

MULTIPLE SIGNED GRAPH LEARNING FOR GENE REGULATORY NETWORK INFERENCE

Abdullah Karaaslanli¹, Satabdi Saha², Tapabrata Maiti² and Selin Aviyente¹

¹Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, US

²Department of Statistics and Probability, Michigan State University, East Lansing, MI, US

ABSTRACT

Many real-world data are represented through the relations between data samples, i.e., a graph structure. Although many datasets come with a pre-existing graph, there is still a large number of applications where the graph structure is not readily available. An essential task for such cases is graph learning (GL), which infers the graph structure from a set of graph signals. Existing GL techniques mostly focus on learning a single graph structure; however, samples are usually connected in multiple different ways. Furthermore, existing works can only handle unsigned graphs, while contemporary tasks require inference of signed graphs, which are better at representing similarity and dissimilarity of samples. In this paper, we propose a framework (mvSGL) for joint estimation of multiple related signed graphs. mvSGL optimizes the total variation of graph signals with respect to graphs while ensuring that the graphs are similar to each other through a consensus graph. mvSGL is employed in the inference of multiple gene regulatory networks (GRN) from single cell datasets that include multiple cell types. Performance evaluation using simulated and real datasets demonstrates the effectiveness of mvSGL in the inference of multiple related GRNs.

Index Terms— Graph Learning, Multiview Graphs, Gene Regulatory Networks

1. INTRODUCTION

In many modern data science applications, relationships between data samples are well described with a graph structure. While many real-world data are intrinsically graph-structured, e.g. social networks, there is still a large number of applications, such as single cell gene expression data, where the graph is not readily available. In the latter case, graphs are effective for revealing the relational structure and may assist in a variety of learning tasks. Graph learning (GL) deals with the construction of a topological structure among entities from a set of observations on these entities, i.e., graph signals.

In recent years, GL problem has been addressed from a graph signal processing (GSP) perspective, where the unknown graph is learned by exploiting the relation between graph signals and graph structure [1, 2]. However, most of the existing work on graph learning considers simple data, where all signals are assumed to be defined on a single graph structure. In many applications, the data may be heterogeneous or mixed and come from multiple related graphs, also known as multiview graphs. Example of this setup includes gene regulatory networks where the pairwise relations between genes varies across different cell types, functional connectivity of the brain where links differ across different frequency bands.

This work was in part supported by National Science Foundation grants CCF-2211645 and CCF-2006800.

Recent works consider the problem of learning multiple related graphs each with a subset of observations, also known as joint inference of multiple graphs [3]. In this setting, it is assumed that for each graph a set of graph signals are provided and unknown graphs are defined over the same nodes with different but related edges. This problem setting has been most widely studied for learning the topology of dynamic networks [4, 5, 6]. Assuming that the evolution of graph structure is smooth across time, the problem learns multiple graphs regularized with a term that promotes variation between consecutive graphs to be small. More recently, the problem of joint inference of multiple graphs has been formulated in [3], where signals are assumed to be stationarity with respect to graph structure and pairwise similarity between all graphs is promoted through regularization. An alternative to these approaches has been proposed in [7], where multiple graphs representing the interactions between nodes in different layers are available and this information is used to learn a global graph structure. Thus, a single graph is learned from multiview graphs rather than learning multiple graphs, simultaneously.

In this paper, we address the multiview GL problem with a specific focus on joint inference of gene regulatory networks (GRN) from multiple cell classes (conditions/disease states) from single cell data [8]. Unlike previous approaches which are limited to unsigned graphs, we will focus on multiview signed graph learning as GRNs are signed. To this end, we propose multiview signed graph learning (mvSGL) that extends previous signed graph learning method in [9] to multiview graph setup with the following contributions:

- mvSGL learns multiple signed graphs jointly while regularizing the learned graphs to be similar to each other,
- Similarity of view graphs is ensured through a learned consensus, which captures the common structure across views.
- Results on simulated and real single cell data illustrate the effectiveness of mvSGL.

The remaining of the paper is organized as follows. In Section 2, background on graphs and GL is provided. Section 3 describes the proposed method for joint learning of multiple signed graphs. Section 4 and 5 include results on simulated and real single cell datasets, respectively. Concluding remarks are in final section.

