyl concerning eigenvalues of linear transformations I. *Proc. Natl. Acad. Sci. USA* **1949**, *35*, 652. [[CrossRef](#)]
34. Boyd, S.; Boyd, S.P.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: London, UK, 2004.
35. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *31*, 210–227. [[CrossRef](#)]
36. Dueck, D.; Frey, B.J. Non-metric affinity propagation for unsupervised image categorization. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–8.
37. Winn, J.; Jojic, N. Locus: Learning object classes with unsupervised segmentation. *Tenth IEEE Int. Conf. Comput. Vis.* **2005**, *1*, 756–763.
38. Hosmer, D.W.; Lemeshow, S.; May, S. Applied survival analysis: Regression modeling of time to event data. *J. Stat. Plan. Inference* **2000**, *91*, 173–175.
39. Verhaak, R.G.; Hoadley, K.A.; Purdom, E.; Wang, V.; Qi, Y.; Wilkerson, M.D.; Miller, C.R.; Ding, L.; Golub, T.; Mesirov, J.P.; et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **2010**, *17*, 98–110. [[CrossRef](#)] [[PubMed](#)]
40. Nounshmehr, H.; Weisenberger, D.J.; Diefes, K.; Phillips, H.S.; Pujara, K.; Berman, B.P.; Pan, F.; Pelloski, C.E.; Sulman, E.P.; Bhat, K.P.; et al. Recognition of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **2010**, *17*, 510–522. [[CrossRef](#)]
41. Brennan, C.W.; Verhaak, R.G.; McKenna, A.; Campos, B.; Nounshmehr, H.; Salama, S.R.; Zheng, S.; Chakravarty, D.; Sanborn, Z.; Berman, S.H.; et al. The somatic genomic landscape glioblastoma. *Cell* **2013**, *155*, 462–477. [[CrossRef](#)] [[PubMed](#)]

42. Lee, D.H.; Lee, T.H.; Jung, C.H.; Kim, Y.H. Wogonin induces apoptosis by activating the AMPK and p53 signaling pathways in human glioblastoma cells. *Cell. Signal.* **2012**, *24*, 2216–2225. [[CrossRef](#)] [[PubMed](#)]
43. Rodriguez, A.; Brown, C.; Badie, B. Chimeric antigen receptor T-cell therapy for glioblastoma. *Transl. Res.* **2017**, *187*, 93–102. [[CrossRef](#)]
44. Villano, J.L.; Collins, C.A.; Manasanch, E.E.; Ramaprasad, C.; van Besien, K. Aplastic anaemia in patient with glioblastoma multiforme treated with temozolomide. *Lancet Oncol.* **2006**, *7*, 436–438. [[CrossRef](#)]
45. Li, Y.; Wu, F.X.; Ngom, A. A review on machine learning principles for multi-view biological data integration. *Brief. Bioinform.* **2018**, *19*, 325–340. [[CrossRef](#)]
46. Liang, X.; Zhang, P.; Yan, L.; Fu, Y.; Peng, F.; Qu, L.; Shao, M.; Chen, Y.; Chen, Z. LRSSL: Predict and interpret drug–disease associations based on data integration using sparse subspace learning. *Bioinformatics* **2017**, *33*, 1187–1196. [[CrossRef](#)]
47. Bashashati, A.; Haffari, G.; Ding, J.; Ha, G.; Lui, K.; Rosner, J.; Huntsman, D.G.; Caldas, C.; Aparicio, S.A.; Shah, S.P. DriverNet: Uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol.* **2012**, *13*, 1–14. [[CrossRef](#)] [[PubMed](#)]
48. Ritchie, M.D.; Holzinger, E.R.; Li, R.; Pendergrass, S.A.; Kim, D. Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* **2015**, *16*, 85–97. [[CrossRef](#)]