NetREX: Network Rewiring using EXpression - Towards Context Specific Regulatory Networks

Yijie Wang[1,⋆], Dong-Yeon Cho[1,⋆], Hangnoh Lee[2], Brian Oliver[2], and Teresa M Przytycka[1]

[1] National Center of Biotechnology Information, National Library of Medicine, NIH, Bethesda MD 20894, USA
przytyck@ncbi.nlm.nih.gov

[2] Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, 50 South Drive, Bethesda, MD 20892, USA

Abstract. Understanding gene regulation is a fundamental step towards understanding of how cells function and respond to environmental cues and perturbations. An important step in this direction is the ability to infer the transcription factor (TF)-gene regulatory network (GRN). However gene regulatory networks are typically constructed disregarding the fact that regulatory programs are conditioned on tissue type, developmental stage, sex, and other factors. Due to lack of the biological context specificity, these context-agnostic networks may not provide insight for revealing the precise actions of genes for a specific biological system under concern. Collecting multitude of features required for a reliable construction of GRNs such as physical features (TF binding, chromatin accessibility) and functional features (correlation of expression or chromatin patterns) for every context of interest is costly. Therefore we need methods that is able to utilize the knowledge about a context-agnostic network (or a network constructed in a related context) for construction of a context specific regulatory network.

To address this challenge we developed a computational approach that utilizes expression data obtained in a specific biological context such as a particular development stage, sex, tissue type and a GRN constructed in a different but related context (alternatively an incomplete or a noisy network for the same context) to construct a context specific GRN. Our method, NetREX, is inspired by network component analysis (NCA) that estimates TF activities and their influences on target genes given predetermined topology of a TF-gene network. To predict a network under a different condition, NetREX removes the restriction that the topology of the TF-gene network is fixed and allows for adding and removing edges to that network. To solve the corresponding optimization problem, which is non-convex and non-smooth, we provide general mathematical framework allowing use of the recently proposed Proximal Alternative Linearized Maximization technique and prove that our formulation has the properties required for convergence.

We tested our NetREX on simulated data and subsequently applied it to gene expression data in adult females from 99 hemizygotic lines of the *Drosophila* deletion (DrosDel) panel. The networks predicted by NetREX showed higher biological consistency than alternative approaches. In addition, we used the list of recently identified targets of the Doublesex (DSX) transcription factor to demonstrate the predictive power of our method.

⋆ Equal contribution

1 Introduction

Cell function, fitness, and survival depend on a complex regulatory program involving interactions between genes and their regulators. Regulatory relationships between transcription factors (TFs) and genes they regulate constitute a gene regulatory network (GRN) that is often represented as a directed bipartite graph. Several experimental and computationally derived types of evidences can be used to infer the topology of such regulatory network including genome-wide chromatin immunoprecipitation (ChIP), gene expression profiling, and motif analysis. Complementing the topology of a GRN, a further level of understanding can be obtained by modeling the quantitative relation between TF activities and expression of their target genes. Specifically, given expression data obtained by engineered perturbations of a reference state, or by tracing expression changes over a number of naturally occurring conditions, the goal is to model the expression changes as a function of changes in activities of TFs when the underlying GRN topology is available. In particular, network component analysis (NCA) has proven to be a powerful method for such modeling [1, 2, 3, 4]. The essence of NCA approach is the assumption that the expression of genes can be modeled by linear combination of TF activities [5]. TF activity is a hidden parameter of each TF that the method infers from the data. While the assumption of linearity of the effects is obviously a simplifying one, it provides a good first approximation and makes the problem tractable.

An important drawback of the NCA approach is that it requires the topology of the GRN to be known. Several computational methods which integrate diverse functional genomics data sets, were developed to infer GRNs and investigate gene regulation at the systems level [6, 7]. Yet, current knowledge of the topology of regulatory networks is not complete, even for simple unicellular organisms such as yeast [8] and for most organisms construction of a regulatory network has not even been attempted. In addition these networks are typically context-agnostic, namely, they were constructed without considering tissue type, development stage, and other relevant conditions. However, for very closely related organisms their regulatory networks can be assumed to be rather similar due to evolutionary conservation. Similarly, for any specific organism, the regulatory interaction in different tissues are expected to overlap significantly. This motivates the need for developing a method that can utilize a network constructed for a closely related organism, stage, or tissue as a starting point for constructing a tissue, stage or organism specific network. Indeed, there is an increasing recognition of the importance of tissue specific analyses and tissue specific networks [9, 10].

To address this challenge we developed Network Rewiring using EXpression (NetREX), a new mathematically rigorous method, that builds on the linear model utilized by the NCA method, but without the assumption that the topology of the regulatory network is fixed. That is, unlike previous methods, NetREX does not restrict the structure of the regulatory network to be hardwired but instead utilizes expression data from a set of perturbations performed in a given context and a prior network that is assumed to be related to the target network by limited changes in the topology to construct a context specific network. We remark that allowing for rewiring in the topology of the prior network adds a whole new level of complexity. Specifically, we use ℓ_0 norm to directly handle the number of removed and newly added edges as well as induce sparse solutions in our formulation. Unlike the widely used strategy, which is replacing the non-convex ℓ_0 norm by its convex relaxation ℓ_1 norm, we focus on the harder problem involving ℓ_0 norm and provide a number of rigorous derivations and results allowing us to adopt the recently proposed Proximal Alternative Linearized Maximization (PALM) algorithm [11]. In addition, we also proved the convergence of the NetREX algorithm.

To evaluate method's performance we first tested NetREX on simulated data. Specifically, we analyzed its performance as the function of the noise in the prior network and in the gene expres-

sion data. We found that under the assumptions of the model, NetREX is able to dramatically improve the accuracy of the regulatory networks as long as the prior network and the gene expression are not very noisy. After testing the method on simulated data, we used NetREX for constructing regulatory networks for adult female flies. We used the network constructed in [6] as the prior network. This network was build by integrating diverse data sets including TF binding, evolutionarily conserved sequence motifs, gene expression across developmental time-course, and chromatin modification data sets. The topology of the network provides an initial wiring diagram that includes TF-target gene interactions predicted from data available at the time of network construction disregarding the context. Starting with this network, we utilized a new expression data set that we collected for adult female flies where perturbations in expression were achieved by genetic deletions. Specifically, the gene expression data of adult females are from 99 hemizygotic lines (deletion/+) of the *Drosophila* deletion collection (DrosDel) project [12, 13]. To evaluate the resulting networks we used a previously applied method [6] to access biological relevance of the networks by using Gene Ontology (GO) annotations [14] and physical protein-protein interactions (PPIs). We compared NetREX with several methods including a correlation based algorithm [15] and GENIE3 [16], the best performer in the DREAM4 *In Silico Multifactorial challenge* [15] We observed dramatic improvements in terms of fold enrichment comparing to all competing algorithms. Subsequently we asked how well the method predicts targets of the TF that have specific roles in adult flies and those targets would be difficult to identify based on cell lines or embryonic data that were predominately used in the construction of the prior network. For this analysis we focused on the Doublesex transcription factor (DSX) whose predicted targets have been recently elucidated [17]. We showed that the target genes predicted by NetREX are in good agreement with the experimentally identified targets.