2. BACKGROUND

A weighted signed graph is $G = (V, E, \mathbf{W})$ where V is the node set with cardinality n , E is the edge set and $\mathbf{W}$ is the adjacency matrix with the weights of edges between two nodes allowed to take on both positive and negative values. A signed graph can be decomposed into two unsigned graphs, $G^+ = (V, E^+, \mathbf{W}^+)$ and $G^- = (V, E^-, \mathbf{W}^-)$, where $W_{ij}^+ = W_{ij}$ ($W_{ij}^- = |W_{ij}|$) if $W_{ij} > 0$

($W_{ij} < 0$), and 0, otherwise. Combinatorial Laplacian matrix of an unsigned graph G is $\mathbf{L} = \mathbf{D} - \mathbf{W}$ where $\mathbf{D}$ is the diagonal matrix with node degrees, i.e. $D_{ii} = \sum_{j=1}^n W_{ij}$. Since $\mathbf{L}$ is a positive semi-definite matrix, its eigendecomposition is $\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ with eigenvalues ordered such that $0 = \Lambda_{11} \leq \Lambda_{22} \leq \dots \leq \Lambda_{nn}$.

A graph signal defined on a graph is a vector $\mathbf{x} \in \mathbb{R}^n$ where x_i is the signal value on the i th node. An unknown unsigned graph G can be learned from a given set of graph signals $\{\mathbf{x}_i \in \mathbb{R}^n\}_{i=1}^p$ that are defined on G using assumptions made about the relation between graph structure and graph signals. Dong et. al. [10] use the assumption that graph signals are smooth with respect to G to learn the structure of G . A graph signal $\mathbf{x}$ is smooth with respect to G if the signal values of two nodes connected with an edge are similar to each other. The smoothness of $\mathbf{x}$ can be quantified by its total variation with respect to G , i.e. $\text{tr}(\mathbf{x}^T \mathbf{L} \mathbf{x})$. Thus, the following optimization problem is proposed in [10] to learn G , where the total variation of signals are minimized with respect to $\mathbf{L}$:

$$\underset{\mathbf{L} \in \mathbb{L}}{\text{minimize}} \text{tr}(\mathbf{X}^T \mathbf{L} \mathbf{X}) + \alpha \|\mathbf{L}\|_F^2 \text{ subject to } \text{tr}(\mathbf{L}) = 2n, \quad (1)$$

where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the data matrix with columns $\mathbf{x}_i$'s, $\mathbb{L} = \{\mathbf{L} : L_{ij} = L_{ji} \leq 0 \forall i \neq j, \mathbf{L} \mathbf{1} = \mathbf{0}\}$ is the set of valid Laplacian matrices. The first term is the total variation of signals, Frobenius norm of $\mathbf{L}$ regularizes the density of the learned graph such that larger values of α result in denser graphs. Finally, the last constraint prevents the trivial solution $\mathbf{L} = \mathbf{0}$.

Recently, we have introduced an extension of the unsigned graph learning problem in (1) to learn an unknown signed graph G based on the assumptions that the graph signals are (i) smooth with respect to G^+ , and (ii) non-smooth with respect to G^- [9]. These assumptions state that graph signals have small total variation on G^+ and large total variation on G^- . Thus, the signed graph G is learned with the following optimization problem:

$$\underset{\mathbf{L} \in \mathbb{L}, \mathbf{L}^- \in \mathbb{L}^-}{\text{minimize}} \sum_{s \in \{+, -\}} \text{tr}(\mathbf{K}^s \mathbf{L}^s) + \alpha_s \|\mathbf{L}^s\|_F^2 \quad (2)$$

subject to $\text{tr}(\mathbf{L}^s) = 2n \forall s$, and $(\mathbf{L}^+, \mathbf{L}^-) \in \mathbb{C}$,

where $\mathbf{L}^+$ and $\mathbf{L}^-$ are the Laplacian matrices of G^+ and G^- . $\mathbf{K}^+ = \mathbf{X}\mathbf{X}^T$, $\mathbf{K}^- = -\mathbf{X}\mathbf{X}^T$ and the cyclic property of trace operation, i.e. $\text{tr}(\mathbf{X}^T \mathbf{L} \mathbf{X}) = \text{tr}(\mathbf{X}\mathbf{X}^T \mathbf{L})$, is employed. In order to make sure that $\mathbf{L}^+$ and $\mathbf{L}^-$ are not non-zero at the same indices, they are constrained to be in the set $\mathbb{C} = \{(\mathbf{L}^+, \mathbf{L}^-) : L_{ij}^+ = 0 \text{ if } L_{ij}^- \neq 0 \text{ and } L_{ij}^- = 0 \text{ if } L_{ij}^+ \neq 0, \forall i \neq j\}$. Finally, trivial zero solutions are prevented as in (1).

