

Xueqian Li xueqianl@andrew.cmu.edu
 Dong Wang dongw1@andrew.cmu.edu
 Yiwen Yuan yiyweny@andrew.cmu.edu

Using Deep Kernels in Time Varying Networks for Reverse-Engineering of Gene Interaction

Background

We explore the performance of deep kernels on graphical models, which incorporates both the structural properties of deep learning architectures with non-parametric flexibility of kernels.

In our project, we aim to solve the task of reverse-engineering of gene interaction⁴. Inspired by KELLER¹, Pairwise MRF is used to model interactions between genes. Deep kernels, constructed from neural network and LSTM are used for learning parameters of the Pairwise MRF.

Limitations and Motivations

1. RBF kernel uses predefined factor which is not changeable during time varying process
2. A strong assumption that the interaction of pair genes is inversely related to their distance
3. The use of deep kernels can be a better candidate for capturing time dependencies
4. Deep kernels are more flexible to learn MRF kernel weights.

Introduction

We adopt the deep kernel design from Deep Kernel Learning⁵, starting from a base kernel $k(x_i, x_j | \theta)$ with hyperparameters θ , we transform the inputs X as

$$k(x_i, x_j | \theta) \rightarrow k(g(x_i, w), g(x_j, w) | \theta, w)$$

where $g(x, w)$ is a neural network and a LSTM. For the base kernel $k(x_i, x_j | \theta)$, we use the RBF kernel.

Methods

1. KELLER (kernel-reweighted logistic regression method)

We express the gene measurement distributions at one time step as a binary pair-wise Markov Random Field (MRF)

$$\mathbb{P}_{\theta^{(t)}}(X^{(t)}) := \frac{1}{Z(\theta^{(t)})} \exp \left(\sum_{(u,v) \in \mathcal{E}^{(t)}} \theta_{uv}^{(t)} X_u^{(t)} X_v^{(t)} \right) \quad \hat{\theta}_{uv}^{(t)} = \underset{\theta_{uv}^{(t)} \in \mathbb{R}^{p-1}}{\operatorname{argmin}} \left(- \sum_{i=1}^n w^{(t)}(t_i) \gamma(\theta_{uv}^{(t)}; x^{(t_i)}) + \lambda \|\theta_{uv}^{(t)}\|_1 \right)$$

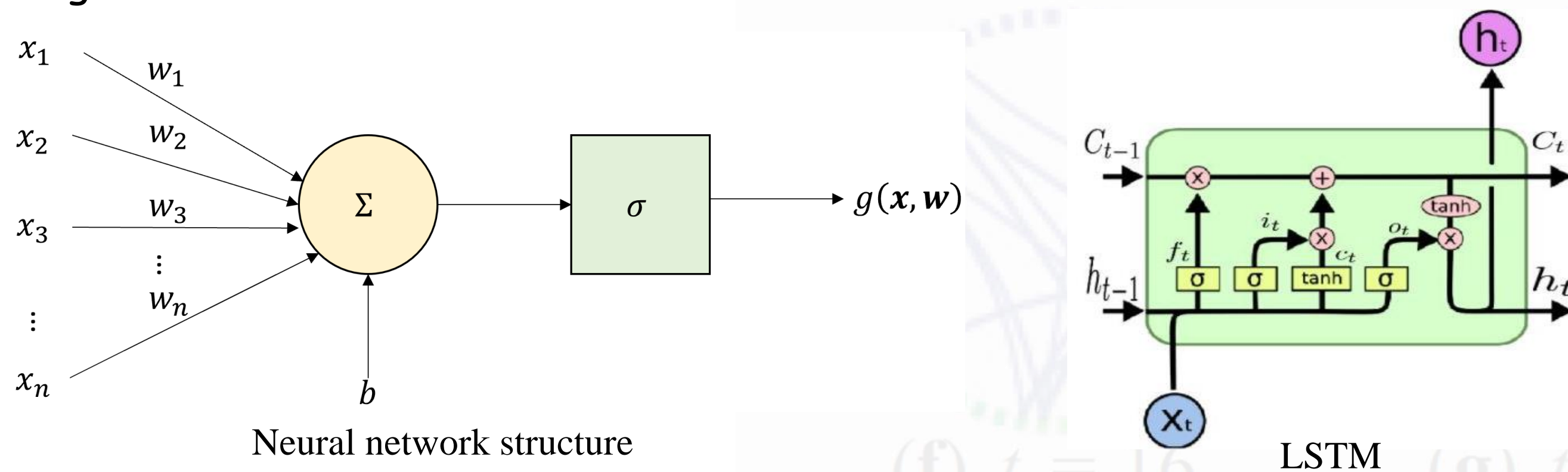
where $\theta_{uv}^{(t)}$ measures the strength of undirected interaction between gene u and v .

