Supplementary Material: Approximate Graph Laplacians for Multimodal Data Clustering

Aparajita Khan and Pradipta Maji, *Senior Member, IEEE*

1 Proofs of Eigenspace Approximation

This section contains proofs for Theorems 2 and 3 and Corollary 1 established in Section 4 of the main paper. The basics of matrix perturbation theory [1] required for the ease of understanding these proofs is provided in the appendix.

Let there be M shifted normalized graph Laplacians $L_m \in \mathbb{R}^{n\times n}$ for $m = 1,\ldots,M$, each of which is a symmetric and positive semi-definite. Let L_m have eigenvalues $\lambda_1^m \geq \ldots \geq \lambda_n^m$ and corresponding eigenvectors $U_m = [u_1^m,\ldots,u_n^m]$. Let joint Laplacian $\mathbf{L} = \sum_{m=1}^{M} \alpha_m L_m$, such that $\alpha_m \geq 0$ and $\sum_{m=1}^{M} \alpha_m = 1$. Let the eigen decomposition of $\mathbf{L}$ be given by $\mathbf{L} = \mathbf{Z}\Gamma\mathbf{Z}^T$ with eigenvalues $\gamma_1 \geq \ldots \geq \gamma_n$. For a fixed integer r, we assume $\gamma_r \neq \gamma_{r+1}$. Let approximate joint Laplacian $\mathbf{L}^{r*} = \sum_{m=1}^{M} \alpha_m L_m^r$ be the convex combination of best rank r approximation L_m^r of individual L_m's. Also, let the approximation of L_m using the eigenpairs $(r+1)$ to n be denoted by $L_m^{r\perp}$. Let $\mathbf{L}^{r*}$ have eigenvalues $\pi_1,\ldots,\pi_n$ and its eigen-decomposition be given by $\mathbf{L}^{r*} = \mathbf{V}\Pi\mathbf{V}^T$. Let $\mathbf{Z}^r$ and $\mathbf{V}^r$ be the matrices formed by the first r columns of $\mathbf{Z}$ and $\mathbf{V}$, respectively.

Theorem 1. *For any unitarily invariant norm $\|\ .\ \|$, the following bound holds on the principal angles between the subspaces defined by $\mathcal{C}(\mathbf{Z}^r)$ and $\mathcal{C}(\mathbf{V}^r)$:*

$$\|\sin\Theta\left(\mathcal{C}(\mathbf{Z}^r),\mathcal{C}(\mathbf{V}^r)\right)\| \leq \frac{\left\|\left(\sum_{m=1}^{M}\alpha_m L_m^{r\perp}\right)\mathbf{V}^r\right\|}{\left(\pi_r - \pi_{r+1} - \sum_{m=1}^{M}\alpha_m\lambda_{r+1}^m\right)}, \quad (1)$$

assuming $\pi_r > \pi_{r+1} + \sum_{m=1}^{M}\alpha_m\lambda_{r+1}^m$.

Proof. The matrices $\mathbf{Z}$ and Γ contain the eigenpairs of $\mathbf{L}$. For the given r, let $\mathbf{Z}$ and Γ be partitioned as

$$\mathbf{Z} = \begin{bmatrix}\mathbf{Z}^r & \mathbf{Z}^{r\perp}\end{bmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix}\Gamma^r & \mathbf{0}\\ \mathbf{0} & \Gamma^{r\perp}\end{bmatrix}. \quad (2)$$

• *A. Khan and P. Maji are with the Biomedical Imaging and Bioinformatics Lab, Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India. E-mail: {aparajitak_r, pmaji}@isical.ac.in.*

Since $\mathbf{Z}^r$ and $\mathbf{Z}^{r\perp}$ contain eigenvectors of $\mathbf{L}$, so,

$$\mathbf{L}\mathbf{Z}^r = \mathbf{Z}^r\Gamma^r \subset \mathcal{C}(\mathbf{Z}^r). \quad (3)$$

This implies that the transformation of any vector $v \in \mathcal{C}(\mathbf{Z}^r)$ lies in $\mathcal{C}(\mathbf{Z}^r)$ itself. So, $\mathbf{Z}^r$ spans an invariant subspace of the matrix $\mathbf{L}$ [1]. Similarly,

$$\mathbf{L}\mathbf{Z}^{r\perp} = \mathbf{Z}^{r\perp}\Gamma^{r\perp} \subset \mathcal{C}(\mathbf{Z}^{r\perp}). \quad (4)$$

So, $\mathbf{Z}^{r\perp}$ also spans an invariant subspace of $\mathbf{L}$. Moreover, the columns of $\mathbf{Z}^{r\perp}$ span the subspace orthogonal to the one spanned by the columns of $\mathbf{Z}^r$. Now, let

$$\mathbf{B}_1 = (\mathbf{Z}^r)^T\mathbf{L}\mathbf{Z}^r = \Gamma^r \text{ and } \mathbf{B}_2 = (\mathbf{Z}^{r\perp})^T\mathbf{L}\mathbf{Z}^{r\perp} = \Gamma^{r\perp}. \quad (5)$$

According to the theory of invariant subspaces [1], $\mathbf{B}_1$ and $\mathbf{B}_2$ are called the representation of $\mathbf{L}$ with respect to the bases $\mathbf{Z}^r$ and $\mathbf{Z}^{r\perp}$, respectively. The matrix $\mathbf{B}_1 = \Gamma^r$ contains eigenvalues $\gamma_1,\ldots,\gamma_r$, while $\mathbf{B}_2 = \Gamma^{r\perp}$ contains eigenvalues $\gamma_{r+1},\ldots,\gamma_n$. Let $\Omega(B)$ denote the set of eigenvalues of a matrix B. Under the assumption that $\gamma_r \neq \gamma_{r+1}$, we have

$$\Omega(\mathbf{B}_1) \cap \Omega(\mathbf{B}_2) = \emptyset. \quad (6)$$

It follows from (6) that the eigenvalues of $\mathbf{B}_1$ and $\mathbf{B}_2$ are non-intersecting. So, $\mathbf{Z}^r$ spans a simple invariant subspace of $\mathbf{L}$ with its complementary subspace being spanned by $\mathbf{Z}^{r\perp}$. Also, $\begin{bmatrix}\mathbf{Z}^r & \mathbf{Z}^{r\perp}\end{bmatrix}$ is unitary and $\mathbf{L}$ can be decomposed as

$$\mathbf{L} = \mathbf{Z}^r\mathbf{B}_1(\mathbf{Z}^r)^T + \mathbf{Z}^{r\perp}\mathbf{B}_2(\mathbf{Z}^{r\perp})^T. \quad (7)$$

The decomposition in (7) is called the spectral resolution of $\mathbf{L}$ along $\mathbf{Z}^r$ and $\mathbf{Z}^{r\perp}$. Now, let $\mathbf{L}$ be written as

$$\mathbf{L} = \sum_{m=1}^{M}\alpha_m L_m = \sum_{m=1}^{M}\alpha_m\left(L_m^r + L_m^{r\perp}\right),$$

$$\Rightarrow \quad \mathbf{L} = \sum_{m=1}^{M}\alpha_m L_m^r + \sum_{m=1}^{M}\alpha_m L_m^{r\perp},$$

$$\Rightarrow \quad \mathbf{L} = \mathbf{L}^{r*} + \mathbf{L}^{r\perp *}, \text{ where } \mathbf{L}^{r\perp *} = \sum_{m=1}^{M}\alpha_m L_m^{r\perp}. \quad (8)$$

Let the eigenvectors and eigenvalues of $\mathbf{L}^{r*}$ be partitioned as

$$\mathbf{V} = \begin{bmatrix}\mathbf{V}^r & \mathbf{V}^{r\perp}\end{bmatrix} \quad \text{and} \quad \Pi = \begin{bmatrix}\Pi^r & \mathbf{0}\\ \mathbf{0} & \Pi^{r\perp}\end{bmatrix}. \quad (9)$$

Since $\mathbf{V}^r$ contains eigenvectors of $\mathbf{L}^{r*}$, so

$$\mathbf{L}^{r*}\mathbf{V}^r = \mathbf{V}^r\Pi^r \subset \mathcal{C}(\mathbf{V}^r). \tag{10}$$

This implies that $\mathbf{V}^r$ spans an invariant subspace of $\mathbf{L}^{r*}$ and Π^r is a Hermitian matrix of order r which gives the representation of $\mathbf{L}^{r*}$ with respect to the basis $\mathbf{V}^r$. According to (8), $\mathbf{L}$ can be written as the sum of $\mathbf{L}^{r*}$ and a perturbation $\mathbf{L}^{r\perp*}$. The perturbation theory [1] analyzes how near is the perturbed subspace $\mathcal{C}(\mathbf{V}^r)$ to an invariant subspace $\mathcal{C}(\mathbf{Z}^r)$ of $\mathbf{L}$, in terms of the perturbation matrix $\mathbf{L}^{r\perp*}$. So, the residual $\mathcal{R}$ of the matrix $\mathbf{L}$, with respect to a perturbed basis $\mathbf{V}^r$ and the Hermitian matrix Π^r, is given by

$$\begin{aligned}
\mathcal{R} &= \mathbf{L}\mathbf{V}^r - \mathbf{V}^r\Pi^r \\
&= \left(\mathbf{L}^{r*} + \sum_{m=1}^{M}\alpha_m L_m^{r\perp}\right)\mathbf{V}^r - \mathbf{V}^r\Pi^r \qquad \text{[from (8)]} \\
&= \mathbf{L}^{r*}\mathbf{V}^r + \left(\sum_{m=1}^{M}\alpha_m L_m^{r\perp}\right)\mathbf{V}^r - \mathbf{V}^r\Pi^r \\
&= \mathbf{V}^r\Pi^r + \left(\sum_{m=1}^{M}\alpha_m L_m^{r\perp}\right)\mathbf{V}^r - \mathbf{V}^r\Pi^r \\
&= \left(\sum_{m=1}^{M}\alpha_m L_m^{r\perp}\right)\mathbf{V}^r.
\end{aligned} \tag{11}$$

The matrices Π^r and $\mathbf{B}_2 = \Gamma^{r\perp}$ consist of the eigenvalues of the perturbed subspace $\mathcal{C}(\mathbf{V}^r)$ and the complementary invariant subspace $\mathcal{C}(\mathbf{Z}^{r\perp})$, respectively. According to the Davis-Kahan theorem [2], the bound on the difference between an invariant subspace $\mathcal{C}(\mathbf{Z}^r)$ and its perturbation $\mathcal{C}(\mathbf{V}^r)$ holds only if the eigenvalues of the perturbed subspace and the complementary invariant subspace are non-intersecting. So, the range in which the eigenvalues of Π^r and $\mathbf{B}_2$ lie are derived.

The matrix Π contains the eigenvalues of $\mathbf{L}^{r*}$ given by $\Pi = diag(\pi_1, \ldots, \pi_r, \pi_{r+1}, \ldots, \pi_n)$ which can be partitioned into Π^r and $\Pi^{r\perp}$ as in (9). So, the eigenvalues of Π^r satisfy

$$\Omega(\Pi^r) \in [\pi_r, \pi_1]. \tag{12}$$

The range of the eigenvalues of $\mathbf{B}_2$ is derived next. Since each L_m is a real symmetric matrix, its low-rank approximations L_m^r and $L_m^{r\perp}$ are also real symmetric matrices. So, each L_m^r and $L_m^{r\perp}$ have the Hermitian property and $\mathbf{L}^{r\perp*}$ is the sum of M Hermitian matrices according to (8). The eigenvalues of $L_m^{r\perp}$ lie in $[\lambda_{r+1}^m, \lambda_n^m]$, and those of $\alpha_m L_m^{r\perp}$ lie in $[\alpha_m\lambda_{r+1}^m, \alpha_m\lambda_n^m]$. Applying Weyl's inequality [1] for the eigenvalues of sum of Hermitian matrices to $\mathbf{L}^{r\perp*}$, we get

$$\Omega(\mathbf{L}^{r\perp*}) \in \left[\sum_{m=1}^{M}\alpha_m\lambda_{r+1}^m, \sum_{m=1}^{M}\alpha_m\lambda_n^m\right]. \tag{13}$$