Our method addresses an important challenge in analysis of gene regulation. It can be applied in many diverse setting including construction of condition specific GRNs and networks for organisms related to a model organism where a preliminary regulatory network exits. As a spin-off of these studies we also developed mathematical underpinning allowing to adopt Proximal Alternative Linearized Maximization (PALM) algorithm to the context of the ℓ_0 elastic net.

2 Mathematical Underpinning of the NetREX method

Before describing the mathematical foundations behind the NetREX method, we provide a brief overview of the traditional (static) NCA method and its various implementations. Next we introduce the formula for the objective function in our NetREX method. Importantly, the objective function is non-convex and non-smooth because of using ℓ_0 norm in our formulation. Rather than relaxing the problem by replacing the non-convex ℓ_0 norm with the convex ℓ_1 norm, we have directly solved the more challenging problem with ℓ_0 norm by adopting the recently proposed Proximal Alternative Linearized Maximization (PALM) algorithm [11] to the original formulation of the problem.

2.1 Network Component Analysis (NCA) and it implementations

The main principle of the NCA is to explain the expression of each gene as a linear combination of activities of TFs regulating that gene weighted by their control strength that exerts on the gene. In case of NCA, the topology of the bipartite GRN is provided as a part of the input. Formally, let $E \in \mathbb{R}^{N \times L}$ be the matrix of expression data of N genes in L experiments. NCA is a special case of a more general problem which is to express E as

$$E = SA + \Gamma, \tag{1}$$

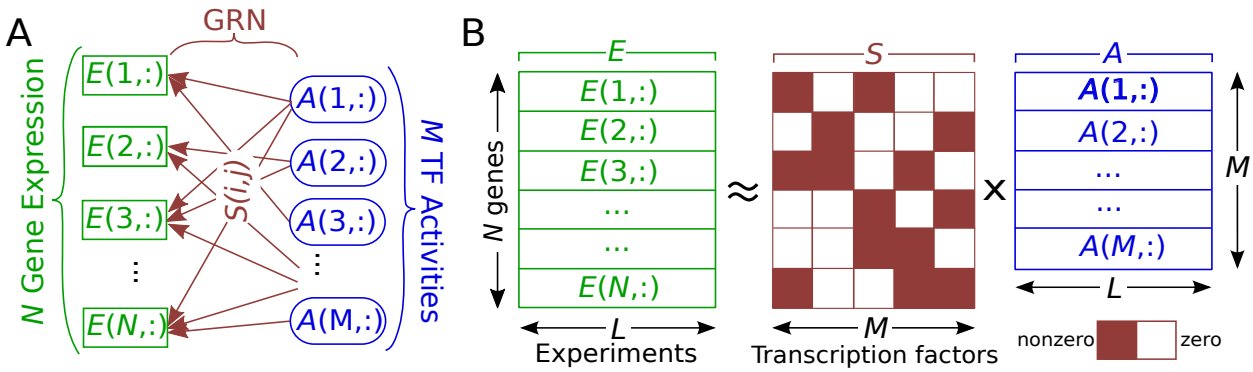


Fig. 1: The NCA model. (A) The graph representation of NCA. $E(i,:)$ is the expression of gene i over L experiments and $A(i,:)$ is the activity of TF i over the same L experiments. $S(i,j)$ is the control strength from TF j to gene i. (B) The algebraic formulation of NCA. E, S and A in (B) correspond to E, S and A in (A).

where $S \in \mathbb{R}^{N\times M}$ is a weighted adjacency matrix of the bipartite GRN $\mathcal{G}(TF, TG, \mathcal{S})$ such that the edges of $\mathcal{G}$ in the edge set $\mathcal{S}$ connect transcription factors in the M element set TF to target genes from the N element set TG. Specifically, for target gene i and transcription factor j, weight $S(i,j)$ defines the control strength that transcription factor j exerts on gene i. The rows of $A \in \mathbb{R}^{M\times L}$ represent the (hidden) TF activities of each TF over the set of experiments, and $\Gamma \in \mathbb{R}^{N\times M}$ represents the noise (Fig. 1).

Many mathematical techniques, such as principle component analysis (PCA), independent component analysis (ICA), non-negative matrix factorization (NMF) [18] and sparse coding (SP) [19], can be used to determine the decomposition of E specified in (1) (for NMF, E needs to be normalized to a non-negative matrix). However, PCA and NMF [20] are unable to find a decomposition of E when $M > L$ (i.e. the number of TFs is larger than the number of experiments). Moreover, PCA and ICA hinge on the assumptions of orthogonality and independence between the signals, which may not hold for TF activities (rows of A). In addition, all of them can not utilize the prior knowledge of the GRN $\mathcal{G}$. In contrast, NCA [5, 21, 22, 23, 24, 25] can deal with the situation when $M > L$, make no assumptions on TF activities and is able to take full advantage of the prior knowledge of the GRN $\mathcal{G}$. Specifically, NCA aims to uncover the matrix A describing the hidden regulatory activities of TFs and matrix S describing control strengths of each TF on target genes assuming that the structure S_0 (unweighted adjacency matrix) of the underlining GRN $\mathcal{G}_0 = (TF, TG, \mathcal{S}_0)$ is known. That is only the entries of S that correspond to edges in $\mathcal{S}_0$ can be non-zero (formally $\text{Support}(S) = \text{Support}(S_0)$, where $\text{Support}(S)$ denotes the support of S, i.e. the positions of its non-zero entries.). Thus NCA recovers the TF activities A and their control strength S by solving the following optimization problem with only the expression data E and the structure S_0 of $\mathcal{G}_0$ as inputs.

$$\begin{aligned} &\min_{S,A} \frac{1}{2}\|E - SA\|_F^2 \\ &s.t.\ \text{Support}(S) = \text{Support}(S_0), \\ &\quad \|S\|_\infty \le a,\ \|A\|_\infty \le b, \end{aligned} \tag{2}$$

where $\|S\|_\infty = \max_{i,j} |S(i,j)|$. The first constraint in the above formulation restricts the structure of the regulatory network $\mathcal{G}$ represented by matrix S to be exactly the same to the input regulatory network $\mathcal{G}_0$. And the rest of the constraints aim to ensure the elements in A and S remain within the domain of biologically sensible values.