3. MULTIVIEW SIGNED GRAPH LEARNING

3.1. Problem Formulation

Let $\{\mathbf{X}^i \in \mathbb{R}^{n \times p_i}\}_{i=1}^N$ be an N -view dataset of graph signals, where the columns of $\mathbf{X}^i$ are the p_i graph signals defined on an unknown signed graph $G^i = (V, E^i, \mathbf{W}^i)$. While the node set is assumed to be common across the different view, the edge sets E^i 's and the corresponding edge weights are assumed to be different but similar to each other across views. Based on this assumption, G^i can be inferred jointly, which will allow information sharing across views during learning. To this end, we propose an optimization problem that extends (2) to learn G^i 's simultaneously while ensuring that they are similar to each other. Similarity of learned G^i 's is achieved through a regularization term that forces each G^i to be close to a

consensus signed graph G , which is also learned by combining information from G^i 's. Thus, the structure of G reflects the common connections shared across G^i 's.

Let $\mathbf{L}^{i,s}$ be the Laplacian matrices of $G^{i,s}$ for all $i \in \{1, \dots, N\}$ and $s \in \{+, -\}$. Also, let the Laplacian matrices of positive and negative part of the consensus graph G be $\mathbf{L}^+$ and $\mathbf{L}^-$. Define sets $\mathcal{L}^+ = \{\mathbf{L}^{1,+}, \dots, \mathbf{L}^{N,+}, \mathbf{L}^+\}$ and $\mathcal{L}^- = \{\mathbf{L}^{1,-}, \dots, \mathbf{L}^{N,-}, \mathbf{L}^-\}$. The optimization problem for joint learning of G^i 's and G is then:

$$\begin{aligned} &\underset{\mathcal{L}^+, \mathcal{L}^-}{\text{minimize}} \sum_{s \in \{+, -\}} \sum_{i=1}^N \{ \text{tr}(\mathbf{K}^{i,s} \mathbf{L}^{i,s}) + \alpha_s \|\mathbf{L}^{i,s}\|_F^2 \\ &\quad + \beta_s \|\mathbf{L}^{i,s} - \mathbf{L}^s\|_{F, \text{off}}^2 \} + \gamma_+ \|\mathbf{L}^+\|_{1, \text{off}} + \gamma_- \|\mathbf{L}^-\|_{1, \text{off}} \\ &\text{subject to } \mathbf{L}^{i,s} \in \mathbb{L}, \text{tr}(\mathbf{L}^{i,s}) = 2n, \forall i, \forall s \in \{+, -\} \quad (3) \\ &\quad (\mathbf{L}^{i,+}, \mathbf{L}^{i,-}) \in \mathbb{C} \forall i, \mathbf{L}^+, \mathbf{L}^- \in \mathbb{L}, (\mathbf{L}^+, \mathbf{L}^-) \in \mathbb{C}, \end{aligned}$$

where $\mathbf{K}^{i,+} = \mathbf{X}^i \mathbf{X}^{iT}$, $\mathbf{K}^{i,-} = -\mathbf{X}^i \mathbf{X}^{iT}$, $\|\cdot\|_{F, \text{off}}$ and $\|\cdot\|_{1, \text{off}}$ are the Frobenius norm and the ℓ_1 -norm of the off-diagonal entries, respectively. Following (1), for each view i , we minimize the total variation of signals with respect to $G^{i,+}$ and maximize the total variation of signals with respect to $G^{i,-}$. The second term controls the density of the learned $G^{i,+}$ ($G^{i,-}$) such that for larger values of α_+ (α_-), we learn denser graphs. The third term is the regularizer that ensures that each view graph G^i is similar to the consensus graph G with β_+ (β_-) controlling amount of the regularization. The last term is added to control the density learned consensus graph with smaller values of γ_+ and γ_- resulting in a denser consensus graph. Finally, each G^i and G are subject to the same constraints as in (2).

3.2. Optimization

The problem in (3) can be written in a vectorized form, where one learns the upper triangular parts of the Laplacian matrices. For this, we first define the following operators: $\text{diag}() : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^n$ is an operator that returns the diagonal of the input matrix. The operator $\text{upper}() : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n(n-1)/2}$ returns the upper triangular part of the input matrix. For an $n \times n$ symmetric matrix $\mathbf{A}$, define $\mathbf{S} \in \mathbb{R}^{M \times n}$ such that $\text{Supper}(\mathbf{A}) = \mathbf{A} \mathbf{1} - \text{diag}(\mathbf{A})$. Finally, let $\mathbf{1}$ and $\mathbf{0}$ be all-one and all-zero vectors. To vectorize (3), let $\mathbf{k}^{i,s} = \text{upper}(\mathbf{K}^{i,s})$, $\mathbf{d}^i = \text{diag}(\mathbf{K}^{i,s})$, $\mathbf{\ell}^{i,s} = \text{upper}(\mathbf{L}^{i,s})$ and $\mathbf{\ell}^s = \text{upper}(\mathbf{L}^s)$ for $s \in \{+, -\}$. Also, let $\mathcal{L}_v^+ = \{\mathbf{\ell}^{1,+}, \dots, \mathbf{\ell}^{N,+}, \mathbf{\ell}^+\}$ and $\mathcal{L}_v^- = \{\mathbf{\ell}^{1,-}, \dots, \mathbf{\ell}^{N,-}, \mathbf{\ell}^-\}$. The vectorized form of (3) is:

$$\begin{aligned} &\underset{\mathcal{L}_v^+, \mathcal{L}_v^-}{\text{minimize}} \sum_{s \in \{+, -\}} \sum_{i=1}^N \{ \langle \mathbf{k}^{i,s} - \mathbf{S}^T \mathbf{d}^{i,s}, \mathbf{\ell}^{i,s} \rangle + \alpha \|\mathbf{S} \mathbf{\ell}^{i,s}\|_2^2 \\ &\quad + 2\alpha \|\mathbf{\ell}^{i,s}\|_2^2 + \beta \|\mathbf{\ell}^{i,s} - \mathbf{\ell}^s\|_2^2 \} + \gamma_+ \|\mathbf{\ell}^+\|_1 + \gamma_- \|\mathbf{\ell}^-\|_1 \quad (4) \\ &\text{subject to } \mathbf{1}^T \mathbf{\ell}^{i,+} = -n, \mathbf{1}^T \mathbf{\ell}^{i,-} = -n, \mathbf{\ell}^{i,+} \leq 0, \mathbf{\ell}^{i,-} \leq 0, \\ &\quad \mathbf{\ell}^{i,+} \perp \mathbf{\ell}^{i,-} \forall i \text{ and } \mathbf{\ell}^+ \leq 0, \mathbf{\ell}^- \leq 0, \mathbf{\ell}^+ \perp \mathbf{\ell}^-, \end{aligned}$$

where the first term in the summation corresponds to the first term in (3), and the correspondence between the remaining terms to the terms in (3) can be deduced using the hyperparameters. First two constraints correspond to the trace constraints in (3). $\mathbf{\ell}^{i,+} \perp \mathbf{\ell}^{i,-}$ together with non-negativity constraint is called complementarity constraint [11] and corresponds to $(\mathbf{L}^{i,+}, \mathbf{L}^{i,-}) \in \mathbb{C}$ in (3).

The problem in (4) is non-convex due to complementarity constraints. ADMM is shown to be convergent for problems with complementarity constraints under some assumptions [12]. To write the problem in ADMM form, introduce auxiliary variables $\mathbf{v}^{i,s} = \mathbf{\ell}^{i,s}$

for all i and $s \in \{+, -\}$. Similarly, introduce $\mathbf{v}^s = \ell^s$ for all s . Also, let $\mathcal{V}^s = \{\mathbf{v}^{1,s}, \dots, \mathbf{v}^{N,s}, \mathbf{v}^s\}$ for $s \in \{+, -\}$. Then, the problem in its standard ADMM form is:

$$\begin{aligned} & \underset{\mathcal{L}_v^+, \mathcal{L}_v^-, \mathcal{V}^+, \mathcal{V}^-}{\text{minimize}} \sum_{i=1}^N \iota_S(\mathbf{v}^{i,+}, \mathbf{v}^{i,-}) + \sum_{s \in \{+, -\}} \sum_{i=1}^N f(\ell^{i,s}, \ell^s) + \iota_H(\ell^{i,s}) \\ & \quad + \iota_S(\mathbf{v}^+, \mathbf{v}^-) + \gamma_+ \|\ell^+\|_1 + \gamma_- \|\ell^-\|_1 \quad (5) \\ & \text{subject to } \mathbf{v}^{i,+} = \ell^{i,+}, \mathbf{v}^{i,-} = \ell^{i,-}, \mathbf{v}^+ = \ell^+, \text{ and } \mathbf{v}^- = \ell^-, \end{aligned}$$

where $f(\ell^{i,s}, \ell^s) = \langle \mathbf{k}^{i,s} - \mathbf{S}^\top \mathbf{d}^{i,s}, \ell^{i,s} \rangle + \alpha \|\mathbf{S} \ell^{i,s}\|_2^2 + 2\alpha \|\ell^{i,s}\|_2^2 + \beta \|\ell^{i,s} - \ell^s\|_2^2$, $\iota_S(\cdot, \cdot)$ is the indicator function for the complementarity set $S = \{(\mathbf{v}, \mathbf{w}) : \mathbf{v} \leq 0, \mathbf{w} \leq 0, \mathbf{v} \perp \mathbf{w}\}$, and $\iota_H(\cdot)$ is the indicator function for the hyperplane $H = \{\ell : \mathbf{1}^\top \ell = -n\}$. Let $\Lambda^s = \{\lambda_{1,s}, \dots, \lambda_{N,s}, \lambda_s\}$ for $s \in \{+, -\}$ be the set of Lagrangian multipliers associated with constraints in (5). Augmented Lagrangian is then:

$$\begin{aligned} L_p(\mathcal{L}_v^+, \mathcal{L}_v^-, \mathcal{V}^+, \mathcal{V}^-, \Lambda^+, \Lambda^-) &= \sum_{i=1}^N \iota_S(\mathbf{v}^{i,+}, \mathbf{v}^{i,-}) \\ &+ \sum_{s \in \{+, -\}} \sum_{i=1}^N \{f(\ell^{i,s}, \ell^s) + \iota_H(\ell^{i,s}) + \lambda_{i,s}^\top (\mathbf{v}^{i,s} - \ell^{i,s}) \\ &+ \frac{\rho}{2} \|\mathbf{v}^{i,s} - \ell^{i,s}\|_2^2\} + \iota_S(\mathbf{v}^+, \mathbf{v}^-) \quad (6) \\ &+ \sum_{s \in \{+, -\}} \{\gamma_s \|\ell^s\|_1 \lambda_s^\top (\mathbf{v}^s - \ell^s) + \frac{\rho}{2} \|\mathbf{v}^s - \ell^s\|_2^2\}, \end{aligned}$$

where ρ is the parameter of augmented Lagrangian. Using augmented Lagrangian, ADMM steps at k th iteration are then:

$$(\hat{\mathbf{V}}^+, \hat{\mathbf{V}}^-) = \underset{\mathcal{V}^+, \mathcal{V}^-}{\text{argmin}} L_p(\hat{\mathcal{L}}_v^+, \hat{\mathcal{L}}_v^-, \mathcal{V}^+, \mathcal{V}^-), \quad (7)$$

$$(\hat{\mathcal{L}}_v^+, \hat{\mathcal{L}}_v^-) = \underset{\mathcal{L}_v^+, \mathcal{L}_v^-}{\text{argmin}} L_p(\mathcal{L}_v^+, \mathcal{L}_v^-, \hat{\mathbf{V}}^+, \hat{\mathbf{V}}^-), \quad (8)$$

$$\hat{\lambda}_{i,s} = \hat{\lambda}_{i,s} + \rho(\hat{\mathbf{v}}^{i,s} - \hat{\ell}^{i,s}), \quad \forall i, s \quad (9)$$

$$\hat{\lambda}_s = \hat{\lambda}_s + \rho(\hat{\mathbf{v}}^s - \hat{\ell}^s), \quad \forall s \quad (10)$$

where $\hat{\cdot}$ and $\hat{\cdot}$ represent the values of variables at k th and $(k-1)$ th iteration, respectively. To solve (7), we use the fact that it can be solved for each $(\mathbf{v}^i, \mathbf{w}^i)$ pair (and $(\mathbf{v}, \mathbf{w})$), separately. This separation leads to a set of optimization problems all of which can be solved by projection onto the complementarity set S . The problem in (8) is separable across $\mathcal{L}_v^+$ and $\mathcal{L}_v^-$, leading to two optimization problems both of which can be solved with Block Coordinate Descent (BCD) [13].

4. SIMULATIONS

In this section, we test mvSGL¹ on simulated single cell data with multiple gene expression datasets generated from multiple related GRNs. We also compare the performance of mvSGL to SGL (see optimization problem in (2)). Since SGL can learn a single signed graph at a time, it is applied to each dataset separately to learn GRNs. Both SGL and mvSGL require the selection of α_s 's, which is set such that the learned $G^{i,s}$ has edge density around 0.1 for all i and

¹Codes can be found at <https://github.com/Single-Cell-Graph-Learning/scMSGSL>

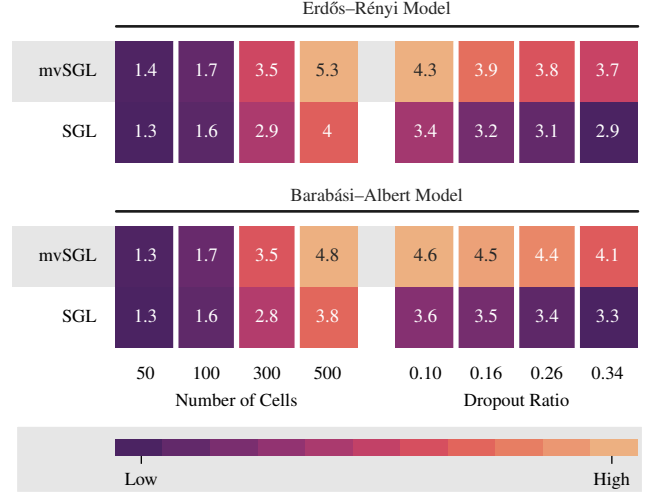


Fig. 1. AUPRC ratios of the methods on simulated data generated by two different random graph models. Left and right panels show how the performance of methods changes with increasing number of cells and increasing dropout ratios, respectively.