The eigenvalues of $\mathbf{L}$ lie in $[\gamma_n, \gamma_1]$, while those of $\mathbf{L}^{r*}$ lie in $[\pi_n, \pi_1]$. The range of eigenvalues of $\mathbf{L}^{r\perp*}$ is given by (13). Again, $\mathbf{L}$ $(= \mathbf{L}^{r*} + \mathbf{L}^{r\perp*})$ is the sum of two Hermitian matrices $\mathbf{L}^{r*}$ and $\mathbf{L}^{r\perp*}$. So, using Weyl's inequality, the eigenvalues of $\mathbf{L}$ satisfy

$$\pi_j + \sum_{m=1}^{M}\alpha_m\lambda_n^m \le \gamma_j \le \pi_j + \sum_{m=1}^{M}\alpha_m\lambda_{r+1}^m, \tag{14}$$

for $j = 1, ..., n$. As stated previously, $\mathbf{B}_2 = \Gamma^{r\perp}$ consists of eigenvalues $\gamma_{r+1}, \ldots, \gamma_n$ of $\mathbf{L}$. Thus, the maximum eigenvalue of $\mathbf{B}_2$ is γ_{r+1}, which using (14) is bounded by

$$\gamma_{r+1} \le \pi_{r+1} + \sum_{m=1}^{M}\alpha_m\lambda_{r+1}^m. \tag{15}$$

According to (12), the minimum eigenvalue of Π^r is π_r. Let δ be the minimum of the separation between the eigenvalues of Π^r and $\mathbf{B}_2$, which is given by

$$\begin{aligned}
\delta &= \min\{\Omega(\Pi^r)\} - \max\{\Omega(\mathbf{B}_2)\} \\
&= \pi_r - \pi_{r+1} - \sum_{m=1}^{M}\alpha_m\lambda_{r+1}^m > 0.
\end{aligned} \tag{16}$$

$$\text{So, } \pi_r - \delta = \pi_{r+1} + \sum_{m=1}^{M}\alpha_m\lambda_{r+1}^m. \tag{17}$$

From (15) and (17), we get $\gamma_{r+1} \le (\pi_r - \delta)$. Moreover, as $\gamma_n \le \gamma_{r+1}$, $\gamma_n \le (\pi_r - \delta)$. Also, $(\pi_1 + \delta) \ge (\pi_r - \delta)$, as $\pi_1 \ge \pi_r$. This implies that the eigenvalues of $\mathbf{B}_2$, that is, $\gamma_{r+1}, \ldots, \gamma_n$ satisfy

$$\Omega(\mathbf{B}_2) \in \mathbb{R} \setminus [\pi_r - \delta, \pi_1 + \delta]. \tag{18}$$

The constraints in (12) and (18) imply that the eigenvalues of Π^r are included in the interval $[\pi_r, \pi_1]$, while those of $\mathbf{B}_2$ are excluded from the interval $[\pi_r - \delta, \pi_1 + \delta]$, where $\delta > 0$. So, for an invariant subspace $\mathcal{C}(\mathbf{Z}^r)$, the eigenvalues of its complementary subspace $\mathcal{C}(\mathbf{Z}^{r\perp})$ and those of its perturbed subspace $\mathcal{C}(\mathbf{V}^r)$ are non-intersecting. Finally, according to the Davis-Kahan theorem [2] which bounds the difference between an invariant subspace and its perturbation, for any unitarily invariant norm $\| \cdot \|$,

$$\|\sin\Theta\left(\mathcal{C}(\mathbf{Z}^r), \mathcal{C}(\mathbf{V}^r)\right)\| \le \frac{\| \mathcal{R} \|}{\delta}. \tag{19}$$

Substituting the value of $\mathcal{R}$ and δ from (11) and (16), respectively, in (19), we get

$$\|\sin\Theta\left(\mathcal{C}(\mathbf{Z}^r), \mathcal{C}(\mathbf{V}^r)\right)\| \le \frac{\left\|\left(\sum_{m=1}^{M}\alpha_m L_m^{r\perp}\right)\mathbf{V}^r\right\|}{\left(\pi_r - \pi_{r+1} - \sum_{m=1}^{M}\alpha_m\lambda_{r+1}^m\right)} \tag{20}$$

This concludes the proof. ■

Corollary 1. *Let $tr(B)$ denote the trace of matrix B. Then,*

$$\|\sin\Theta\left(\mathcal{C}(\mathbf{Z}^r), \mathcal{C}(\mathbf{V}^r)\right)\|_F^2 \le \frac{tr\left((\mathbf{V}^r)^T\left(\sum_{m=1}^{M}\alpha_m L_m^{r\perp}\right)^2\mathbf{V}^r\right)}{\left(\pi_r - \pi_{r+1} - \sum_{m=1}^{M}\alpha_m\lambda_{r+1}^m\right)}, \tag{21}$$

Proof. The Frobenius norm of a matrix B, given by $\|B\|_F = \sqrt{tr(B^TB)}$, is an unitarily invariant norm. The squared Frobenius norm of $\mathcal{R}$ in (11) is given by

$$\|\mathcal{R}\|_F^2 = tr\left((\mathbf{V}^r)^T\left(\sum_{m=1}^{M}\alpha_m L_m^{r\perp}\right)^2\mathbf{V}^r\right). \tag{22}$$

The Davis-Kahan theorem holds for any unitarily invariant norm. So, substituting the value of δ and the Frobenius norm of $\mathcal{R}$ in (19), the required bound in (21) is obtained. ■

The eigenvalues of $\mathbf{L}^r$ and $\mathbf{L}^{r*}$ are given by the elements of the diagonal matrices Γ and Π, respectively. The bound on the difference between the eigenvalues is given as follows.

Theorem 2. *The eigenvalues of* $\mathbf{L}$ *and* $\mathbf{L}^{r*}$ *satisfy the following bound:*

$$\sum_{j=1}^{n} (\gamma_j - \pi_j)^2 \leq \sum_{j=r+1}^{n} \sum_{m=1}^{M} \alpha_m (\lambda_j^m)^2. \tag{23}$$

Proof. The decomposition of $\mathbf{L}$ in (8) gives $\mathbf{L} = \mathbf{L}^{r*} + \mathbf{L}^{r\perp*}$. Both $\mathbf{L}^{r*}$ and $\mathbf{L}^{r\perp*}$ are low-rank approximations of the real-symmetric matrix $\mathbf{L}$ using its eigenpairs. So, $\mathbf{L}^{r*}$ and $\mathbf{L}^{r\perp*}$ are also real and symmetric. The eigenvalues of $\mathbf{L}^{r*}$ are given by $\pi_1, \ldots, \pi_n$, while those of $\mathbf{L}^{r\perp*}$ are given by

$$\sum_{m=1}^{M} \alpha_m \lambda_{r+1}^m, \ldots, \sum_{m=1}^{M} \alpha_m \lambda_n^m, \tag{24}$$

according to (13). $\mathbf{L}$ is the sum of two real-symmetric matrices and has eigenvalues $\gamma_1, \ldots, \gamma_n$. The squared Frobenius norm of $\mathbf{L}^{r\perp*}$, given by the sum of squares of its eigenvalues, is

$$\left\| \mathbf{L}^{r\perp*} \right\|_F^2 = \sum_{j=r+1}^{n} \sum_{m=1}^{M} \alpha_m (\lambda_j^m)^2. \tag{25}$$

According to the Weilandt-Hoffman theorem [3], the sum of squares of the difference between the eigenvalues of $\mathbf{L}$ and $\mathbf{L}^{r*}$ is bounded by the squared Frobenius norm of the residual $\mathbf{L}^{r\perp*}$. Therefore,

$$\sum_{j=1}^{n} (\gamma_j - \pi_j)^2 \leq \sum_{j=r+1}^{n} \sum_{m=1}^{M} \alpha_m (\lambda_j^m)^2. \tag{26}$$

This proves the bound on the eigenvalues. ■

2 Description of Datasets

This subsection presents the description of five multi-omics cancer data sets and four benchmark multi-view data sets used for the evaluation of the proposed and the existing algorithms.

2.1 Omics Data Sets

Five real-life multi-omics cancer data sets from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/), are used in this study. EXperimental results of four data sets, namely, colorectal carcinoma (CRC), lower grade glioma (LGG), stomach adenocarcinome (STAD), and breast invasive carcinoma (BRCA) are provided in the main paper. All results of fifth data set, ovarian carcinoma (OV) and additional results of CRC, LGG, STAD, and BRCA data sets are provided in the supplementary material. The five TCGA data sets used in this study are described as follows:

1) **Colorectal carcinoma** (**CRC**): It is the third most commonly diagnosed cancer in both men and women and account for nine percent of all cancer deaths [4]. The colon and rectum are parts of the digestive system and cancer forms in the colon and/or the rectum. There are 307 samples in the OV data set. Depending on the site of origin, the samples of OV are divided into two subtypes, namely, colon carcinoma and rectum carcinoma, having 236 and 71 samples, respectively.
2) **Lower grade glioma** (**LGG**): Diffuse low-grade and intermediate-grade gliomas which together make up the lower-grade gliomas have highly variable clinical behaviour that is not adequately predicted on the basis of histological class. Integrative analysis of data from RNA, DNA-copy-number, and DNA-methylation platforms has uncovered three prognostically significant subtypes of lower-grade glioma [5]. The LGG data set consists of 267 samples. The first subtype has 134 samples which exhibit IDH mutation and no 1p/19q codeletion. The second subtype exhibits both IDH mutation and 1p/19q codeletion and has 84 samples. The third one is called the wild-type IDH subtype and has 49 samples.
3) **Stomach adenocarcinoma** (**STAD**): Stomach/Gastric cancer was the worlds third leading cause of cancer mortality in 2012, responsible for 723,000 deaths [6]. TCGA research network has proposed a molecular classification dividing gastric cancer into four subtypes [7]. The STAD data set has 199 samples which consists of 38 samples from microsatellite unstable tumours, which show elevated mutation rates, 20 samples of tumours showing positivity for EpsteinBarr virus, 97 samples of tumours having chromosomal instability, and 44 samples of genomically stable tumors.
4) **Breast invasive carcinoma** (**BRCA**): Breast cancer is one of the most common cancers with greater than 1,300,000 cases and 450,000 deaths each year worldwide [8]. During the last 15 years, four intrinsic molecular subtypes of breast cancer (Luminal A, Luminal B, HER2-enriched, and Basal-like) have been identified and intensively studied [9], [10], [8]. The BRCA data set consists of 398 samples comprising of 171, 98, 49, and 80 samples of LuminalA, LuminalB, HER2-enriched, and Basal-like subtype, respectively.
5) **Ovarian carcinoma** (**OV**): Ovarian cancer is the eighth most commonly occurring cancer in women and there were nearly 300,000 new cases in 2018 [11]. Ovarian cancer encompasses a heterogeneous group of malignancies that vary in etiology, molecular biology, and numerous other characteristics. TCGA researchers have identified four robust expression subtypes of high-grade serous ovarian cancer [12]. The OV data set consists of 334 samples. The four subtypes are termed as immunoreactive, differentiated, proliferative, and mesenchymal, consisting of 74, 91, 90, and 79 samples, respectively.

These subtypes have been shown to be clinically relevant and provide roadmap for patient stratification and trials of targeted therapies.

Data pre-processing: For all the data sets, four different omic modalities are considered, namely, DNA methylation (mDNA), gene expression (RNA), microRNA expression (miRNA), and reverse phase protein array expression (RPPA). n order to avoid considering features with too many

TABLE S1
Summary of Omics Data Sets

Different Data Sets	No. of Samples (n)	No. of Features					Sample to Feature Ratio	No. of Clusters (k)
		mDNA	RNA	miRNA	RPPA	Total		
CRC	464	2000	2000	291	178	4469	0.103826	2
LGG	267	2000	2000	333	209	4542	0.058784	3
STAD	242	2000	2000	291	218	4509	0.053670	4
BRCA	398	2000	2000	278	212	4490	0.088641	4

missing values, for all the omic modalities, those features for which the corresponding omic expression value is not present for more than 5% of the total number of samples are excluded. For the remaining features, missing values are replaced using 0.

For CRC, LGG, STAD, and BRCA data sets, sequence based RNA and miRNA expression data from Il- lumina HiSeq and Illumina GA platforms are used. The RNA and miRNA modalities contain expression signals for 20, 502 annotated genes and 1046 miRNAs, respectively. However, fil- tering out miRNAs with more than 5% missing values reduced the number miRNAs for the these data sets to around 300. The under- lying assumption of the proposed work is that the data follows multivariate Gaussian distribution. However, the sequence based RNA and miRNA expression modalities of the data sets contain normalized RPKM (reads per kilobase of exon per million) counts for the genes. Count data are known to follow a skewed distribution and have the property that the variance depends on the mean value [13]. It is observed that genes having larger mean expression values also tend to have larger variances and are not normally distributed. Log transformation is generally performed on the sequence based expression data to make the data more or less normally distributed [13]. The degree of normality attained depends on the skewness of the data before transformation. Therefore, for modalities with sequence based count data, the 0 entries are replaced by 1, and then the data is log-transformed using base 10. On the other hand, for the OV data set, array based RNA and miRNA expression data from AgilentG4502A_07_3 and H-miRNA_8x15Kv2 platforms are used. As the RNA and miRNA expression data for the OV data set is observed on microarray based platforms which contain log-ratio based expression data, so the data is not log-transformed as in case of the other four data sets. The RNA modality of OV data set consists of expression for 17,814 genes amongst which 2,000 most variable genes are considered. The miRNA expression data is available for 799 microRNAs.