The first method [5] to solve problem (2) can provide a unique solution only if the following conditions are met: (i) the matrix S should have full-column rank; (ii) each column of S should have at least $M - 1$ zeros; (iii) the matrix A should have full row-rank. Under these conditions,

S and A are estimated using an iterative two-step least-squares algorithm [5]. Tran et al. [21] expanded NCA by allowing the specification of the zero pattern of A as well as S. Galbraith et al. [22] modified the NCA method by revising the third criterion for NCA which cannot be tested before solving the problem. Chang et al. [23] treated NCA as an unconstrained optimization problem and employed singular value decomposition (SVD) to find a closed form solution for S without time-consuming iterations. Jacklin et al. [24] also proposed a non-iterative algorithm for NCA resorting to convex optimization methods. All these methods are vulnerable to the presence of small number of outliners in expression data. To deal with these outliers, Noor et al. [25] proposed ROBust Network Component Analysis (ROBNCA) where an additional sparse matrix was used for explicitly modeling the outliers.

2.2 The formulation of the optimization problem behind NetREX

Independently on numerous variants of the NCA, the assumption that the GRN must be known in advance is a significant drawback of the NCA method. NetREX relaxes this restriction under the assumption that a prior regulatory network that is not too far from the underlining true regulatory network is given. Therefore, it is possible to recover the underlining regulatory network by limited changes to the prior network. Note that this is a very reasonable assumption in many practical applications where the prior network could come from a related organism or a related tissue or even from the same organism but without sufficient data. Additionally, to guide the network reconstruction, we assume that genes with highly correlated expression are likely to be regulated by the same TFs. The correlations between genes can be encoded in the gene correlation network $\mathcal{G}^E$ constructed based on gene expression data E. Thus in the new optimization problem (3) we remove the constraint that the structure of the network is fixed ($\text{Support}(S) = \text{Support}(S_0)$) but introduce a penalty term that limits the number of added and removed edges with respect to the prior network, along with the terms encouraging consistent treatment of co-expressed genes, and network sparsity. We devote the rest of this subsection to explaining the roles of the added terms.

$$\begin{aligned} &\min_{S,A} \frac{1}{2}\left\|E - SA\right\|_F^2 + \lambda\left(\left\|S_0\right\|_0 - \left\|S \odot S_0\right\|_0 + \left\|S \odot \bar{S}_0\right\|_0\right) + \kappa \text{tr}(S^T L S) + \eta\left\|S\right\|_0 + \xi\left\|S\right\|_F^2 + \mu\left\|A\right\|_F^2 \\ &s.t.\ \left\|S\right\|_\infty \leq a,\ \left\|A\right\|_\infty \leq b. \end{aligned} \tag{3}$$

where $\lambda, \kappa, \eta, \xi, \mu$ are the parameters controlling the strength of the corresponding terms.

The term controlled by λ restricts the number of edge changes. Here $\bar{S}_0$ is the adjacency matrix of the complement graph of $\mathcal{G}_0$ and therefore $\bar{S}_0 + S_0 = \mathbf{1}_{N\times M}$ and $\left\|X\right\|_0$ is the ℓ_0 norm that computes the number of non-zero entries in X. $\odot$ is the Hadamard product. We note that $\left\|S_0\right\|_0 - \left\|S \odot S_0\right\|_0$ denotes the exact number of regulations removed from $\mathcal{G}_0$ and $\left\|S \odot \bar{S}_0\right\|_0$ is the number of regulations added to the prior network $\mathcal{G}_0$. λ controls the change in topology of the regulatory network. Larger λ indicates that only small number of edges can be added and removed controlling how far our predicted network $\mathcal{G}$ is from the prior network $\mathcal{G}_0$.

The term controlled by κ (the graph embedding term [26]) encourages $S(i,k)$ and $S(j,k)$ to have similar control strength if genes i and j are correlated. In Supplementary Materials A.1 we provide derivations demonstrating that:

$$\frac{1}{2}\sum_{i,j}\sum_{k} W(i,j)\left(S(i,k) - S(j,k)\right)^2 = \text{tr}(S^T L S), \tag{4}$$

where tr() is the trace of a matrix and W and L are the adjacency matrix and the Laplacian matrix of the correlation network $\mathcal{G}^E$, respectively.

The term of equation (3) that is controlled by parameter η encourages sparsity of the final network (note the ℓ_0 norm that computes the number of non-zero elements). However we note that there might exist correlations between TF activities (rows of A), which would imply relations between TFs and enforcing the sparsity constrain might weaken them. This means that, for a gene, only one TF can be selected from a group of TFs whose activities are highly correlated even though all TFs in the group regulate the gene. Therefore, we have an additional term (controlled by parameter ξ) using Frobenius norm to ensure that all regulating TFs have non-zero values in S. For the reader familiar with the elastic net model, we point that $\eta \|S\|_0 + \xi \|S\|_F^2$ is analogous to ℓ_1 elastic net [27], and we can refer to it as ℓ_0 elastic net.

Finally, the last term controlled by the variable μ enforces smoothness of activities in A by avoiding many elements in A reach to the limit $\{-b, b\}$.