$s \in \{+, -\}$. Similarly, γ_s parameters are selected such that G^s has edge density around 0.1. β_s is selected such that the correlation between G^i and G^j is around 0.5 $\forall i \neq j$. The multiclass version of area under the precision-recall curve (AUPRC), i.e. the average of AUPRC+ and AUPRC-, is used as the evaluation metric, where the activating and inhibitory edges of the learned GRNs are compared to the activating and inhibitory edges of the ground truth GRNs, respectively. We report the ratio of AUPRC values obtained by the methods to those of a random estimator.

Data Generation: We simulate gene expression data from a multivariate zero-inflated negative binomial (ZINB) distribution, which is shown to accurately capture the characteristics of single cell datasets [14, 15]. Given a known graph structure, we generate synthetic datasets using an algorithm developed by [16] and illustrated in [17, 9]. In order to generate multiple single cell datasets, we first create a baseline graph G with 100 nodes using either an Erdős-Rényi (ER) model or a Barabási-Albert (BA) model. G is then converted to a signed graph by randomly selecting half of its edges as negative, while the remaining ones are set as positive. Next, $\{G_i\}_{i=1}^5$ are generated by adding $0.9 \times \binom{n}{2} \times \eta$ new edges to the baseline graph G . We set $\eta = 0.1$ such that 90% of the edges are common across G^i 's. Half of the added edges are set as negative edges, while the other half is positive. From each G^i , simulated gene expressions from p cells are generated using the ZINB simulator to create $\mathbf{X}^i$. The three parameters of the ZINB distribution were determined using a real scRNA-seq dataset [18]. Each simulation is repeated 10 times and the average performance over 10 realizations is reported.

Experiment 1: In the left panel of Fig.1, we report how the performance of the methods changes with varying number of cells p . It is seen that for the different cell numbers, mvSGL has higher AUPRC ratios than SGL, with the difference in performance increasing with increasing number of cells. This indicates that mvSGL shares valuable information across views, thus leading to improved performance. As expected, the performance of both methods improves with increasing number of cells. These observations hold for both random graph models.

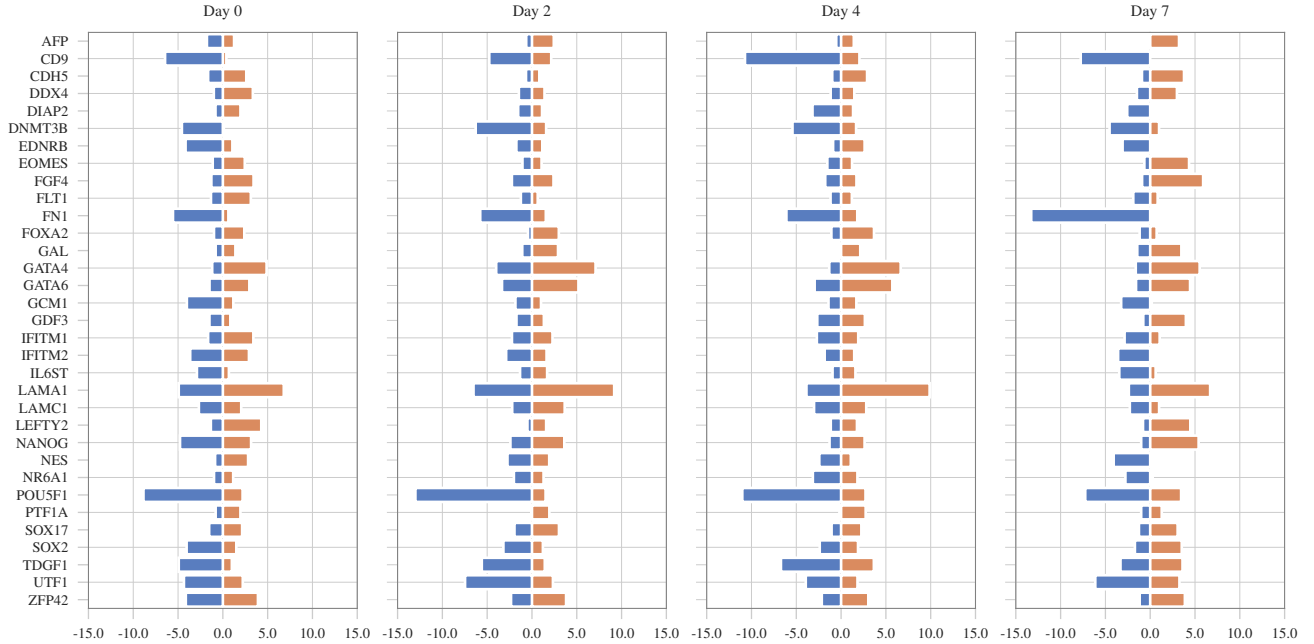


Fig. 2. Genes with the highest node strengths. Orange and blue bars indicate that the strength is calculated using activating and inhibitory edges, respectively. Only genes whose activating or inhibitory strength is among the top 10 genes in any view are shown.