For the DNA methylation modality, methylation β-values from Illumina HumanMethylation450 and HumanMethylation450 beadarray platforms are used. The HumanMethylation450 beadarray gives methylation β-values of 485,577 CpG sites, while HumanMethylation27 beadarray covers 27,578 CpG sites. These two platforms share a common set of 25,978 CpG locations. Over 94% of loci present on HumanMethylation27 array are included in the HumanMethylation450 array content. Moreover, the correlation between the β-value measurements across the two platforms were compared in https://cancergenome.nih.gov/abouttcga/aboutdata/platformdesign/illuminamethylation450 and [14] which showed strong correlation of $R^2 > 0.97$. Therefore, for all the data set, methylation data across those common 25,978 CpG locations are considered from both the platforms. Additionally, CpG locations with missing gene information were filtered out from the study. The top 2,000 most variable CpG sites are used for clustering. For protein modality, reverse phase protein array data from the MDA_RPPA_Core platform having approximately 220 proteins is used. These four modalities, measured on different platforms represent a wide variety of biological information. The summary of the data sets in terms of their sample size, dimension of their individual modalities, and their number of clusters is provided in Table S1.

2.2 Multimodal Benchmark Data Sets

Four multimodal bechmark data sets, namely, Football, Plotics-uk, Rugby, and Digits from different application domains are used to evaluate the performance of the proposed and the existing algorithms. Among them, Football, Politics-uk, and Rugby are social networking based Twitter data sets, while Digits is an image data set. The benchmark data sets are described as follows:

2.2.1 Twitter Data Sets

A brief description of the three benchmark Twitter data sets used in this work are is as follows:

1) **Football**: This data set is a collection of 248 English Premier League football players and clubs active on Twitter. The disjoint ground truth communities correspond to the 20 individual clubs in the league.
2) **Politics-uk**: This data set consists of 419 Members of Parliament (MPs) in the United Kingdom. The ground truth consists of five groups, corresponding to political parties.
3) **Rugby**: The Rugby data set is a collection of 854 international Rugby Union players, clubs, and organizations currently active on Twitter. The ground truth consists of over- lapping communities corresponding to 15 countries. In the case of players, these user accounts can potentially be assigned to both their home nation and the nation in which they play club rugby. As the full names or screen names of the Twitter users are not available, so the overlapping Rugby players are assigned either to their country or their club.

For each dataset, a heterogeneous collection of nine network and content-based modalities are available. In all cases, cosine similarity is used to compute the pairwise similarities between the Twitter users. All the Twitter data sets are publicly available at http://mlg.ucd.ie/aggregation/.

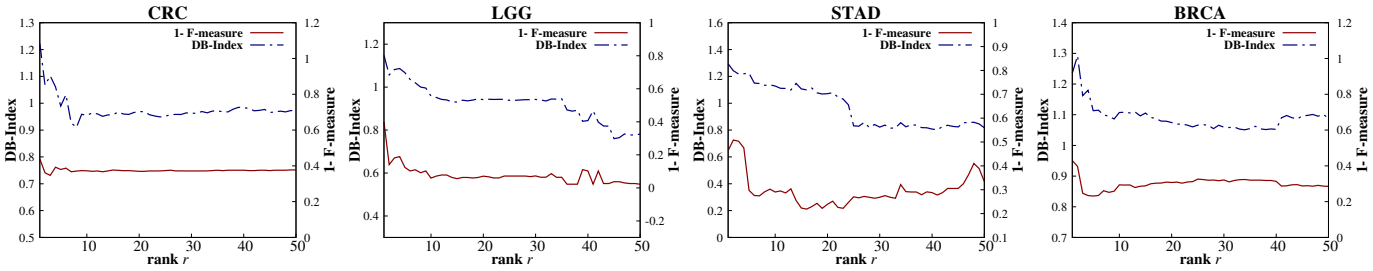


Fig. S1. Variation of DB index and $(1-\text{F-measure})$ for different values of rank r for CRC, LGG, STAD, and BRCA data sets.

Description of the nine different modalities of each Twitter data set is given below:

1) **Tweets500**: User content profiles, constructed from the concatenation of the 500 most recently-posted tweets for each user.
2) **Lists500**: List content profiles, constructed from the concatenation of both the names and the descriptions of the 500 Twitter lists to which each user has most recently been assigned.
3) **Follows**: From the unweighted directed follower graph, construct binary user profile vectors based on the users whom they follow (i.e. out-going links).
4) **Followed-by**: From the unweighted directed follower graph, construct binary user profile vectors based on the users that follow them (i.e. incoming links). A pair of users are deemed to be similar if they are frequently co-followed by the same users.
5) **Mentions**: From the weighted directed mention graph, construct user profile vectors based on the users whom they mention.
6) **Mentioned-by**: From the weighted directed mention graph, construct binary user profile vectors based on the users that mention them. A pair of users are deemed to be similar if they are frequently co-mentioned by the same users.
7) **Retweets**: From the weighted directed retweet graph, construct user profile vectors based on the users whom they retweet.
8) Retweeted-by: From the weighted directed retweet graph, construct user profile vectors based on the users that retweet them. Users are deemed to be similar if they are frequently co-retweeted by the same users.
9) **ListMerged500**: Based on Twitter user list memberships, construct an unweighted bipartite graph, such that an edge between a list and a user indicates that the list contains the specified user. A pair of users are deemed to be similar if they are frequently linked to the same lists. Again, we only consider the 500 lists to which each user has been assigned most recently.

2.2.2 *Image Data Set*

The **Digits** data set is an image data set which consists of features of handwritten numerals ('0'-'9') extracted from a collection of Dutch utility maps with 200 patterns per class (for a total of 2,000 patterns) have been digitized in binary images. The data set is publicly available at https://archive.ics.uci.edu/ml/datasets/Multiple+Features. The samples are represented in terms of the following six feature sets:

1) mfeat-fou: 76 Fourier coefficients of the character shapes.
2) mfeat-fac: 216 profile correlations.
3) mfeat-kar: 64 Karhunen-Love coefficients.
4) mfeat-pix: 240 pixel averages in 2 x 3 windows.
5) mfeat-zer: 47 Zernike moments.
6) mfeat-mor: 6 morphological features.

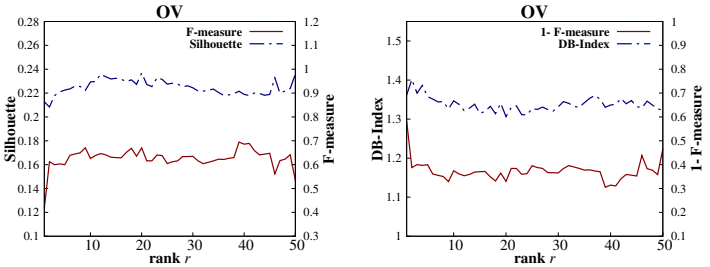


Fig. S2. Variation of Silhouette and DB index with that of F-measure for different values of rank r for OV data sets.

3 Results on Omics Data Sets

This section presents additional results from the four omics data sets, CRC, LGG, STAD, and BRCA which are used to evaluate the performance of the proposed CoALa algorithm. Furthermore, the performance of the proposed and the existing algorithms is studied on the ovarian carcinoma (OV) data set. This section also reports the experimental results on the OV data set.

3.1 Rank Estimation using Davies Bouldin Index

In Section 5.2 of the main paper, the optimum value of the rank r of the individual Laplacians is selected using the Silhouette index. It is also demonstrated in Fig. 1 of the main paper that with the change in rank r, the value of Silhouette index which is an internal cluster evaluation index varies in a similar fashion as that of F-measure, which is an external one. However, other internal cluster validity indices which evaluate the compactness and separability of clusters can also be used to estimate the rank. This subsection illustrates

TABLE S2
Effect of Row-normalization on Individual Laplacians of Omics Data

Data Set	Index	mDNA_RNrm	mDNA	RNA_RNrm	RNA	miRNA_RNrm	miRNA	RPPA_RNrm	RPPA
CRC	F-measure	0.5505426	**0.5849894**	**0.5397796**	**0.5397796**	**0.5694956**	0.5673758	0.5580409	**0.5741394**
	Purity	**0.7370690**	**0.7370690**	**0.7370690**	**0.7370690**	**0.7370690**	**0.7370690**	**0.7370690**	**0.7370690**
	Rand	**0.4996742**	0.4989573	**0.4991528**	**0.4991528**	**0.5026439**	0.5022809	0.4998510	**0.5007448**
	Jaccard	0.3813969	**0.3925508**	**0.3789509**	**0.3789509**	**0.3820386**	0.3818306	0.3810029	**0.3853947**
	Dice	0.5521902	**0.5637867**	**0.5496220**	**0.5496220**	**0.5528624**	0.5526446	0.5517771	**0.5563681**
LGG	F-measure	0.8146853	**0.8269248**	0.5442741	**0.5875701**	0.4593501	**0.4717221**	**0.4432318**	0.4326018
	Purity	0.8164794	**0.8352060**	0.5468165	**0.5917603**	0.5243446	**0.5318352**	**0.5280899**	**0.5280899**
	Rand	0.7706063	**0.7861508**	0.5983780	**0.6149925**	0.5591788	**0.5593760**	**0.5511250**	0.5476050
	Jaccard	0.5294865	**0.5814133**	0.2881813	**0.3235367**	0.2376175	**0.2476680**	0.2312886	**0.2328447**
	Dice	0.6923716	**0.7353085**	0.4474235	**0.4888972**	0.3839918	**0.3970095**	0.3756854	**0.3777356**
STAD	F-measure	0.4576854	**0.5469686**	0.4336250	**0.4781377**	0.4116322	**0.3998266**	0.4260470	**0.4469459**
	Purity	0.5537190	**0.5867769**	0.5371901	**0.5495868**	0.5000000	**0.4917355**	0.4917355	**0.4917355**
	Rand	0.6261445	**0.6509722**	0.6155825	**0.6239155**	0.6008710	**0.5989164**	0.5918521	**0.5883543**
	Jaccard	0.2295951	**0.2869053**	0.2062031	**0.2234653**	0.1911182	**0.1994524**	0.1919343	**0.2076045**
	Dice	0.3734483	**0.4458841**	0.3419044	**0.3652989**	0.3209055	**0.3325725**	0.3220551	**0.3438286**
BRCA	F-measure	0.5910402	**0.5982526**	0.7467001	**0.7690661**	0.4751092	**0.5105008**	0.5073433	**0.5630781**
	Purity	0.6482412	**0.6532663**	0.7412060	**0.7688442**	0.5552764	**0.5703518**	0.5804020	**0.5879397**
	Rand	0.7174031	**0.7193018**	0.7848056	**0.7995519**	0.6479754	**0.6455071**	0.6748478	**0.6689493**
	Jaccard	0.3235980	**0.3318872**	0.4420047	**0.4857607**	0.2442047	**0.2672039**	0.2637432	**0.3132549**
	Dice	0.4889672	**0.4983713**	0.6130420	**0.6538882**	0.3925475	**0.4217221**	0.4174000	**0.4770664**
OV	F-measure	0.3816637	**0.3857003**	**0.6652254**	0.6444234	0.4120201	**0.4186585**	0.3705716	**0.3858236**
	Purity	0.3892216	**0.3922156**	**0.6676647**	0.6497006	**0.4221557**	0.4131737	**0.3712575**	**0.3712575**
	Rand	0.6455737	**0.6469224**	**0.7622233**	0.7536459	**0.6552121**	0.6387046	**0.6415098**	0.6357196
	Jaccard	0.1681087	**0.1705390**	**0.3576703**	0.3517861	0.1864047	**0.1927033**	0.1650892	**0.1725011**
	Dice	0.2878306	**0.2913855**	**0.5268883**	0.5204760	0.3142346	**0.3231370**	0.2833932	**0.2942447**

the use of Davies Bouldin (DB) index [15] index for selection of the optimum value of rank parameter r. DB index is a minimization based index, while, the external index F-measure is maximization based. The value of F-measure lies between $[0, 1]$, while DB index is unbounded. So, the value of F-measure is subtracted from its maximum value, 1, and the difference is compared with that of DB index. Fig. S1 shows the variation in the value of DB index with that of $(1-\text{F-measure})$ with the increase in rank r for CRC, LGG, STAD, and BRCA data sets. The plots in Fig. S1 show that for all these four omics data sets the variation in DB index is very similar to that of $(1-\text{F-measure})$. Since the external and internal measures are found to vary similarly, the optimum value of DB index is likely to produce a nearly optimum value of F-measure for the same parameter configuration. For the OV data set, the variation in F-measure and Silhouette index with increase in rank r is shown in Fig. S2a, while that of $(1-\text{F-measure})$ and DB index is shown in Fig. S2b. The plots in Fig. S2a and S2b indicate that the variation of F-measure with change in rank r is consistent with change in both Silhouette and DB indices for the OV data sets. The similarly varying DB index and $(1-\text{F-measure})$ curves in Fig. S1 and S2b justify that DB index can also be used for the choice of optimal rank.