After some linear algebra (Supplementary Materials A.2), we obtain our final formulation as follow. We require $\eta - \lambda \geq 0$, otherwise the above formulation would preserve all regulations in $\mathcal{G}_0$.

$$\begin{aligned} &\min_{S,A} \frac{1}{2} \|E - SA\|_F^2 + (\eta - \lambda) \|S \odot S_0\|_0 + (\eta + \lambda) \left\|S \odot \bar{S}_0\right\|_0 + \kappa \mathrm{tr}(S^T L S) + \xi \|S\|_F^2 + \mu \|A\|_F^2 \\ &s.t.\ \|S\|_\infty \leq a,\ \|A\|_\infty \leq b. \end{aligned} \tag{5}$$

2.3 Solving the optimization problem underlying the NetREX algorithm

Our algorithm to solve optimization problem (5) relies on the recently proposed proximal alternative linearized maximization (PALM) [11] algorithm. The PALM method can solve a general optimization problem formulated as

$$\min : H(S, A) = F(S, A) + \Phi(S) + \Psi(A) \text{ over } S \in \Upsilon, A \in \Omega, \tag{6}$$

where $F(S, A)$ has to be smooth but $\Phi(S)$ and $\Psi(A)$ do not need to have the convexity and smoothness properties. Υ and Ω are constraint sets for S and A, respectively. The PALM algorithm alternatively applies technique known as proximal forward-backward scheme to both S and A. Specifically, at iteration k, the proximal forward-backward mappings of $\Phi(S)$ and $\Psi(A)$ on $S \in \Upsilon$ and $A \in \Omega$ for given S^k and A^k are the solutions for the following sub-problems, respectively.

$$S^{k+1} \in \arg\min_{S \in \Upsilon} \left\{ \left\langle S - S^k,\ \nabla_S F(S^k, A^k) \right\rangle + \frac{c^k}{2} \left\| S - S^k \right\|_F^2 + \Phi(S) \right\}; \tag{7a}$$

$$A^{k+1} \in \arg\min_{A \in \Omega} \left\{ \left\langle A - A^k,\ \nabla_A F(S^{k+1}, A^k) \right\rangle + \frac{d^k}{2} \left\| A - A^k \right\|_F^2 + \Psi(A) \right\}, \tag{7b}$$

where $\langle X, Y \rangle = \mathrm{tr}(X^T Y)$, c^k and d^k are positive real numbers and $\nabla_S F(S^k, A^k)$ is the derivative of $F(S, A^k)$ with respect to S at point S^k for fixed A^k and $\nabla_A F(S^{k+1}, A^k)$ is the derivative of $F(S^{k+1}, A)$ with respect to A at point A^k for fixed S^{k+1}. It has been proven that the sequence $\left\{(S^k, A^k)\right\}_{k \in \mathbb{N}}$ generated by PALM converges to a critical point when it is bounded [11].

Casting our optimization problem (5) into the PALM algorithm framework (6) introduced above, we have $F(S, A) := \frac{1}{2} \|E - SA\|_F^2 + \kappa \mathrm{tr}(S^T L S)$, $\Psi(A) := \mu \|A\|_F^2$ and $\Phi(S) := (\eta + \lambda) \left\|\bar{\mathbb{S}}_0 \odot S\right\|_0 + (\eta - \lambda) \|\mathbb{S}_0 \odot S\|_0 + \xi \|S\|_F^2$. The constraint sets Υ and Ω are respectively $\Upsilon = \{S \mid \|S\|_\infty \leq a\}$ and $\Omega = \{A \mid \|A\|_\infty \leq b\}$. We note that $F(S, A)$, $\Psi(A)$ and $\Phi(S)$ satisfy the requirements of the PALM algorithm. Namely, $F(S, A)$ is smooth, $\Psi(A)$ is convex and smooth but, as allowed in the PALM approach, $\Phi(S)$ is non-convex and non-smooth. Hence, we can apply the PALM algorithm to our problem as long as we can efficiently solve the proximal forward-backward mappings for our

specific $\Phi(S)$ and $\Psi(A)$. Proving that we can actually do it is mathematically the most challenging component of the development of the method. Due to technicality of the derivations we leave most of them to the supplement and in what follows we only point to the most critical components of the argument.

It is easy to confirm that the NetREX problem (5) can be solved by alternatively applying the following proximal forward-backward mappings (8a) and (8b), which are derived from (7a) and (7b) by casting our specific $F(S, A)$, $\Phi(S)$, $\Psi(A)$, Υ and Ω and some linear algebra (derivations can be found in Supplementary Materials A.3).:

$$S^{k+1} \in \arg\min_{\|S\|_\infty \leq a} \left\{ \Phi(S) + \frac{c^k}{2} \left\| S - U^k \right\|_F^2 \right\}; \tag{8a}$$

$$A^{k+1} \in \arg\min_{\|A\|_\infty \leq b} \left\{ \Psi(A) + \frac{d^k}{2} \left\| A - V^k \right\|_F^2 \right\}, \tag{8b}$$

where

$$U^k = S^k - \frac{1}{c^k} \nabla_S F(S^k, A^k) \text{ and } V^k = A^k - \frac{1}{d^k} \nabla_A F(S^{k+1}, A^k). \tag{9}$$

The derivatives $\nabla_S F(S^k, A^k)$ and $\nabla_A F(S^{k+1}, A^k)$ can be computed by

$$\nabla_S F(S^k, A^k) = (S^k A^k - E)(A^k)^T + 2\kappa L S^k \text{ and } \nabla_A F(S^{k+1}, A^k) = (S^{k+1})^T (S^{k+1} A^k - E), \tag{10}$$

which are Lipschitz continuous with $L(A^k) = \left\| A^k (A^k)^T \right\|_F + 2\kappa \|L\|_F$ and $L(S^{k+1}) = \left\| (A^k)^T A^k \right\|_F$ as Lipschitz constants, respectively. As suggested by [11], we set $c^k = \max\left\{v,\ L(A^k)\right\},\ v > 0$ and $d^k = \left\{v,\ L(S^{k+1})\right\},\ v > 0$ to make sure the formulas in (9) are well defined.

The closed form solution of the proximal forward-backward mapping (8a) can be obtained based on Proposition 1, the Proximal Mapping of ℓ_0 Elastic Net Under $\|\|_\infty$ Constraint Proposition, and its corollary (Corollary 1). The proposition and the corollary and their proofs can be found in the Supplementary Materials B.1 and B.2. We emphasize that Proposition 1 provides the closed form solution for the proximal mapping of ℓ_0 elastic net under $\|\|_\infty$ constraint and thus it has broader applications to diverse feature selection approaches [28, 29].

With the help of Proposition 1 and Corollary 1, (8a) can be efficiently computed by (11a). And (8b) can be computed by (11b).

$$S^{k+1} \in \text{prox}_{\|\cdot\|_\infty \leq a} \left(U^k, \frac{2(\eta + \lambda)}{c^k}, \frac{2(\eta - \lambda)}{c^k}, \frac{2\xi}{c^k} \right); \tag{11a}$$

$$A^{k+1} = \mathbb{P}_{\|\cdot\|_\infty \leq b} \left(\frac{1}{1 + \frac{2\mu}{d^k}} V^k \right), \tag{11b}$$

where the definitions of $\text{prox}_{\|\cdot\|_\infty \leq a}(\cdot)$ and $\mathbb{P}_{\|\cdot\|_\infty \leq b}(\cdot)$ can be found in Corollary 1 and Proposition 1, respectively. The derivations of (11a) and (11b) can be found in the Supplementary Materials A.4.