Experiment 2: An important feature of single cell data is dropout phenomena, where a high percentage of genes are not expressed in cells due to technical noise or genuine biological variability [19]. To observe how methods are influenced from the dropout ratio of single cell datasets, we generate datasets with increasing dropout ratio while $p = 400$. Right panel of Fig. 1 shows the AUPRC ratios for mvSGL and SGL for both random graph models. Similar to cell sensitivity analysis, scMSGSL performs better compared to SGL. Increasing dropout ratio causes a drop in the performance of both methods. This is expected as increasing the number of zeros in the datasets due to technical or biological noise makes learning harder. As in the previous experiment, these observations hold for both ER and BA models.

5. REAL DATA

In this section, mvSGL is employed to study the differentiation process in mouse embryonic stem cells (mESC) [20]. Single cell datasets from [21] generated using high-throughput droplet-microfluidic approach were used to study differentiation in mESC before and after leukemia inhibitory factor (LIF) withdrawal. The dataset contains cells sampled from 4 states: before LIF withdrawal (day 0) and after the withdrawal (days 2, 4 and 7). Each cell type has 72 genes and the number of cells in each subgroup are respectively: 933, 303, 683 and 798. Hyperparameters of mvSGL are selected the same way as in simulated datasets.

Since there is no ground truth, we validate the results by analyzing the centrality of different nodes in the learned graphs. In Fig. 2, the positive and negative edge strengths of central genes in the learned GRNs are plotted. NANOG, SOX2, ZFP42 are found to be the central inhibitory genes in the early stages of differentiation with their inhibitory strength diminishing as cells proceed to a mature state. POU5F1 and UTF1 also exhibit higher number of inhibitory

relationships in the the first few days and their strength reduces in Days 4 and 7. Centrality of these genes in early stages of the development is inline with previous works [22, 23, 24]. Reduction in the expression of NANOG has been shown to be correlated with the induction of genes GATA4 and GATA6 which initiate differentiation of pluripotent cells [25] and therefore GATA4 and GATA6 has been correctly identified as central in Days 2, 4 and 7. Collectively, these results indicate that mvSGL manages to find GRNs whose central nodes are inline with previous research.

6. CONCLUSIONS

In this work, we proposed mvSGL for learning multiple related signed graphs with a specific focus on GRN inference from single cell datasets containing gene expressions from multiple cell types. Compared to previous work on multiview graph learning, mvSGL learns signed graphs, which are important structures to represent similarity and dissimilarity between nodes. Furthermore, mvSGL can learn a consensus signed graph, that captures the shared structure across multiview graphs. Although we did not explicitly study the properties of learned consensus graphs, they can be beneficial when the focus is on aggregation of multiview graphs as in [7]. Our results on simulated and real single cell datasets that contain gene expressions from multiple cell types indicate the effectiveness of the method in learning multiple related signed graphs.

There are some open problems that can be considered as future work. First, in our current formulation the node set is assumed to be the same across views. However, in some applications some of the nodes may be hidden in different views. Second, we used squared Frobenius norm for the regularization term to ensure that the view graphs are close to the consensus graph. Other norms, such as L1 norm, can be considered to promote the sparsity of the variation across views.