3.2 Advantage of Averting Row-Normalization on Individual Laplacians

As stated in Section 5.4.4 of the main paper, row-normalization in spectral clustering is advantageous for those cases where the similarity graph can be easily partitioned into component subgraphs. However, in case of real-life data sets, where the clusters in the data set are not well-separated, the similarity graphs tend to be densely connected and row-normalization fails to provide any added advantage. To study this at the level of individual graph Laplacians, the clustering performance of the individual modalities with and without row-normalization is reported in Table 5.4.4. In Table 5.4.4 mDNA_RNrm, RNA_RNrm, miRNA_RNrm, and RPPA_RNrm represents the row-normalized counterparts of mDNA, RNA, miRNA, and RPPA modalities, respectively. The results reported in Table 5.4.4 show that for all four component modalities of the STAD and BRCA data sets, and for mDNA, RNA, and miRNA modalities of LGG data set spectral clustering without row-normalization provides better performance compared to the one with row-normalization, in terms of all five external indices. For the CRC data set, for RNA both the approaches have the same performance, while for mDNA and RPPA averting row-normalization provided better performances for majority of the external indices. Only for the miRNA modality of the CRC data set, row-normalization slightly outperforms the one without it. For the OV data set also, only for the RNA modality, row-normalization gives slightly better performance, while for the other three modalities, namely, mDNA, miRNA, and RPPA avoiding row-normalization gives better performance for majority of the external indices. Summarily, spectral clustering without the row-normalization step gives better clustering performance in 77 cases out of the total 100 cases. Thus, even when performing spectral clustering on the individual graph Laplacians, it is better to avoid the row-normalization step for real-life omics data sets.

3.3 Results on OV Data Set

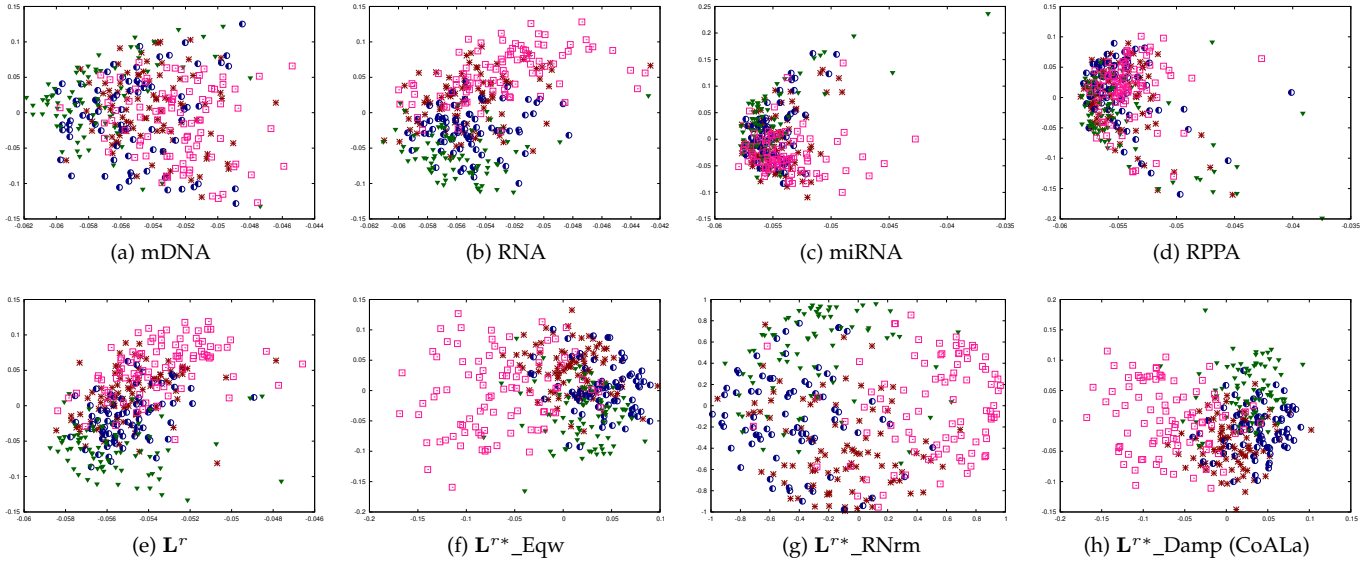
This subsection presents results on the TCGA ovarian carcinoma (OV) data set which could not be provided in the main paper due to space constraints. The OV data set contains 334 samples divided into four molecular subtypes.

TABLE S3
Comparative Performance Analysis of Individual Modalities, Variants of Joint Subspace, and Proposed Algorithm for OV Data Set

Data Set	Index	mDNA	RNA	miRNA	RPPA	$\mathbf{L}^r$	$\mathbf{L}^{r*}$_Eqw	$\mathbf{L}^{r*}$_RNrm	$\mathbf{L}^{r*}$_Damp (CoALa)
OV	F-measure	0.3857003	0.6444234	0.4186585	0.3858236	0.6664847	0.5372622	**0.7029289**	0.6700660
	Purity	0.3922156	0.6497006	0.4131737	0.3712575	0.6616766	0.5688623	**0.7005988**	0.6736527
	Rand	0.6469224	0.7536459	0.6387046	0.6357196	0.7465969	0.6830124	**0.7623312**	0.7379295
	Jaccard	0.1705390	0.3517861	0.1927033	0.1725011	0.3316576	0.2638129	**0.3543549**	0.3303621
	Dice	0.2913855	0.5204760	0.3231370	0.2942447	0.4981124	0.4174873	**0.5232822**	0.4966499

TABLE S4
Comparative Performance Analysis of Proposed and Existing Approaches on OV Data Set

Data Set		Measure	COCA	LRAcluster	JIVE (PERM)	A-JIVE	iCluster	PCA-con	SNF	NormS	CoALa
OV $n = 334$; $k = 4$; $M = 4$		Subspace Rank	-	2	32	64	2	3	3	14	3
	External	F-measure	0.5966656	0.6384046	0.5709916	0.4872798	0.4808256	0.6868295	0.6260039	**0.6910392**	0.670066
		Purity	0.5892215	0.6287425	0.5778443	0.4955089	0.5119760	0.6946108	0.6287425	**0.6976048**	0.6736527
		Rand	0.7002086	0.7322472	0.6910323	0.6852043	0.6916078	0.7734621	0.7164949	**0.7766269**	0.7379295
		Jaccard	0.3126523	0.3157798	0.2614657	0.2451469	0.2568036	0.3880307	0.2961293	**0.3930125**	0.3303621
		Dice	0.4761614	0.4799888	0.4145427	0.3906669	0.4086615	0.5591097	0.4569441	**0.5642627**	0.4966499
	Internal	Silhouette	-	0.3749983	0.3373489	0.3474126	0.4084831	0.3730159	**0.4439673**	0.3705642	0.3658571
		Dunn	-	0.0206385	0.0147134	0.0148862	0.0130433	0.0241898	0.0116611	**0.0401888**	0.0198808
		DB	-	0.8879858	1.0133890	0.8654302	**0.8079155**	0.8902714	0.8397593	0.8931774	0.8662991
		Xie-Beni	-	137.27130	106.51260	209.29040	138.38600	122.30580	448.9055	**41.933410**	88.616350
		Time (in sec)	41.75	33.05	3650.52	570.50	2076.36	**0.86**	2.48	1.72	15.58



(a) mDNA (b) RNA (c) miRNA (d) RPPA

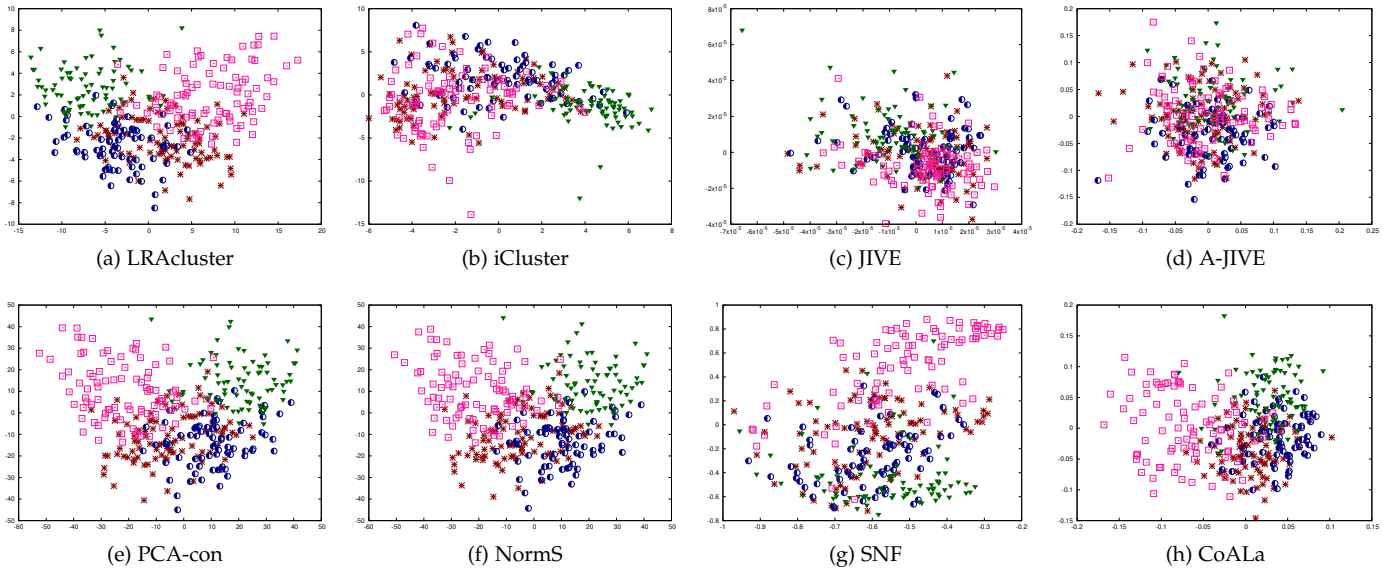
(e) $\mathbf{L}^r$ (f) $\mathbf{L}^{r*}$_Eqw (g) $\mathbf{L}^{r*}$_RNrm (h) $\mathbf{L}^{r*}$_Damp (CoALa)

Fig. S3. Scatter plots using first two components of individual Laplacians, different variants of joint subspace, and the proposed CoALa algorithm for OV data set

3.3.1 *Comparison with Individual and Joint Laplacians*

Table S3 compares the performance of the proposed CoALa algorithm with that of the individual Laplacians and different variants of the joint subspace like the full-rank, equally weighted, and the row-normalized ones. The four individual modalities are mDNA, RNA, miRNA, and RPPA. The full-rank subspace is denoted by $\mathbf{L}^r$ which is formed by the convex combination of all the eigenpairs of each individual Laplacian. The subspace $\mathbf{L}^{r*}$_Eqw denotes the approximate subspace formed by the equally weighted combination of r most informative (largest) eigenpairs of each Laplacian, while $\mathbf{L}^{r*}$_Damp denotes the one formed by the proposed damped weighted combination introduced in Section 3.5 of the main paper. The subspace $\mathbf{L}^{r*}$_Damp corresponds to the proposed CoALa algorithm and $\mathbf{L}^{r*}$_RNrm denotes the row-normalized variant of the proposed $\mathbf{L}^{r*}$_Damp subspace.

The results reported in Table S3 show that the proposed algorithm has outperformed all four individual modalities. Thus integration of information from multiple modalities preserves better cluster structure compared to unimodal analysis. Among the individual modalities, RNA has the best performance followed by miRNA. The two remaining modalities, mDNA and RPPA have close enough performances, while there is a significant difference between the performances of the most relevant (RNA) and the second most relevant one (miRNA). The scatter plots using the two largest eigenvectors of for the shifted normalized graph Laplacians of mDNA, RNA, miRNA, and RPPA are given in Fig. S3a, S3b, S3c, and S3d, respectively. Other than the



(a) LRAcluster (b) iCluster (c) JIVE (d) A-JIVE

(e) PCA-con (f) NormS (g) SNF (h) CoALa

Fig. S4. Scatter plots for first two components of different low-rank based approaches for OV data set

scatter plot for RNA in Fig. S3b, objects from different subtypes are nearly inseparable from each other in Fig. S3a, S3c, and S3d. This is also evident from the poor performance of mDNA, miRNA, and RPPA in Table S3 across all five external indices. Table S3 also presents the comparative clustering performance of different variants of the joint subspace. Table S3 shows that the proposed approximate subspace with relevance based damped weighting, $\mathbf{L}^{r*}$_Damp outperforms the full-rank subspace, $\mathbf{L}^{r}$, as well as the equally weighted one, $\mathbf{L}^{r*}$_Eqw. The two-dimensional scatter plots of $\mathbf{L}^{r}$, $\mathbf{L}^{r*}$_Eqw, $\mathbf{L}^{r*}$_RNrm, and the proposed $\mathbf{L}^{r*}$_Damp subspaces are provided in Fig. S3e, S3f, S3g, and S4h, respectively. The plots in Fig. S3e-S4h show that the objects in $\mathbf{L}^{r}$, $\mathbf{L}^{r*}$_Eqw, and $\mathbf{L}^{r*}$_Damp subspaces lack inter-cluster separability. However, in the row-normalized approximate subspace $\mathbf{L}^{r*}$_RNrm, since row-normalization shifts the objects from different subtypes away from the origin into different directions, so the objects have higher inter-cluster separability as compared to the other three subspaces. This is also evident from the best performance of $\mathbf{L}^{r*}$_RNrm in Table S3 compared to all other subspaces.