2. Neural Network

We use logistic regression, i.e. express the distribution of each gene conditioned on other genes.

3. LSTM (Long short-term memory)

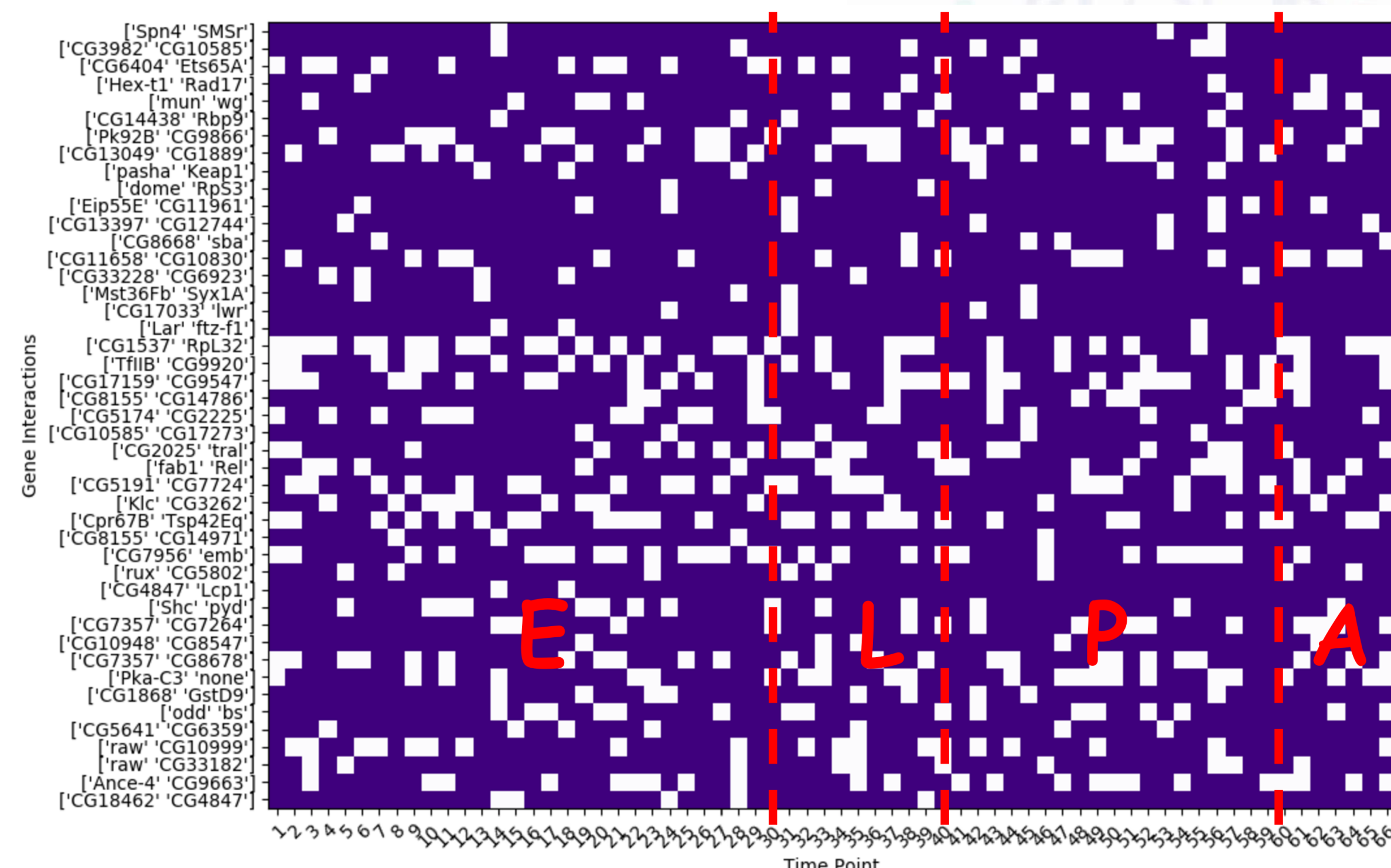
Using LSTM to generate the MRF factors dynamically, we can encapsulate dynamic interactions between genes into the kernel factors and thus recover more flexible and reasonable results.