We now have all the ingredients for our NetREX algorithm. Hence, we describe the NetREX algorithm in Algorithm 1 in Supplementary Materials C. We note that the constraints for both S and A ($\|S\|_\infty \leq a$ and $\|A\|_\infty \leq b$) make sure that the sequence $\left\{(S^k, A^k)\right\}_{k \in \mathbb{N}}$ is bounded. Thus we state that the sequence produced by the NetREX algorithm converges to a critical point of the optimization problem (5), which is described in Proposition 2 in Supplementary Materials B.3.

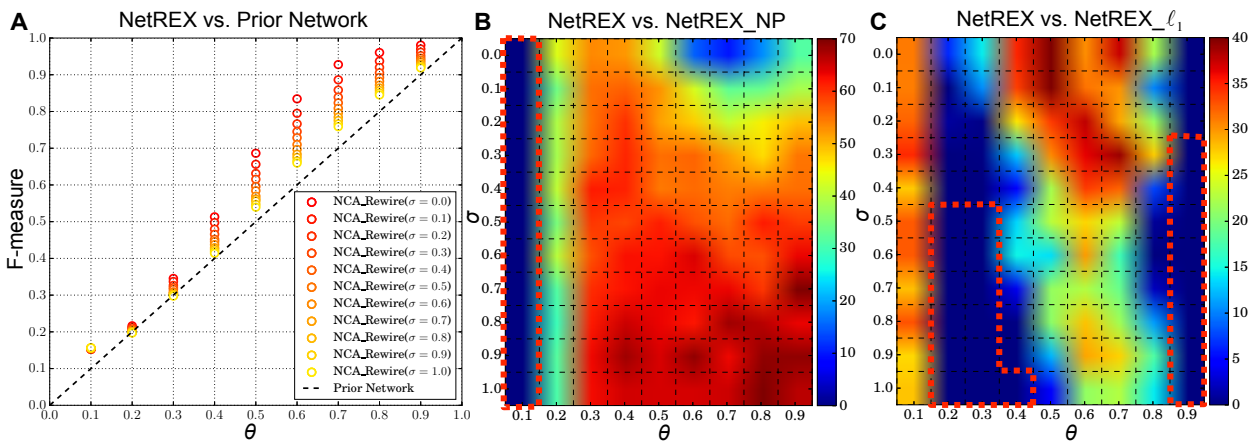


Fig. 2: (A) Comparison between F-measures of the networks predicted by NetREX and prior networks. The x-coordinate denotes percentages of true edges in prior networks and the black dashed line denotes the F-measures of the prior networks. The circles are the average F-measures of the networks predicted by NetREX under different σ and θ over 50 random inputs. (B) Comparison between F-measures of the networks predicted by NetREX and NetREX_NP. The color in each dashed block indicates the $-\log$ p-value for corresponding (σ, θ), where the p-values are obtained from one-sided paired t-test between F-measures of the compared algorithms. The warmer the color is, the significantly larger F-measures of the networks predicted by NetREX are than NetREX_NP. The red dashed line circles the (σ, θ) pairs where NetREX_NP achieves larger F-measure at significant level 0.01. (C) Comparison between F-measures of the networks predicted by NetREX and NetREX_ℓ_1. The color coding is the same as in the panel (B).

3 Validation and Experimental Results

3.1 Results on Simulated Data

To validate our approach, we applied NetREX to the simulated data generated based on linear model (1). We first randomly generated the ground truth adjacency matrix S of the regulatory network $\mathcal{G}(TF, TG, \mathcal{S})$ and TF activities A. Then, the simulated expression data was generated as following

$$E(i,j) = \sum_p S(i,p)A(p,j) + \Gamma(i,j), \tag{12}$$

where $\sum_p S(i,p)A(p,j)$ is the noiseless data arising from known A and S matrices and the noise $\Gamma(i,j) \sim \mathbf{N}(0, \sigma^2)$ obeys a normal distribution with 0 mean and σ^2 variance. We assigned the prior network $\mathcal{G}_0$ the same number of edges as the ground truth network $\mathcal{G}$ has but only θ percentage of edges in $\mathcal{G}_0$ are true edges. We can tune the difficulty of the network rewiring task by using different σ and θ.

We compared NetREX with its two natural variants on the simulation data. The first variant is NetREX_NP (NetREX with No edge Perturbation term) that has the same formulation as NetREX but with $\lambda = 0$. The difference between NetREX and NetREX_NP is that NetREX penalizes the number of edges added and removed from the prior network but NetREX _NP does not. Here we want to mention that NetREX _NP and sparse coding have similar formulations (Supplementary Materials E.1). The other related algorithm in our comparison is NetREX_ℓ_1, which estimates the ℓ_0 norm in NetREX using ℓ_1 norm. We note, that substituting ℓ_0 norm by ℓ_1 norm makes the sub-problems convex and thus easier to solve. The implementation of these two algorithms are introduced in Supplementary Materials E.1.

We evaluated the performance of the compared algorithms in terms of F-measure (defined in Supplementary Materials D.1). To avoid the effect of parameter selection, for each algorithm, under certain noise level (σ, θ), we first found its optimal parameters in terms of F-measure on one simulated data set through grid search. Then we ran the algorithm on another 50 randomly generated simulated data sets under the same (σ, θ) using its optimal parameters. We can further

test whether one method is statistically better than another method under a specific noise level by computing the p-value from one-side paired t-test between their 50 paired F-measures.

The comparisons between NetREX and others are shown in Fig. 2. Fig. 2A shows the comparison between networks predicted by NetREX and the prior networks, in which we found a tendency that when the expression data is less noisy (σ is small) and the prior network is closer to the ground truth (θ is large), the network predicted by NetREX achieves higher F-measures than the prior networks. Additionally, we note that NetREX exhibits, by a larger margin, higher F-measure comparing to the prior networks after $\theta \geq 0.3$. However, for $\theta < 0.3$ the networks predicted by NetREX only marginally better than the prior network, which also implies that if we use random networks that do not have much overlap with the ground truth as the prior networks, we can not obtain promising results. The comparison between NetREX and NetREX_NP is displayed in Fig. 2B. We note that NetREX significantly outperforms NetREX_NP after $\theta > 0.1$. In Fig. 2C, we observe that NetREX_ℓ_1 performs better in certain cases where the noise in the expression data is large (σ is large) because ℓ_1 norm is robust to noise. However, for most of the noise levels, NetREX achieved significantly higher F-measures comparing to NetREX_ℓ_1 demonstrating that ℓ_0 norm is superior to ℓ_1 norm on selecting sparse contributing components.