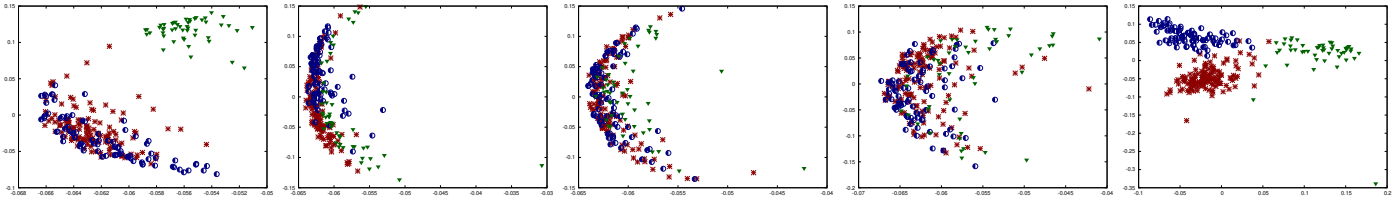
3.3.2 Comparison with Existing Approaches

For the OV data set, the performance of the proposed algorithm is compared with that of eight existing integrative clustering approaches, namely, cluster of cluster analysis (COCA) [16], LRAcluster [17], joint and individual variance explained (JIVE) [18], angle-based JIVE (A-JIVE) [19], iCluster [20], principal component analysis (PCA) on the concatenated data (PCA-con) [21], similarity network fusion (SNF) [22], and normality based low rank subspace (termed as NormS) [23]. The comparative results are reported in Table S4. The results in Table S4 show that the NormS algorithm has the best performance among all the approaches, while the proposed CoALa algorithm has the third best performance after PCA-con, in terms of the external indices. In terms of the internal indices, NormS has the best performance for Dunn and Dunn Xie-Beni indices, while iCluster and SNF have the best performance for DB and Silhouette indices, respectively. The two-dimensional scatter plots for the existing and the proposed approach are given in Fig. S4. The plots for PCA-con and NormS in Fig. S4e and S4f, respectively, are close to each other which is also evident from their external evaluation results in Table S4.

3.4 Scatter Plot Analysis

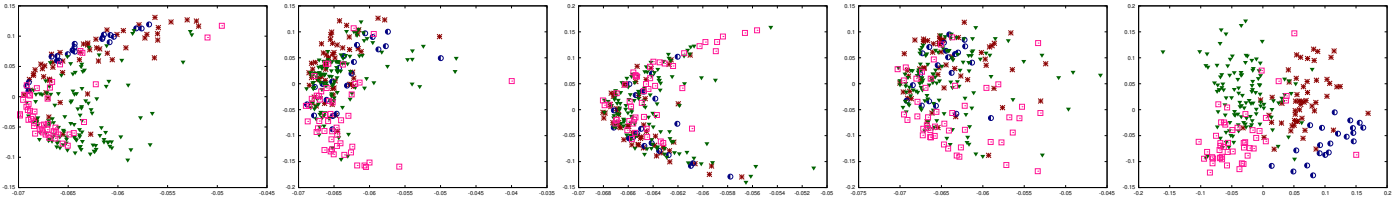
In this subsection the scatter plots for the first two dimensions of the individual modalities and the existing low-rank based approaches are compared with those of the proposed approach for CRC, LGG, STAD, and BRCA data sets. Most of the scatter plots for LGG and STAD data sets are presented in the main paper in Fig. 3 and 4, respectively. This subsection analyzes additional scatter plots for the omics data sets. Two-dimensional scatter plots provide an interesting way for visual and intuitive analysis of the cluster structure reflected in different subspaces. The comparison of the proposed subspace is first made with the individual subspaces and several variants of the joint subspace, followed by those of existing integrative clustering approaches.

Two-dimensional scatter plots of the individual modalities are compared with that of the proposed approach, CoALa, in Fig. S5, S6, S7, and S8 for LGG, STAD, BRCA, and CRC data sets, respectively. For the LGG data set, Fig. S5 shows that only for mDNA in Fig. S5a, one of the clusters marked in green is well-separated from the other two, while for the other modalities, Fig. S5b, S5c, and S5d show that all objects from all three established subtypes are projected close to each other exhibiting poor separability. On the contrary, Fig. S5e shows that all three subtypes of LGG data set are compact and well separated in the scatter plot for the first two dimensions of the proposed subspace. For the BRCA data set, Fig. S7 shows that among the individual modalities, the four previously established TCGA subtypes are best reflected in RNA. For RNA, Fig. S7b shows that two clusters marked in blue and brown are



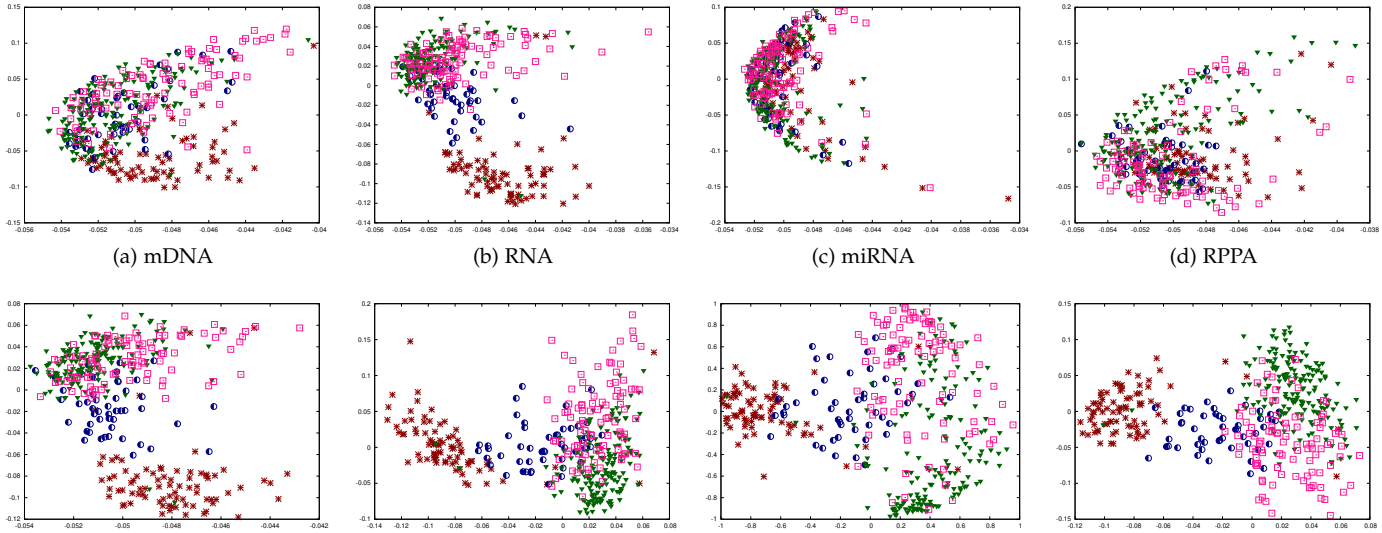
(a) mDNA (b) RNA (c) miRNA (d) RPPA (e) CoALa

Fig. S5. Scatter plots for first two components of individual Laplacians and CoALa algorithm for LGG data set



(a) mDNA (b) RNA (c) miRNA (d) RPPA (e) CoALa

Fig. S6. Scatter plots for first two components of individual Laplacians and CoALa algorithm for STAD data set



(a) mDNA (b) RNA (c) miRNA (d) RPPA

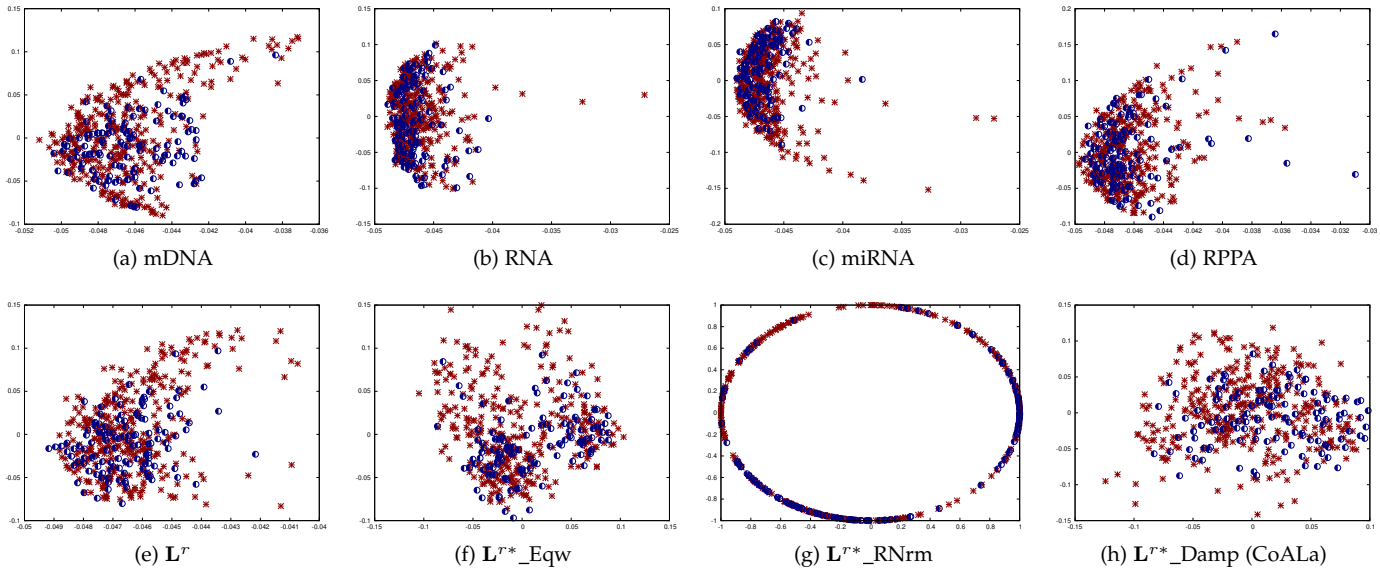
(e) $\mathbf{L}^r$ (f) $\mathbf{L}^{r*}$_Eqw (g) $\mathbf{L}^{r*}$_RNrm (h) $\mathbf{L}^{r*}$_Damp (CoALa)

Fig. S7. Scatter plots using first two components of individual Laplacians, different variants of joint subspace, and the proposed CoALa algorithm for BRCA data set

well separated in the projected two-dimensional subspace, while the other two clusters marked in pink and green have poor separability. On the other hand, Fig. S9h shows that in the proposed subspace, the clusters marked in blue and brown continue to remain well-separated and the one marked in brown is more compact in the proposed subspace compared to its projection in RNA in Fig. S7b. The scatter plots for different variants of the joint subspace, like, the full-rank ($\mathbf{L}^r$), the equally-weighted ($\mathbf{L}^{r*}$_Eqw), and the row-normalized ($\mathbf{L}^{r*}$_RNrm) one are provided in Fig. S7e, S7f, and S7g, respectively, for BRCA data set, and in Fig. S8e, S8f, and S8g, respectively, for CRC data set. Among these three variants, for the BRCA data set, objects in the $\mathbf{L}^{r*}$_RNrm subspace (Fig. S7g) show maximum inter-cluster separation compared to the other two, because of the row-normalization step. The scatter plots the existing low-rank based approaches along with the proposed algorithm are provided in Fig. S9 and S10 for BRCA and CRC data sets. Most of the scatter plots for LGG and STAD data sets are presented in the main paper in Fig. 3 and 4, respectively, those for some of the remaining approaches are provided in Fig. S11.

4 Results on Benchmark Data Sets

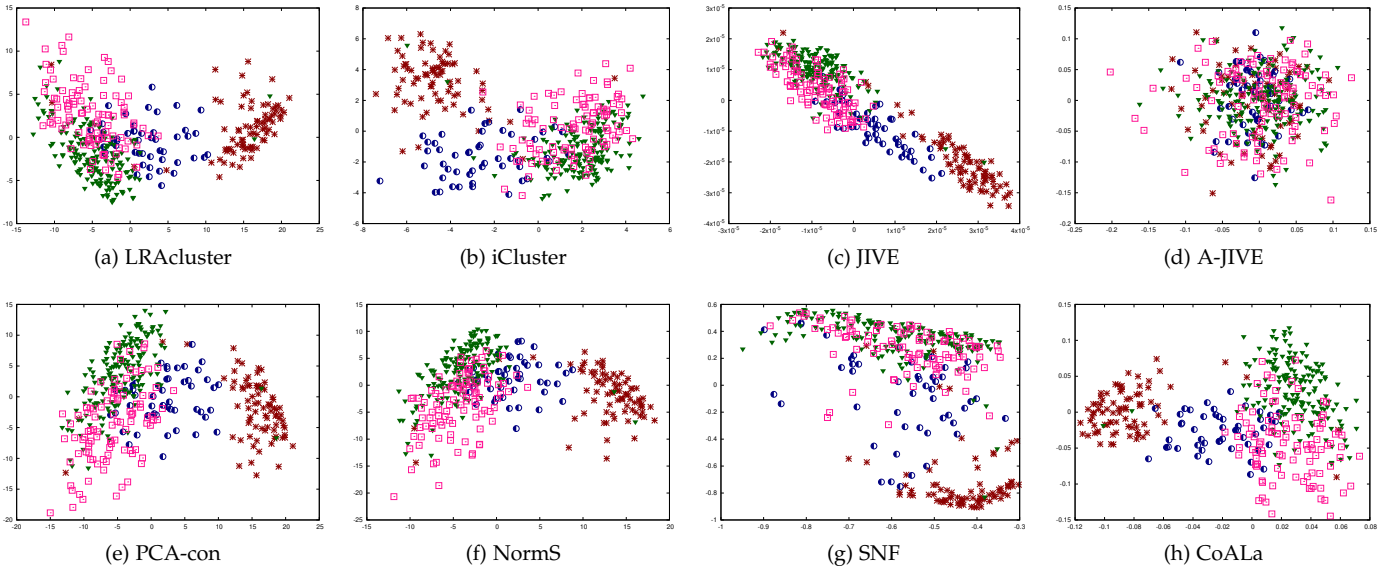
This section presents additional results on the four benchmark multimodal data sets, namely, Football, Politics-uk, Rugby, and Digits used in the main paper to establish the generality of the proposed approach. The scatter plots for the first dimensions of different low-rank subspaces of Politics-uk and Digits data set are given in Fig. S14 and S15, respectively.