Experiments

Gene Interaction

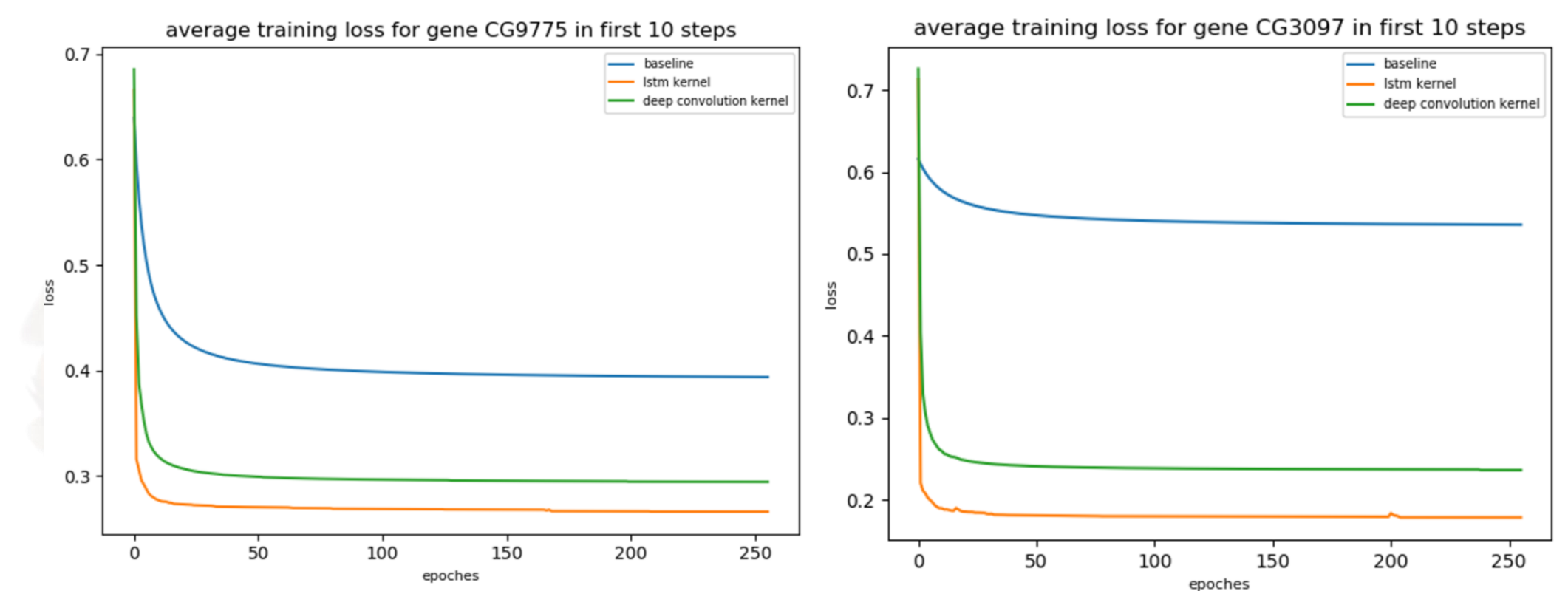
The task that we aim at solving is the problem of reverse engineering the time-evolving network between 588 genes involved in the developmental process during the life cycle of *Drosophila melanogaster*⁴ in a 66-step time series, which covers the E, L, P, A 4 different developmental stages.