7. REFERENCES

- [1] Gonzalo Mateos, Santiago Segarra, Antonio G Marques, and Alejandro Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 16–43, 2019.
- [2] Xiaowen Dong, Dorina Thanou, Michael Rabbat, and Pascal Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 44–63, 2019.
- [3] Madeline Navarro, Yuhao Wang, Antonio G Marques, Caroline Uhler, and Santiago Segarra, "Joint inference of multiple graphs from matrix polynomials," *Journal of Machine Learning Research*, vol. 23, no. 76, pp. 1–35, 2022.
- [4] Vassilis Kalofolias, Andreas Loukas, Dorina Thanou, and Pascal Frossard, "Learning time varying graphs," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Ieee, 2017, pp. 2826–2830.
- [5] Koki Yamada, Yuichi Tanaka, and Antonio Ortega, "Time-varying graph learning based on sparseness of temporal variation," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 5411–5415.
- [6] Stefania Sardellitti, Sergio Barbarossa, and Paolo Di Lorenzo, "Online learning of time-varying signals and graphs," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 5230–5234.
- [7] Eda Bayram, Dorina Thanou, Elif Vural, and Pascal Frossard, "Mask combination of multi-layer graphs for global structure inference," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 394–406, 2020.
- [8] Antoine-Emmanuel Saliba, Alexander J Westermann, Stanislaw A Gorski, and Jörg Vogel, "Single-cell rna-seq: advances and future challenges," *Nucleic acids research*, vol. 42, no. 14, pp. 8845–8860, 2014.
- [9] Abdullah Karaaslanli, Satabdi Saha, Selin Aviyente, and Tapabrata Maiti, "scsgl: kernelized signed graph learning for single-cell gene regulatory network inference," *Bioinformatics*, vol. 38, no. 11, pp. 3011–3019, 2022.
- [10] Xiaowen Dong, Dorina Thanou, Pascal Frossard, and Pierre Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.
- [11] Holger Scheel and Stefan Scholtes, "Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity," *Mathematics of Operations Research*, vol. 25, no. 1, pp. 1–22, 2000.
- [12] Yu Wang, Wotao Yin, and Jinshan Zeng, "Global convergence of admm in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, no. 1, pp. 29–63, 2019.
- [13] Hao-Jun Michael Shi, Shenyinying Tu, Yangyang Xu, and Wotao Yin, "A primer on coordinate descent algorithms," *arXiv preprint arXiv:1610.00040*, 2016.
- [14] Davide Risso, Fanny Perraudeau, Svetlana Gribkova, Sandrine Dudoit, and Jean-Philippe Vert, "Zinb-wave: A general and flexible method for signal extraction from single-cell rna-seq data," *bioRxiv*, p. 125112, 2017.
- [15] Christoph Hafemeister and Rahul Satija, "Normalization and variance stabilization of single-cell rna-seq data using regularized negative binomial regression," *Genome biology*, vol. 20, no. 1, pp. 1–15, 2019.
- [16] Inbal Yahav and Galit Shmueli, "On generating multivariate poisson data in management science applications," *Applied Stochastic Models in Business and Industry*, vol. 28, no. 1, pp. 91–102, 2012.
- [17] Bochao Jia, Suwa Xu, Guanghua Xiao, Vishal Lamba, and Faming Liang, "Learning gene regulatory networks from next generation sequencing data," *Biometrics*, vol. 73, no. 4, pp. 1221–1230, 2017.
- [18] Volker Hovestadt, Kyle S Smith, Laure Bihannic, Mariella G Filbin, McKenzie L Shaw, Alicia Baumgartner, John C DeWitt, Andrew Groves, Lisa Mayr, Hannah R Weisman, et al., "Resolving medulloblastoma cellular architecture by single-cell genomics," *Nature*, vol. 572, no. 7767, pp. 74–79, 2019.
- [19] Justin D Silverman, Kimberly Roche, Sayan Mukherjee, and Lawrence A David, "Naught all zeros in sequence count data are the same," *Computational and structural biotechnology journal*, vol. 18, pp. 2789–2798, 2020.
- [20] Hirotaka Matsumoto, Hisanori Kiryu, Chikara Furusawa, Minoru SH Ko, Shigeru BH Ko, Norio Gouda, Tetsutaro Hayashi, and Itoshi Nikaido, "Scode: an efficient regulatory network inference algorithm from single-cell rna-seq during differentiation," *Bioinformatics*, vol. 33, no. 15, pp. 2314–2321, 2017.
- [21] Allon M Klein, Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A Weitz, and Marc W Kirschner, "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells," *Cell*, vol. 161, no. 5, pp. 1187–1201, 2015.
- [22] Ian Chambers, Douglas Colby, Morag Robertson, Jennifer Nichols, Sonia Lee, Susan Tweedie, and Austin Smith, "Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells," *Cell*, vol. 113, no. 5, pp. 643–655, 2003.
- [23] Kaoru Mitsui, Yoshimi Tokuzawa, Hiroaki Itoh, Kohichi Segawa, Mirei Murakami, Kazutoshi Takahashi, Masayoshi Maruyama, Mitsuyo Maeda, and Shinya Yamanaka, "The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and es cells," *Cell*, vol. 113, no. 5, pp. 631–642, 2003.
- [24] Qing Zhou, Hiram Chipperfield, Douglas A Melton, and Wing Hung Wong, "A gene regulatory network in mouse embryonic stem cells," *Proceedings of the National Academy of Sciences*, vol. 104, no. 42, pp. 16438–16443, 2007.
- [25] Shelley R Hough, Ian Clements, Peter J Welch, and Kristin A Wiederholt, "Differentiation of mouse embryonic stem cells after rna interference-mediated silencing of oct4 and nanog," *Stem cells*, vol. 24, no. 6, pp. 1467–1475, 2006.