(a) mDNA (b) RNA (c) miRNA (d) RPPA

(e) $\mathbf{L}^r$ (f) $\mathbf{L}^{r*}$_Eqw (g) $\mathbf{L}^{r*}$_RNrm (h) $\mathbf{L}^{r*}$_Damp (CoALa)

Fig. S8. Scatter plots for first two components of individual Laplacians, different variants of joint subspace, and the proposed CoALa algorithm for CRC data set



(a) LRAcluster (b) iCluster (c) JIVE (d) A-JIVE

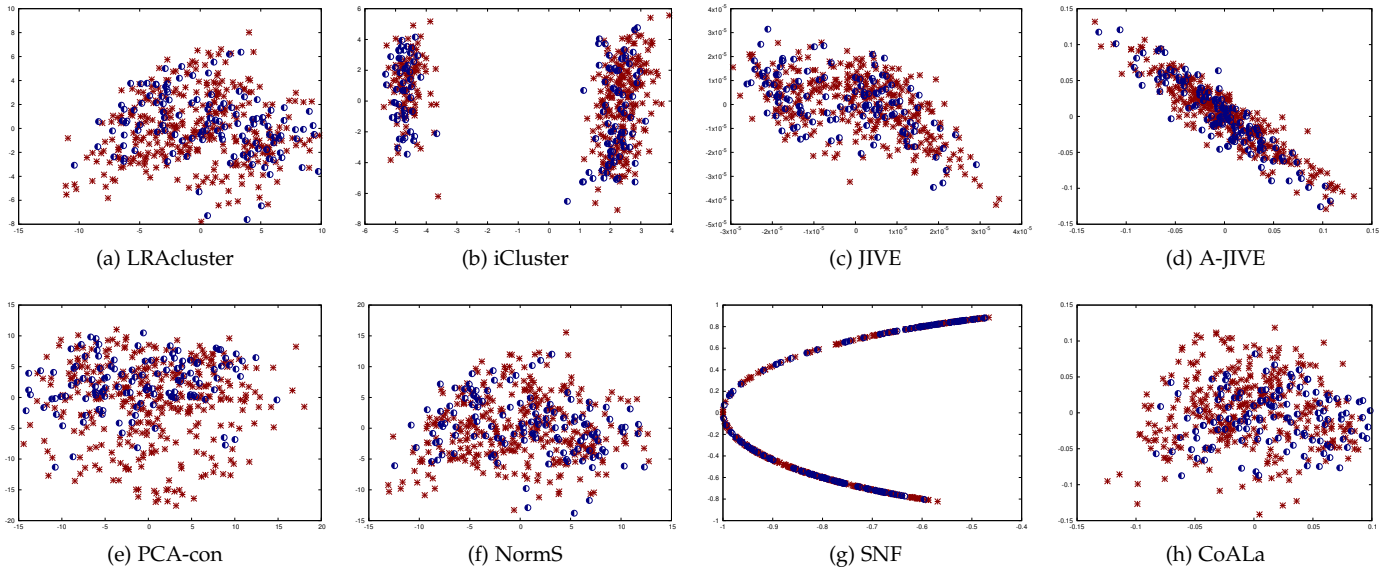
(e) PCA-con (f) NormS (g) SNF (h) CoALa

Fig. S9. Scatter plots for first two components of different low-rank based approaches for BRCA data set

4.1 Estimation of Optimal Rank

Similar to the multi-omics data sets, the rank r of the individual Laplacians of the benchmark data sets is selected using Silhouette index, as described in Section 5.2 of the main paper. For each multimodal data set having M modalities and k clusters, the proposed algorithm selects r eigenpairs from each of the M individual Laplacian and constructs a joint eigenspace of rank rM. Since the proposed algorithm used spectral clustering [24], so clustering is performed on the k largest eigenvectors of the final eigenspace. Therefore, the rank r of the individual Laplacians should be $r \geq \lceil k/M \rceil$. The value of rank r is varied from $\lceil k/M \rceil$ to 50 and for each value of rank r, the Silhouette index $\mathcal{S}(r)$ is evaluated for clustering on the k largest eigenvectors of the final eigenspace. The optimal value of rank is selected to be the one which maximizes the value of Silhouette index over different values of r. Fig. S12 shows the variation in Silhouette index as well as F-measure with the increase in rank r for different benchmark data sets. Fig. S12 shows that the curves for Silhouette index and F-measure vary in a similar fashion over the entire range of r values for the benchmark data sets. Based on the Silhouette index, the optimal ranks selected for Football, Politics-uk, Rugby, and Digits data sets are 22, 45, 7, and 6, respectively. For Football, Politics-uk, and Digits data sets, the value of F-measure corresponding to the rank selected using Silhouette index, coincides with the maximum F-measure obtained over different values of rank r.

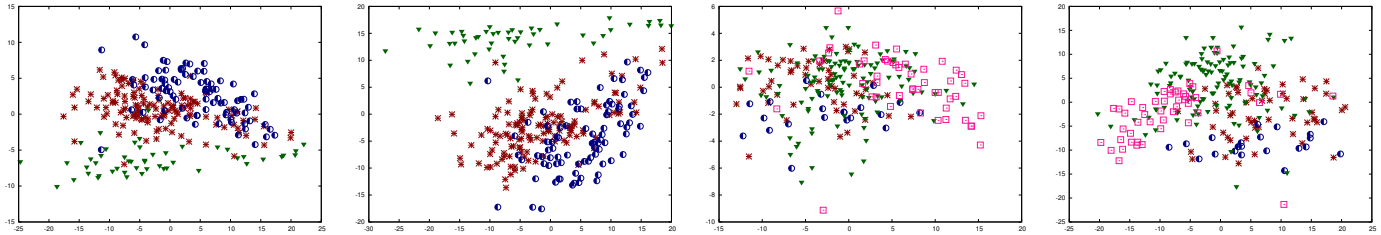
The rank parameter can also be tuned using the DB index. Similar to Fig. S1, the variation of $(1-\text{F-measure})$ and DB index with increase in rank r is shown in Fig. S13 for



(a) LRAcluster (b) iCluster (c) JIVE (d) A-JIVE

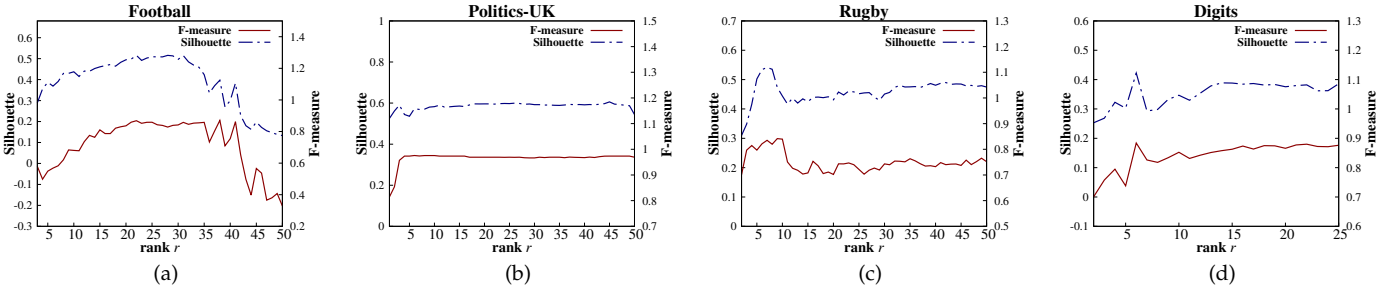
(e) PCA-con (f) NormS (g) SNF (h) CoALa

Fig. S10. Scatter plots for first two components of different low-rank based approaches for CRC data set



(a) LRAcluster for LGG (b) PCA-con for LGG (c) LRAcluster for STAD (d) PCA-con for STAD

Fig. S11. Scatter plots using first two components of some low-rank based approaches for LGG and STAD data sets.



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

Fig. S12. Variation of Silhouette index and F-measure for different values of rank parameter r on benchmark data sets.

different benchmark data sets. Since, $(1-\text{F-measure})$ and DB index are found to vary similarly, so, optimal rank selected using DB index is most likely to also maximize F-measure over different values of r.

4.2 Importance of Multimodal Integration

The three Twitter data sets, namely, Football, Politics-uk, and Rugby have nine different modalities, while the image data set, Digits has six. This subsection shows that integration of information from multiple modalities has huge advantage over unimodal analysis. Table S5 and S6 compares the performance of clustering on the k largest eigenvectors of the individual shifted Laplacians with that of the proposed approximate subspace for the Twitter and the Digits data set, respectively. From the results of Table S5 and S6, it is evident that the proposed CoALa algorithm consistently and significantly outperforms all the individual modalites across all four benchmark data sets. Amongst the Twitter data sets, Table S5 shows that the information about the set of users who follow a particular user or the incoming links (followed-by modality) to the users gives better performance compared to the other modalities for Football and Politics-uk. On the other hand, for the Rugby data set, the set of profiles that a user follows or outgoing links (follows) gives better performance than other modalities on majority of external indices. Results from

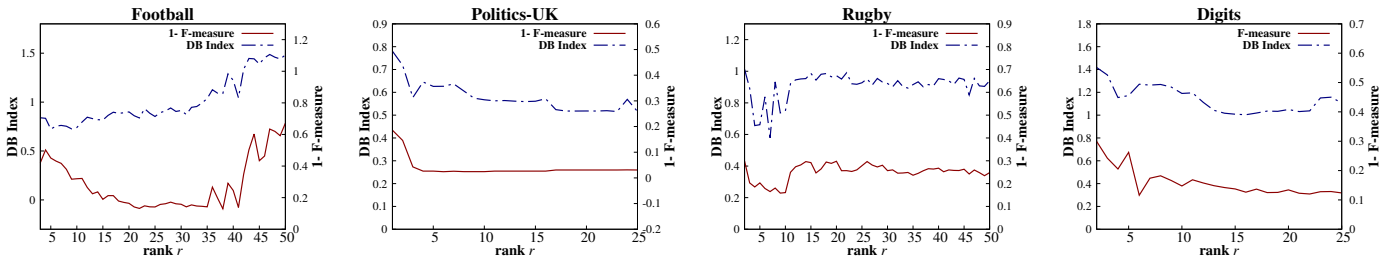


Fig. S13. Variation of DB index and $(1-\text{F-measure})$ for different values of rank parameter r on benchmark data sets.