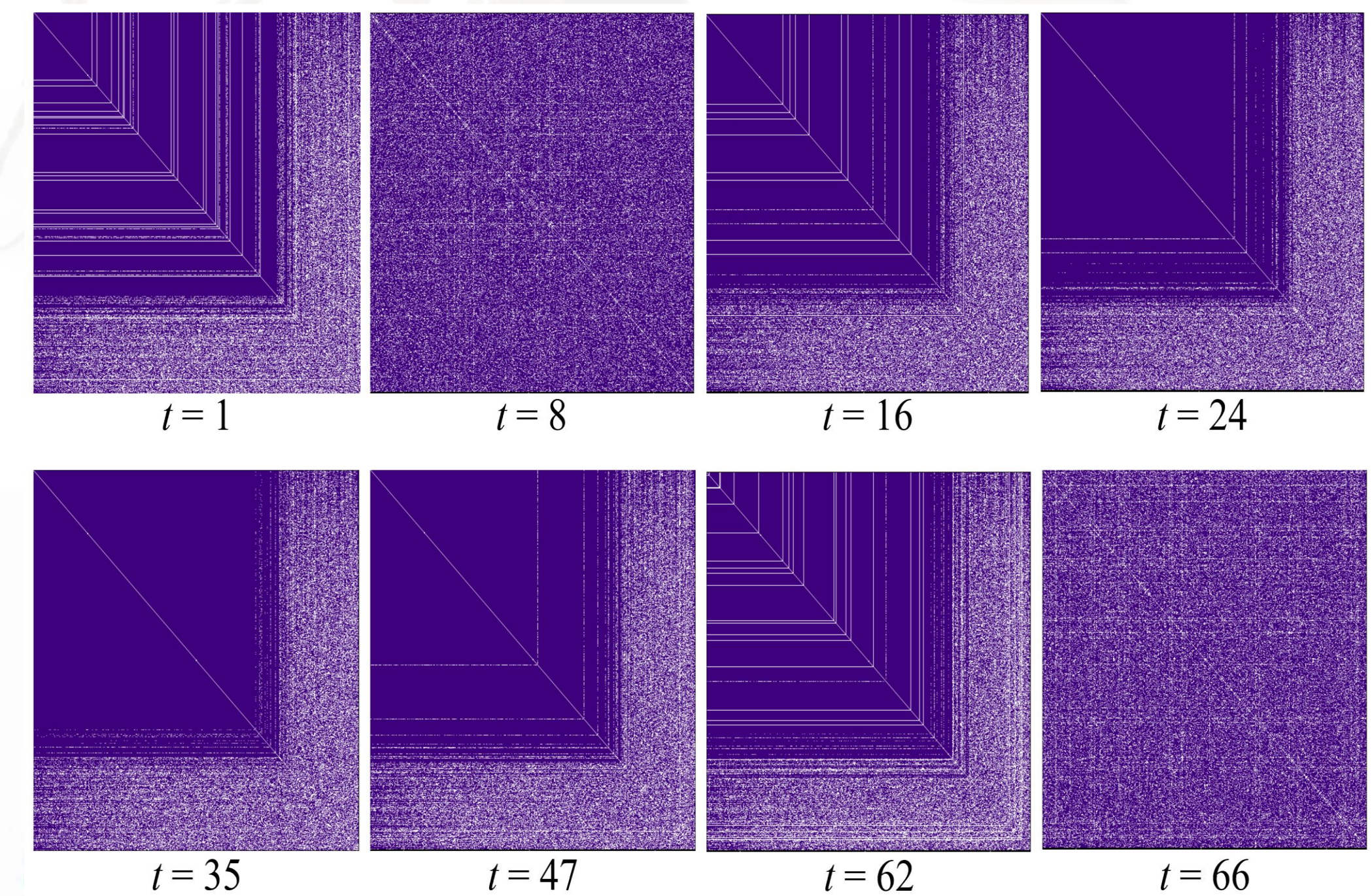
Interactions of gene pairs at different time point during embryonic stage (E), larval stage (L), pupal stage (P) and adult stage (A). The purple block indicates the interactions, where the white block denotes absent interactions.

Reference:

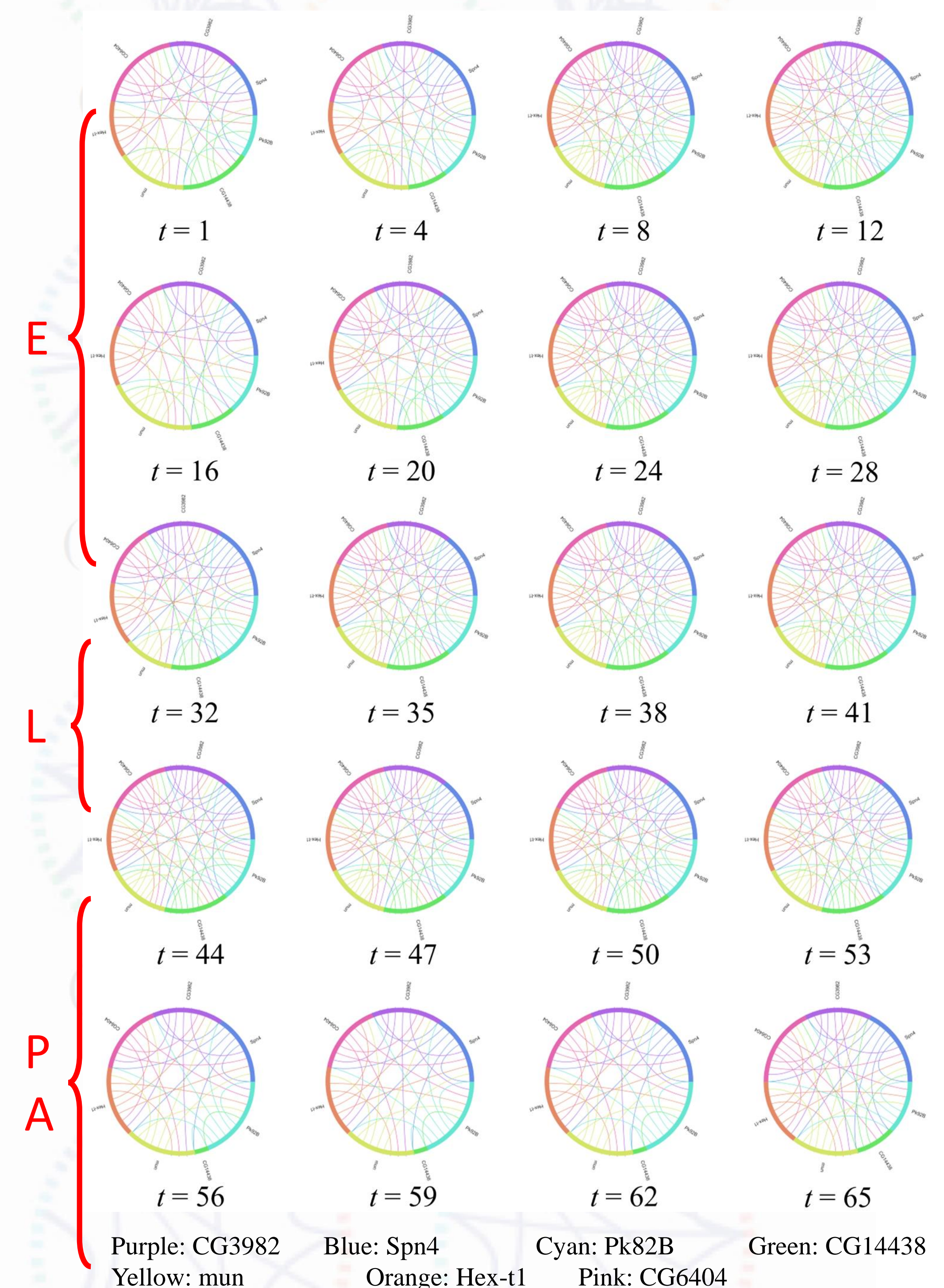
1. Song, Le, Mladen Kolar, and Eric P. Xing. "KELLER: estimating time-varying interactions between genes." *Bioinformatics* 25.12 (2009): i128-i136.
2. Kolar, Mladen, et al. "Estimating time-varying networks." *The Annals of Applied Statistics* 4.1 (2010): 94-123.
3. Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." *nature* 393.6684 (1998): 440.
4. Arbeitman, Michelle N., et al. "Gene expression during the life cycle of *Drosophila melanogaster*." *Science* 297.5590 (2002): 2270-2275.
5. Andrew Gordon Wilson, Zhiting Hu, Ruslan R Salakhutdinov, and Eric P Xing. Stochastic variational deep kernel learning. In *Advances in Neural Information Processing Systems*, pages 2586-2594, 2016b.



Training loss over epochs for 3 proposed methods



Dot plots of the adjacency matrices of the networks.



Interactions of 7 selected developmental genes in different time steps.



Interactions between genes in different development stage

Results

For each timestep, as we proceed to genes, using weights generated from deep kernels enables faster convergence than the non-parametric RBF kernels. In particular, LSTM does a better job in learning sequential weights than one-layer neural network.

Conclusions

We proposed the use of deep kernels in learning graphical models, in particular, pairwise MRF. The kernels have a structure and learning biases of deep architectures, which is more suitable for learning time-varying sequences, especially the ones where the relationship between time and influence is unknown. In our experiment, the deep kernel outperforms non-parametric RBF kernel in learning true data. In short, the deep kernel provides a natural mechanism for graphical models, especially in tasks like estimating gene interaction where using non-parametric weights limit the learning capabilities of the graphical model.