3.2 Results on Real Experimental Data from DrosDel Study

Next we applied NetREX to gene expression data in the adult female flies from 99 hemizygotic lines (deletion/+) of the *Drosophila* deletion collection (DrosDel) project covering 68% of chromosome 2L. Specifically, in each of the DrosDel lines, a different chromosomal fragment has been deleted leaving the organism with only one copy of genes for the deleted region [13]. We used the network constructed in [6] as the prior network, which is constructed through integrating diverse functional genomics data sets (such as transcription factor (TF) binding, evolutionary conserved sequence motifs and etc.) in a supervised learning framework. The data used in [6] typically comes from cell lines and expression profiles of developmental stages. NetREX predicted regulatory networks for the adult female flies. And we verified these networks using GO functional annotations [14], physical protein-protein interactions (PPIs) and *Drosophila* Doublesex transcription factor (DSX) target genes [17].

We compared our predicted networks with the prior networks and the TF-Gene correlation networks that were built based on the Pearson coefficient between expression of TFs and genes using previously reported expression measurements in DrosDel data [30]. Furthermore, we compared with GENIE3 [16], the best performer in the DREAM4 *In Silico Multifactorial challenge* [15], which predicts GRNs using only expression data. We also tried to compare with NMF based algorithms, which can not be applied because the dimensional requirement of NMF [20] is not satisfied (i.e. the number of TFs in the target GRN is larger than the number of hemizygotic lines in DrosDel data). To demonstrate performance of NetREX under different parameters, and to choose the parameters in a manner that does not depend on tested data (GO annotations and PPIs), we developed a simple heuristic that ranks the models performance using a quality score based on the objective function (5) (Supplementary Materials E.3). We then used top twenty models with respect to this ranking. In addition, since performance of the correlation based algorithm and GENIE3 might depend on different cut-offs, in the evaluation we showed the performance of the TF-Gene correlation networks and networks predicted by GENIE3 with different cutoffs (Supplementary Materials E.3). Finally, we note that, unlike other networks in this comparison, the co-expression network is not a regulatory network, however, we embedded the information about this network in our objective function, thus we need to show that the good performance of our method is not merely reflection of embedding this information in the objective function.

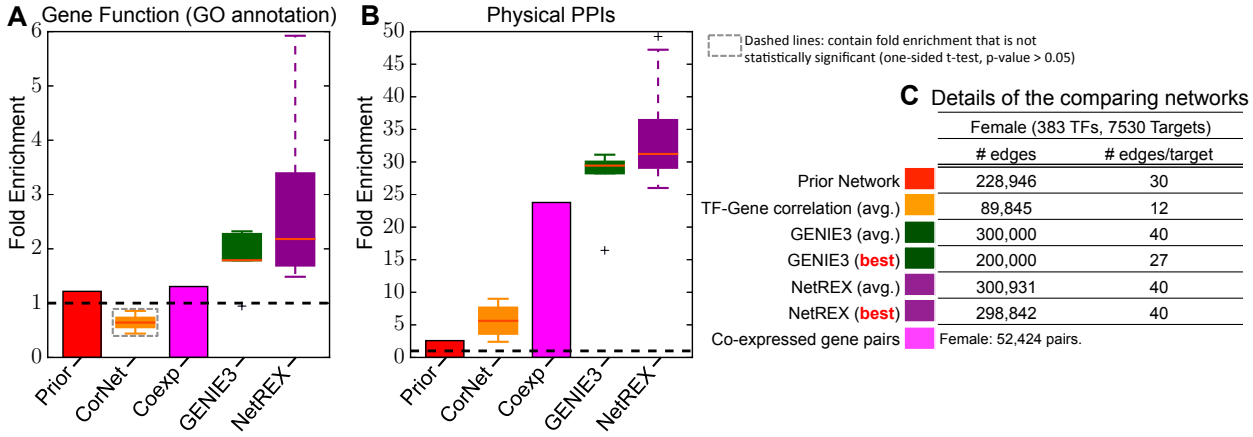


Fig. 3: Functional enrichment and the details of the compared networks. For the correlation based algorithm, GENIE3 and NetREX, we showed the performance of these three algorithms under different parameters (details in Supplementary Materials E.3). (A) is functional enrichment of co-regulated genes for annotation with similar functions. (B) is the fold enrichment of physical interactions between co-regulated genes. (C) is the details of all compared networks. The networks predicted by NetREX show stronger enrichment for both gene functions and physical PPIs.

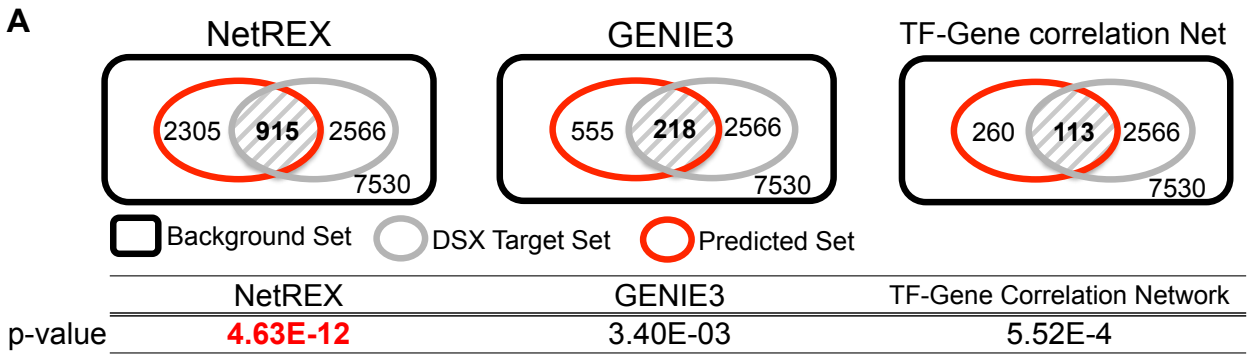
Functional Enrichment of the Predicted Regulatory Network We assessed the biological relevance of the predicted regulatory networks through checking whether genes co-regulated by similar TFs exhibit similar functional properties. We used the measures (a brief review is in Supplementary Materials D.2 and D.3) proposed in [6] to evaluate the enrichment of co-regulated genes in terms of GO functional annotations [14] and experimentally derived physical Protein Protein interactions (PPIs) extracted from DroID database [31], respectively. In addition to the prior networks, the TF-Gene correlation networks and the networks predicted by GENIE3, we also computed the same measures for the co-expression network inputted as the graph embedding term in NetREX.