TABLE S5
Comparative Performance Analysis of Individual Modalities and Proposed Approach On Twitter Data Sets

Index		Followed-By	Follows	Mentioned-By	Mentions	Retweeted-By	Retweets	Tweets500	ListMerged500	Lists500	CoALa
F-measure	Football	0.7747023	0.7042013	0.7241344	0.7109046	0.5537196	0.5202768	0.2022110	0.7232265	0.6606393	**0.8683491**
Purity		0.7282258	0.6766129	0.7362903	0.7092741	0.5447580	0.5008064	0.2072580	0.6931451	0.6399193	**0.8584677**
Rand		0.9472965	0.9197825	0.9356405	0.9384256	0.8593378	0.7958926	0.7691328	0.9322776	0.9147218	**0.9739682**
Jaccard		0.3963918	0.2587792	0.3440107	0.3650386	0.1671624	0.1140070	0.0507533	0.3313234	0.2555575	**0.6005824**
Dice		0.5667814	0.4109037	0.5114635	0.5344279	0.2860945	0.2046035	0.0966025	0.4968149	0.4069673	**0.7504383**
F-measure	Politics-uk	0.9175316	0.8836935	0.8660595	0.7619363	0.8346957	0.7991772	0.5804394	0.8635673	0.8464556	**0.9736129**
Purity		0.9713604	0.9021479	0.8778042	0.7823389	0.8477326	0.8138425	0.6658711	0.9021480	0.8782816	**0.9785203**
Rand		0.9196880	0.8728323	0.8429422	0.7114181	0.7991423	0.7510534	0.6330178	0.8562195	0.8346941	**0.9826084**
Jaccard		0.8019766	0.7251893	0.6719116	0.5161645	0.6134922	0.5532727	0.2923704	0.6507932	0.6018154	**0.9559279**
Dice		0.8901077	0.8400708	0.8037591	0.6800488	0.7601752	0.7121101	0.4524560	0.7884612	0.7514167	**0.9774674**
F-measure	Rugby	0.7113898	0.6643790	0.6873041	0.6705410	0.7078636	0.6856623	0.3737361	0.3460789	0.7426962	**0.8349647**
Purity		0.8474238	0.8435597	0.8274004	0.8121780	0.7915691	0.7816159	0.4871194	0.4566745	0.7796253	**0.8606557**
Rand		0.8609769	0.8580120	0.8562299	0.8482375	0.8560331	0.8406967	0.7177268	0.5223523	0.8672685	**0.9067597**
Jaccard		0.3294506	0.3033581	0.3222504	0.3115874	0.4473977	0.4013710	0.1605544	0.1501404	0.4447761	**0.5982183**
Dice		0.4948458	0.4655016	0.4871073	0.4737759	0.6181705	0.5724485	0.2766858	0.2610802	0.6155136	**0.7486065**

TABLE S6
Comparative Performance Analysis of Individual Modalities and Proposed Approach On Digits Data Sets

		fac	fou	kar	mor	pix	zer	CoALa
F-measure	Digits	0.6451628	0.7209662	0.7022988	0.5651531	0.6829546	0.5545294	**0.8839913**
Purity		0.6223000	0.7100000	0.7027000	0.5414000	0.6890000	0.5350500	**0.8835000**
Rand		0.8994301	0.9173923	0.9156842	0.8655854	0.9108559	0.8757654	**0.9576618**
Jaccard		0.3595262	0.4163257	0.4134869	0.2871402	0.3999145	0.2481882	**0.6502019**
Dice		0.5288984	0.5878948	0.5850591	0.2402663	0.5713413	0.3976774	**0.7880271**

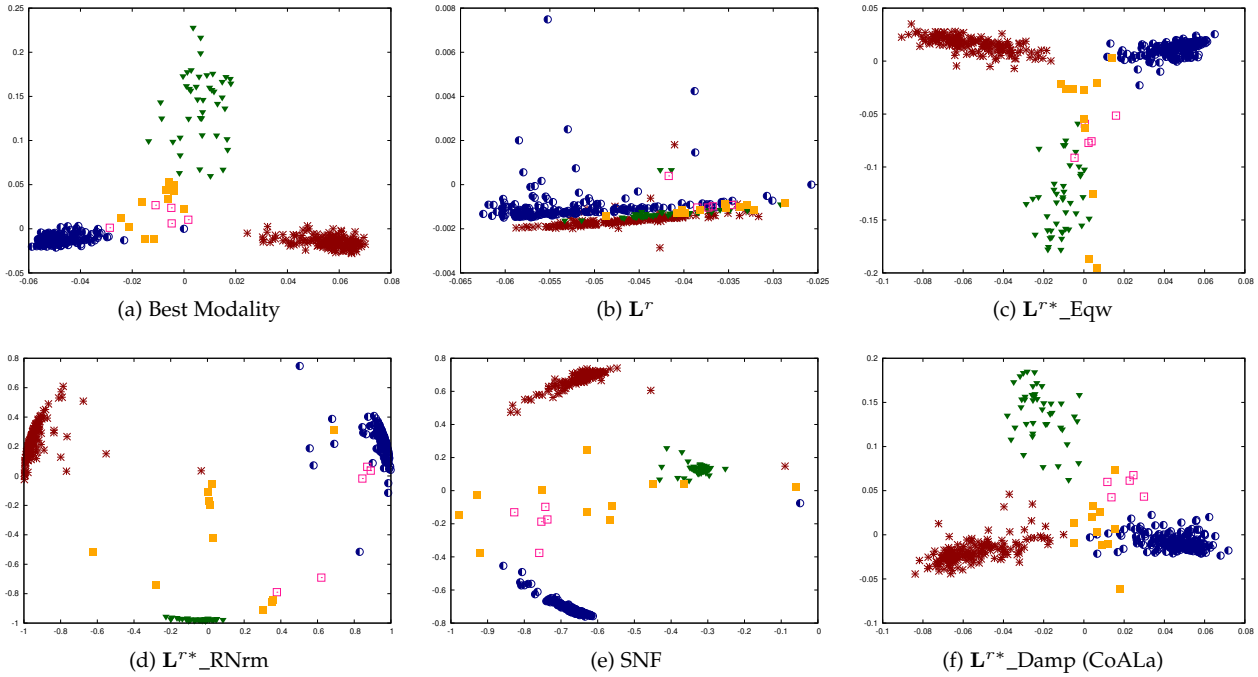
three Twitter data sets imply that 'follows' and 'followed-by' are important relationships for identification of communities in social networks. Moreover, paired modalities like 'follows' and 'followed-by', 'mentions' and 'mentioned-by', 'retweets' and 'retweeted-by' have close performances across all the Twitter data sets. For the Digits data set, Table S6 shows that all six component modalities have significantly lower than that of the proposed approach. The 'fou' modality consisting of 76 Fourier coefficients of the character shapes of the images has the best performance amongst the component modalities. The 64 Karhunen-Love coefficients computed in 'kar' modality has performance close to those based on Fourier coefficients. Summarily, for all the benchmark data sets, integration of multiple modalities always beats the performance of individual Laplacians by a wide margin.

4.3 Choice of Weight Parameter α

The weight parameter $\boldsymbol{\alpha}$ determines the influence of the individual modalities during data integration. Section 3.5 of the main article introduces a relevance based damping strategy for choice of $\boldsymbol{\alpha}$. This damped weighing referred to as $\mathbf{L}^{r*}$_Damp is compared with $\mathbf{L}^{r*}$_Eqw, where all the component modalities are equally weighted. The comparative results for the benchmark data sets are reported in Table S7. It can be observed in Section 3.5 of the main article that for a majority of omics data sets, damped weighting of modalities based on relevance outperforms the equally weighted one. On the contrary, the results in Table S7 shows that the equally weighted strategy gives better performance that the damped one on the Twitter based Football and Politics-uk data sets. One possible explanation is that most of the component modalities of the Twitter data sets are similar to each other and have close performances. For instance, 'follows' and 'followed-by', both are network based modalities where 'follows' captures the outgoing links from the nodes, while 'followed-by' captures the incoming links to the nodes. Other pairs of modalities like 'mentions' and 'mentioned-by', and 'retweets' and 'retweeted-by' are also very similar to each other. In the damped weighting introduced in Section 3.5, slight differences in the relevance values of these similar modalities would dampen the effect of the one with lower relevance by a factor of β. This leads to degraded cluster structure in eigenspace of the joint

TABLE S7
Comparative Performance Analysis of Equally and Damped Weighted Combination on Benchmark Data Sets

Index		$\mathbf{L}^{r*}$_Eqw	$\mathbf{L}^{r*}$_Damp		$\mathbf{L}^{r*}$_Eqw	$\mathbf{L}^{r*}$_Damp		$\mathbf{L}^{r*}$_Eqw	$\mathbf{L}^{r*}$_Damp		$\mathbf{L}^{r*}$_Eqw	$\mathbf{L}^{r*}$_Damp
F-measure	Football	**0.8848290**	0.8683491	Politics-uk	0.9735519	**0.9736129**	Rugby	0.8288040	**0.8349647**	Digits	0.8746977	**0.8839913**
Purity		**0.8778225**	0.8584677		**0.9785203**	**0.9785203**		**0.8621780**	0.8606557		0.8310000	**0.8835000**
Rand		**0.9760741**	0.9739682		**0.9828368**	0.9826084		0.8972515	**0.9067597**		0.9564677	**0.9576618**
Jaccard		**0.6328100**	0.6005824		**0.9564689**	0.9559279		0.5487269	**0.5982183**		**0.6571682**	0.6502019
Dice		**0.7747074**	0.7504383		**0.9777501**	0.9774674		0.7086141	**0.7486065**		**0.7931219**	0.7880271



(a) Best Modality (b) $\mathbf{L}^{r}$ (c) $\mathbf{L}^{r*}$_Eqw

(d) $\mathbf{L}^{r*}$_RNrm (e) SNF (f) $\mathbf{L}^{r*}$_Damp (CoALa)

Fig. S14. Scatter plots using first two components of low-rank based subspaces for Politicsuk data set

TABLE S8
Effect of Row-Normalization on Benchmark Data Sets

Index		$\mathbf{L}^{r*}$_RNrm	CoALa		$\mathbf{L}^{r*}$_RNrm	CoALa		$\mathbf{L}^{r*}$_RNrm	CoALa		$\mathbf{L}^{r*}$_RNrm	CoALa
F-measure	Football	0.8679092	**0.8683491**	Politics-uk	0.9571452	**0.9736129**	Rugby	0.6737320	**0.8349647**	Digits	0.8629902	**0.8839913**
Purity		**0.8911290**	0.8584677		0.9715990	**0.9785203**		0.8545667	**0.8606557**		0.8565000	**0.8835000**
Rand		**0.9785490**	0.9739682		0.9746971	**0.9826084**		0.8606810	**0.9067597**		**0.9629410**	0.9576618
Jaccard		**0.6403940**	0.6005824		0.9356337	**0.9559279**		0.2977746	**0.5982183**		**0.6872517**	0.6502019
Dice		**0.7807807**	0.7504383		0.9667067	**0.9774674**		0.4588952	**0.7486065**		**0.8146404**	0.7880271

Laplacian for two Twitter data sets when using the damped weighted strategy. For Rugby and Digits data sets, damped weighted strategy $\mathbf{L}^{r*}$_Damp has better performance compared to equally weight $\mathbf{L}^{r*}$_Eqw one for majority of the external indices.

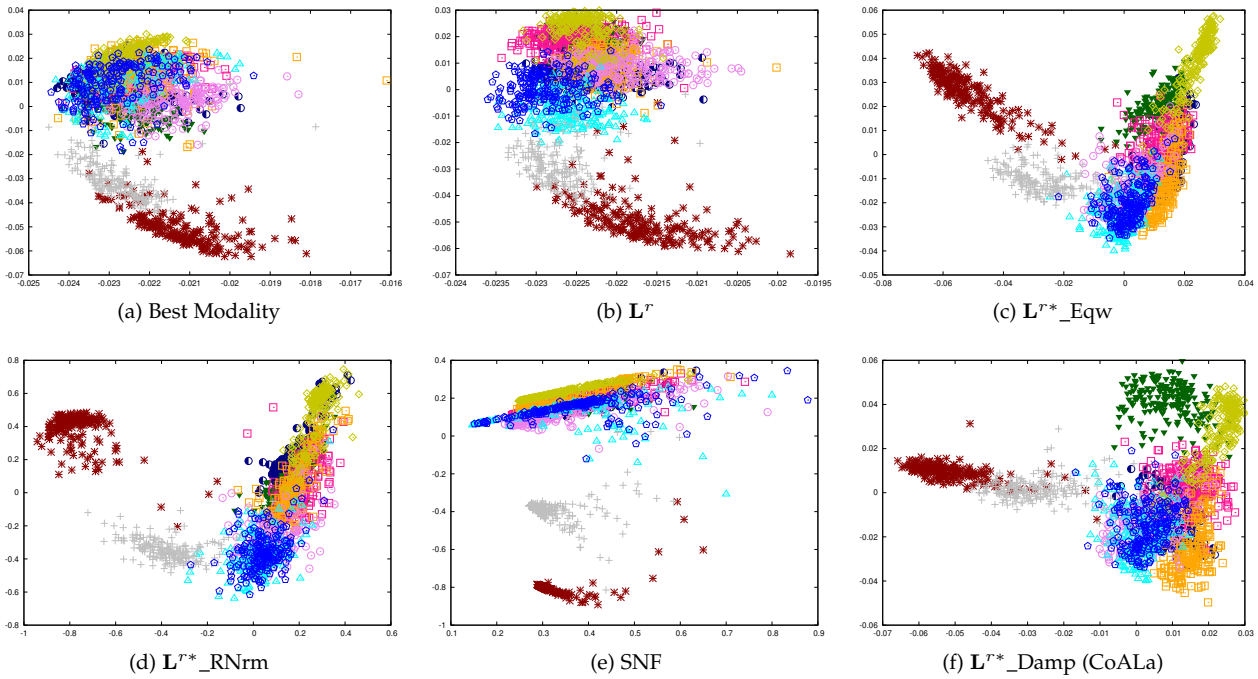
4.4 Effect of Row-Normalization on Benchmark Data Sets

Table S8 compares the performance of the proposed approximate subspace with and without the row-normalization step. In Table S8, the subspace $\mathbf{L}^{r*}$_RNrm corresponds to the row-normalized one. Table S8 shows that for Politics-uk and Rugby data sets, avoiding row-normalization gives better performance for all the external indices. On the other hand, for Football and Digits data set, majority of the external indices gives better performance with row-nomarlization. Scatter plots for the first dimensions of $\mathbf{L}^{r*}$_RNrm and the proposed CoALa algorithm are given in Fig. S14 and S15 for the Politics-uk and the Digits data set, respectively.

5 Experimental Setup for Existing Algorithms

The performance of the proposed CoALa algorithm is compared with six existing integrative clustering based approaches, namely, cluster of cluster analysis (COCA) [16], LRAcluster [25], joint and individual variance explained (JIVE) [26], iCluster [27], principal component analysis (PCA) on concatenated data (PCA-con) [28], and similarity network fusion (SNF) [22]. The experimental setup used for these algorithms is briefly outlined as follows:

1) **COCA** [16]: This is a consensus clustering based approach which first cluster each modality separately and the individual clustering solutions are



(a) Best Modality (b) $\mathbf{L}^r$ (c) $\mathbf{L}^{r*}$_Eqw

(d) $\mathbf{L}^{r*}$_RNrm (e) SNF (f) $\mathbf{L}^{r*}$_Damp (CoALa)

Fig. S15. Scatter plots using first two components of low-rank based subspaces for Digits data set

then combined to get the final cluster assignments. For the COCA approach, k-means clustering is first performed on each modality separately with k clusters. Clusters identified from each modality are encoded into a series of indicator variables for each cluster. Consensus clustering [29] is then performed on the indicator matrix of 0's and 1's using ConsensusClusterPlus R package [30] version 1.40.0. Parameters used for consensus clustering are 80% sample resampling with 1000 iterations of hierarchical clustering based on a Pearson correlation distance metric, as suggested in [16]. The COCA algorithm uses re-sampling based technique to find the clusters, so its performance varies on different executions of the algorithm. The average performance of the COCA algorithm over 10 executions is reported in this work.

2) **LRAcluster** [25]: The low-rank based approach which models each modality of a multimodal data set using a separate probability distribution having its own set of parameters. In this work, four omic modalities are considered for each cancer data set. For Gene and miRNA modalities, sequence based count data are considered, while for DNA and Protein modalities, array based expression data is considered. Therefore, as suggested by the authors of this algorithm, the count based Gene and miRNA modalities are modelled using Poisson distribution, while array based DNA and Protein modalities are modelled using Gaussian distribution [25]. The rank of the lower dimensional subspace is optimized using the likelihood based "explained variation" criteria [25], as suggested by the authors. According to this criteria, the value of explained variance is observed for different values of rank varying between 0 to 10. The optimal value of rank is chosen to be the one having the maximum change in explained variance. The change in explained variance for different values of rank is given in Fig. S16 for different data sets. Based on this criteria, the optimal rank obtained for the CRC, LGG, STAD, BRCA, and OV data sets are 3, 2, 1, 2, and 2, respectively. After obtaining the optimal low-rank subspace, k-means clustering is performed in that subspace to identify the clusters.

3) **JIVE (PERM)** [26] and **A-JIVE** [19]: The JIVE (PERM) and A-JIVE algorithms extracts two low-rank representations for each modality, one encodes the shared joint structure, while the other encodes modality specific structure. The ranks of the joint and the individual structures are automatically determined using a permutation (PERM) test based approach for the JIVE algorithm. After obtaining the joint rank, say j, and the joint and individual structures for each modality, the overall joint structure of all the modalities is obtained by concatenating the j largest principal components of the joint structure from each of the modalities. Then k-means is performed on the concatenated joint structure. The joint rank selected by the A-JIVE algorithm for the CRC, LGG, STAD, OV, and BRCA data sets are 32, 48, 64, and 64, respectively, while the individial ranks for all these data sets are selected to be zero. The joint and individual ranks obtained by the JIVE algorithm using the permutation based rank selection criterion are given in Table S9 for different omics data sets.

4) **iCluster** [27]: This is a low-rank based approach

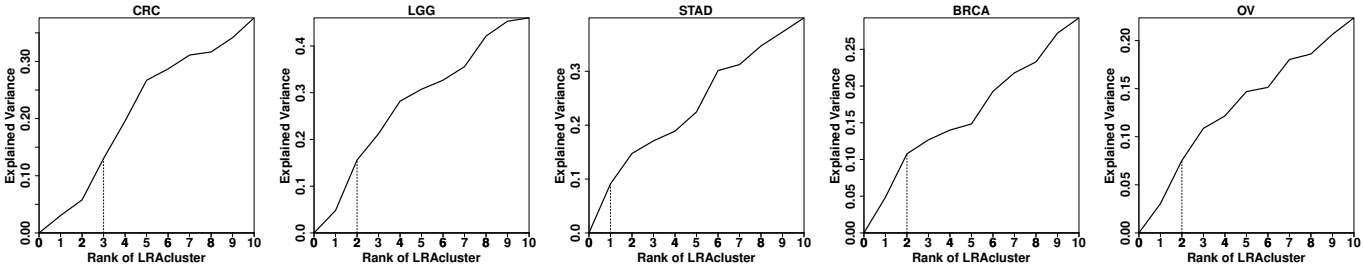


Fig. S16. Optimal rank estimation of LRAcluster for different omics data sets

TABLE S9
Joint and Individual Ranks Estimated by JIVE (PERM) Algorithms

Different Datasets	Joint Rank	Individual Ranks			
		mDNA	RNA	miRNA	RPPA
CRC	16	21	31	23	10
LGG	8	12	23	18	12
STAD	8	15	21	15	8
BRCA	12	30	36	15	15
OV	32	34	50	33	23

TABLE S10
Penalty Parameter λ Selected by iCluster Algorithm

Different Datasets	Rank of Subspace	Penalty λ			
		mDNA	RNA	miRNA	RPPA
CRC	1	0.5488599	0.1188925	0.0602605	0.5977198
LGG	2	0.5228013	0.0244299	0.0928338	0.9657980
STAD	3	0.9201954	0.7149837	0.0960912	0.1026058
BRCA	3	0.3338762	0.0895765	0.8289902	0.8843648
OV	3	0.9332247	0.2622149	0.0798045	0.4185667

which uses Gaussian latent variable model to extract a $(k-1)$ dimensional joint subspace of a multimodal data set, where k is the number of clusters in the data set. The k-means clustering is performed in the $(k-1)$ dimensional joint subspace extracted by the iCluster algorithm Hence, the dimensions of low-rank subspaces extracted by iCluster for CRC, LGG, STAD, BRCA, and OV data sets are 1, 2, 3, 3, and 3, respectively. The iCluster R-package available at https://CRAN.R-project.org/package=iCluster is used to evaluate the performance of the iCluster algorithm. For each modality, iCluster has a lasso penalty parameter (λ), which varies between 0 and 1. The value 0 represents the non-sparse solution where all features are selected, while 1 represents the null model where no features are included. The optimal value of λ is selected using the proportion of deviance (POD) statistic [20]. The POD statistic lies between 0 and 1. Small values of POD indicate strong cluster separability, and large values of POD indicate poor cluster separability. The value of λ that minimizes the POD statistic is selected to be the optimal one. The uniform sampling design (UD) approach of Fang and Wang [31] is used to generate different combination of λ values that are scattered uniformly across the search domain as suggested in [32]. The penalty parameter λ selected for the individual modalities of the multi-omics data sets are provided in Table S10.

5) **PCA-con** [28]: In the PCA-con approach, genomic features from all the available modalities are concatenated and then PCA is performed on the concatenated data to extract the principal subspace. For a comparative study, the number of principal components considered for PCA-con approach is same as the dimension of the joint subspace extracted by the proposed approach, that is, the number of clusters k. For all the low-rank based approaches, namely, LRAcluster, JIVE, A-JIVE, iCluster, PCA-con, NormS, and the proposed approach, k-means clustering is performed 30 times and the cluster solution corresponding to the minimum objective function is used for comparative analysis.

6) **SNF** [22]: This is graph-theoretic approach which constructs a fused network from similarity networks corresponding to individual modalities. On every iteration of the SNF algorithm the weights of the fused network are updated so as to make the fused network more similar to each of the individual modalities. Finally, normalized spectral clustering [33] is performed on the final fused network to obtain the clusters. The SNFtool package available at https://cran.r-project.org/web/packages/SNFtool/index.html is used for study the performance of the SNF algorithm. The algorithm is evaluated at the default parameter setting.

REFERENCES

[1] G. W. Stewart and J.-g. Sun, *Matrix perturbation theory*. Academic press New York, 1990.

[2] C. Davis and W. Kahan, "The rotation of eigenvectors by a perturbation. III," *SIAM Journal on Numerical Analysis*, vol. 7, no. 1, pp. 1–46, 1970.

[3] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[4] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray. (2013) GLOBOCAN 2012 v1.0, cancer incidence and mortality worldwide: IARC cancerbase no. 11. Accessed on January 15, 2014. [Online]. Available: http://globocan.iarc.fr

[5] TCGA Research Network, "Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas," *The New England Journal of Medicine*, vol. 372, no. 26, pp. 2481–2498, 2015.

[6] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012," *Int. J. Cancer*, vol. 136, no. 5, pp. E359–386, Mar 2015.

[7] TCGA Research Network, "Comprehensive molecular characterization of gastric adenocarcinoma," *Nature*, vol. 513, no. 7517, pp. 202–209, 2014.
[8] TCGA Network, "Comprehensive molecular portraits of human breast tumours," *Nature*, vol. 490, no. 7418, pp. 61–70, Oct 2012.
[9] Z. Hu *et al.*, "The molecular portraits of breast tumors are conserved across microarray platforms," *BMC Genomics*, vol. 7, p. 96, Apr 2006.
[10] T. Sorlie *et al.*, "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 19, pp. 10 869–10 874, Sep 2001.
[11] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin*, Sep 2018.
[12] TCGA Research Network, "Integrated genomic analyses of ovarian carcinoma," *Nature*, vol. 474, no. 7353, pp. 609–615, Jun 2011.
[13] I. Zwiener, B. Frisch, and H. Binder, "Transforming rna-seq data to improve the performance of prognostic gene signatures," *PloS one*, vol. 9, no. 1, p. e85150, 2014.
[14] M. Bibikova, B. Barnes, C. Tsan, V. Ho, B. Klotzle, J. M. Le, D. Delano, L. Zhang, G. P. Schroth, K. L. Gunderson, J.-B. Fan, and R. Shen, "High density dna methylation array with single cpg site resolution," *Genomics*, vol. 98, no. 4, pp. 288 – 295, 2011, new Genomic Technologies and Applications.
[15] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, April 1979.
[16] K. A. Hoadley, C. Yau *et al.*, "Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin," *Cell*, vol. 158, pp. 929–944, 2014.
[17] D. Wu, D. Wang, M. Q. Zhang, and J. Gu, "Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: application to cancer molecular classification," *BMC genomics*, vol. 16, no. 1, p. 1022, 2015.
[18] E. F. Lock, K. A. Hoadley, J. S. Marron, and A. B. Nobel, "Joint and individual variation explained (jive) for integrated analysis of multiple data types," *The annals of applied statistics*, vol. 7, no. 1, pp. 523–542, 2013.
[19] Q. Feng, M. Jiang, J. Hannig, and J. Marron, "Angle-based joint and individual variation explained," *Journal of Multivariate Analysis*, vol. 166, pp. 241 – 265, 2018.
[20] R. Shen, A. B. Olshen, and M. Ladanyi, "Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis," *Bioinformatics*, vol. 25, no. 22, pp. 2906–2912, 2009.
[21] I. Jolliffe, *Principal Component Analysis*, ser. Springer Series in Statistics. Springer, 2002.
[22] B. Wang *et al.*, "Similarity network fusion for aggregating data types on a genomic scale," *Nature methods*, vol. 11, no. 3, pp. 333–337, 2014.
[23] A. Khan and P. Maji, "Low-rank joint subspace construction for cancer subtype discovery." *IEEE/ACM transactions on computational biology and bioinformatics*, 2019, doi: 10.1109/TCBB.2019.2894635.
[24] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.
[25] D. Wu, D. Wang, M. Q. Zhang, and J. Gu, "Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: application to cancer molecular classification," *BMC genomics*, vol. 16, no. 1, p. 1022, 2015.
[26] E. F. Lock, K. A. Hoadley, J. S. Marron, and A. B. Nobel, "Joint and individual variation explained (jive) for integrated analysis of multiple data types," *The annals of applied statistics*, vol. 7, no. 1, pp. 523–542, 2013.
[27] R. Shen, A. B. Olshen, and M. Ladanyi, "Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis," *Bioinformatics*, vol. 25, no. 22, pp. 2906–2912, 2009.
[28] I. Jolliffe, *Principal Component Analysis*, ser. Springer Series in Statistics. Springer, 2002.
[29] S. Monti, P. Tamayo, J. Mesirov, and T. Golub, "Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data," *Machine Learning*, vol. 52, pp. 91–118, 2003.
[30] Wilkerson, M. D., Hayes, and D. Neil, "Consensusclusterplus: a class discovery tool with confidence assessments and item tracking," *Bioinformatics*, vol. 26, no. 12, pp. 1572–1573, 2010.
[31] K. T. Fang and Y. Wang, *Number-Theoretic Methods in Statistics*. Chapman and Hall/CRC, 1993.
[32] R. Shen, Q. Mo, N. Schultz, V. E. Seshan, A. B. Olshen, J. Huse, M. Ladanyi, and C. Sander, "Integrative subtype discovery in glioblastoma using icluster," *PloS one*, vol. 7, no. 4, p. e35236, 2012.
[33] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, 2002, pp. 849–856.