First, we examined whether co-regulated genes have similar GO annotations. The comparison results are illustrated in Fig. 3A. NetREX clearly outperforms all other approaches demonstrating benefit of using both the prior network and the condition specific gene expression data.

Next, we evaluated whether physically interacting genes are tend to be co-regulated in the respective networks. Fig. 3B shows the PPI enrichment comparison. Using this enrichment as a measure of network quality as it has been proposed in [6] we observed that NetREX also outperforms all other methods.

Agreement with DSX Targets To further validate the predictions obtained from different methods, we concentrated on target genes of *Drosophila* Doublesex transcription factor (DSX) which involves in the sex determination system as different isoforms in flies [32]. Recently, Clough et al. [17] reported a rich set of DSX targets based on a series of genome-wide experiments and analysis. Thus, we checked how well the predicted DSX targets of NetREX are in agreement with genes identified in [17]. We selected networks giving the best fold enrichment in both GO annotation and physical PPIs. As shown in Fig. 4A, NetREX outperforms other methods and there is statistically significant overlap between targets predicted by NetREX and targets inferred from experimental data (p-value are computed by the hyper geometric test).

Prediction and validation of DSX regulators We then used NetREX to predict the regulators of *doublesex* gene (*dsx*). Fig. 4B illustrates the results of our prediction for *dsx* in female flies. Our predictions include multiple probable transcription factors that for example, *Trl* (*Trithorax-like*) gene encodes *Drosophila* GAGA factor that has reported roles in sex chromosome dosage compensation [33]. In addition, the *rnet* locus is required to repress male courtship behaviors in



	NetREX	GENIE3	TF-Gene Correlation Network
p-value	**4.63E-12**	3.40E-03	5.52E-4

B Predicted *dsx* regulators

Symbol	*klu*	***CG6175***	*Adf1*	*jing*	*Mrtf*	*cwo*	*CG9932*	*ci*	*jim*
Strength	0.203	-0.202	0.202	-0.187	0.174	0.156	0.155	0.153	0.138
Symbol	***bab2***	*crol*	*CG4404*	*CG7987*	*vri*	***bbx***	*su(Hw)*	*CG15073*	*pnt*
Strength	0.137	-0.132	-0.124	0.117	-0.112	0.109	0.107	-0.100	0.097
Symbol	*Hnf4*	*CG32772*	*emc*	*Snoo*	*Hr78*	*CG11247*	***retn***	*Atf3*	***Trl***
Strength	-0.094	0.093	0.093	0.090	-0.090	-0.084	0.071	0.069	0.069

Fig. 4: (A) Comparison of agreement with DSX Targets using Venn diagrams with p-values. P-values are computed by the hyper geometric test. NetREX achieved the lowest p-value. In the prior network, there are 2 target genes of DSX but none of them overlapped with the target set reported in [17]. (B) *dsx* regulators predicted by NetREX for female flies. We ranked the regulators based on the absolute values of their control strength. Genes in red have evidence in literature.

females [34] while *bab1* and *bab2* genes have overlapping functions controlling sex-specific morphology [35]. We also observed many other predicted genes in the list have sex-specific expression, as like *bbx* (*boby sox*) and *CG6175*, which demonstrate testis-specifically repressed, but ovary-expressed patterns [36].

4 Conclusion

Regulatory networks embed key information needed for modeling and interpreting experimental data. Currently, regulatory networks are constructed by combining information from various tissues, stages, and conditions [6, 37, 38] to obtain a context-agnostic network. However the importance of constructing tissue / stage specific networks is now being increasingly recognized [9, 10]. And constructing such network from scratch for every relevant tissue and/or condition is not realistic. In addition, data obtained from different tissues and conditions might, provide additional information that context specific analysis might not be sufficiently empowered to detect. For these reasons it is fundamental to be able to utilize regulatory networks constructed in context independent way as a starting point for context specific network construction. The NetREX method proposed in this paper fulfills this critical need. Importantly, our construction is mathematically rigorous. We proved convergence of the method and validate its performance on both simulation data and real world data. The experimental results demonstrate that NetREX is capable to recover the biological meaningful condition specific TF-gene regulatory networks.

5 Acknowledgments

This work was supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine (YW, TMP, DYC) and National Institute of Diabetes and Digestive and Kidney Diseases National (HK, BO). The authors thanks Roded Sharan, and the members of Przytycka and Oliver groups for helpful discussions.

References

[1] Ashish Misra and Ganesh Sriram. "Network component analysis provides quantitative insights on an Arabidopsis transcription factor-gene regulatory network." In: *BMC Systems Biology* 7.1 (2013), p. 126.

[2] Mario L. Arrieta-Ortiz et al. "An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network." In: *Molecular systems biology* 11.11 (2015), p. 839.

[3] Joerg Martin Buescher et al. "Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism." In: *Science (New York, NY)* 335.6072 (2012), pp. 1099–1103.

[4] Ziv Bar-Joseph, Anthony Gitter, and Itamar Simon. "Studying and modelling dynamic biological processes using time-series gene expression data". In: *Nature Reviews Genetics* 13.8 (2012), pp. 552–564.

[5] James C Liao et al. "Network component analysis: reconstruction of regulatory signals in biological systems." In: *Proceedings of the National Academy of Sciences of the United States of America* 100.26 (2003), pp. 15522–15527.

[6] Daniel Marbach et al. "Predictive regulatory models in Drosophila melanogaster by integrative inference of transcriptional networks". In: *Genome Research* 22.7 (2012), pp. 1334–1349.

[7] Noa Novershtern, Aviv Regev, and Nir Friedman. "Physical Module Networks: An integrative approach for reconstructing transcription regulation". In: *Bioinformatics* 27.13 (2011), pp. 177–185.

[8] Timothy R. Hughes and Carl G. de Boer. "Mapping yeast transcriptional networks". In: *Genetics* 195.1 (2013), pp. 9–36.

[9] Christopher Y. Park et al. "Tissue-aware data integration approach for the inference of pathway interactions in metazoan organisms". In: *Bioinformatics* 31.7 (2015), pp. 1093–1101.

[10] Casey S Greene et al. "Understanding multicellular function and disease with human tissue-specific networks." In: *Nature genetics* 47.6 (2015), pp. 569–76.

[11] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. "Proximal alternating linearized minimization for nonconvex and nonsmooth problems". In: *Mathematical Programming* 146.1-2 (2014), pp. 459–494.

[12] Edward Ryder et al. "The DrosDel collection: A set of P-element insertions for generating custom chromosomal aberrations in Drosophila melanogaster". In: *Genetics* 167.2 (2004), pp. 797–813.

[13] Edward Ryder et al. "The DrosDel deletion collection: A Drosophila genomewide chromosomal deficiency resource". In: *Genetics* 177.1 (2007), pp. 615–629.

[14] J. a. Blake et al. "Gene ontology consortium: Going forward". In: *Nucleic Acids Research* 43.D1 (2015), pp. D1049–D1056.

[15] Daniel Marbach et al. "Wisdom of crowds for robust gene network inference". In: *Nature Methods* 9.8 (2012), pp. 796+.

[16] Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. "Inferring regulatory networks from expression data using tree-based methods". In: *PLoS ONE* 5.9 (2010), pp. 1–10.

[17] Emily Clough et al. "Sex- and tissue-specific functions of drosophila doublesex transcription factor target genes". In: *Developmental Cell* 31.6 (2014), pp. 761–773. arXiv: `NIHMS150003`.

[18] Dd Lee and Hs Seung. "Algorithms for non-negative matrix factorization". In: *Advances in neural information processing systems* 1 (2001), pp. 556–562. arXiv: `0408058v1 [arXiv:cs]`.

[19] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. "Online dictionary learning for sparse coding". In: *Proceedings of the 26th International Conference on Machine Learning* (2009), pp. 1–8. arXiv: `0908.0050`.

[20] Yu-Xiong Wang and Yu-Jin Zhang. "Nonnegative Matrix Factorization: A Comprehensive Review". In: *IEEE Transactions on Knowledge and Data Engineering* 25.6 (2013), pp. 1336–1353.

[21] Linh M. Tran et al. "gNCA: A framework for determining transcription factor activity based on transcriptome: Identifiability and numerical implementation". In: *Metabolic Engineering* 7.2 (2005), pp. 128–141.

[22] Simon J. Galbraith, Linh M. Tran, and James C. Liao. "Transcriptome network component analysis with limited microarray data". In: *Bioinformatics* 22.15 (2006), pp. 1886–1894.

[23] Chunqi Chang, Zhi Ding, Yeung Sam Hung, and Peter Chin Wan Fung. "Fast network component analysis (FastNCA) for gene regulatory network reconstruction from microarray data". In: *Bioinformatics* 24.11 (2008), pp. 1349–1358.

[24] Neil Jacklin, Zhi Ding, Wei Chen, and Chunqi Chang. "Noniterative convex optimization methods for network component analysis". In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 9.5 (2012), pp. 1472–1481.

[25] Amina Noor et al. "ROBNCA: Robust network component analysis for recovering transcription factor activities". In: *Proceedings - IEEE International Workshop on Genomic Signal Processing and Statistics* 29.19 (2013), pp. 19–22.

[26] Mikhail Belkin and Partha Niyogi. "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation". In: *Neural Computation* 15.6 (2003), pp. 1373–1396. arXiv: `arXiv:1011.1669v3`.

[27] H Zou and T Hastie. "Regularization and variable selection via the elastic-net". In: *Journal of the Royal Statistical Society* 67 (2005), pp. 301–320.

[28] Nanne Aben, Daniel J. Vis, Magali Michaut, and Lodewyk F.a. Wessels. "TANDEM: a two-stage approach to maximize interpretability of drug response models based on multiple molecular data types". In: *Bioinformatics* 32.17 (2016), pp. i413–i420.

[29] Jishnu Das et al. "ENCAPP: elastic-net-based prognosis prediction and biomarker discovery for human cancers." In: *BMC genomics* 16.1 (2015), p. 263.

[30] Hangnoh Lee et al. "Effects of Gene Dose, Chromatin, and Network Topology on Expression in Drosophila melanogaster". In: *PLOS Genetics* 12.9 (2016), e1006295.

[31] Jingkai Yu, Svetlana Pacifico, Guozhen Liu, and Russell L Finley. "DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions." In: *BMC Genomics* 9 (2008), p. 461.

[32] B S Baker. "Sex in flies: the splice of life." In: *Nature* 340.6234 (1989), pp. 521–524.

[33] Anthony J. Greenberg, Judith L. Yanowitz, and Paul Schedl. "The Drosophila GAGA Factor is Required for Dosage Compensation in Males and for the Formation of the Male-Specific-Lethal Complex Chromatin Entry Site at 12DE". In: *Genetics* 166.1 (2004), pp. 279–289.

[34] Lynn M Ditch et al. "Drosophila retained/dead ringer is necessary for neuronal pathfinding, female receptivity and repression of fruitless independent male courtship behaviors." In: *Development (Cambridge, England)* 132.1 (2005), pp. 155–164.

[35] Jean-Louis Couderc et al. "The bric à brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila." In: *Development (Cambridge, England)* 129 (2002), pp. 2419–2433.

[36] Scott W. Robinson, Pawel Herzyk, Julian a T Dow, and David P. Leader. "FlyAtlas: Database of gene expression in the tissues of Drosophila melanogaster". In: *Nucleic Acids Research* 41.D1 (2013), pp. 744–750.

[37] Serdar Turkarslan et al. "A comprehensive map of genome-wide gene regulation in Mycobacterium tuberculosis". In: *Scientific data* 2 (2015), p. 150010.

[38] Nikolai Hecker et al. "Optimizing RNA structures by sequence extensions using RNAcop". In: *Nucleic Acids Research* (2015), gkv813.

[39] L. Grippo and M. Sciandrone. "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints". In: *Operations Research Letters* 26.3 (2000), pp. 127–136.

[40] Neal Parikh and Stephen Boyd. "Proximal Algorithms". In: *Foundations and Trends in Optimization* 1.3 (2013), pp. 123–231.

Supplementary Materials I:
Mathematical Derivations and Proofs

A Mathematical Derivations

A.1 The Graph Embedding

The derivation of the graph embedding term (4) is shown below.

$$
\begin{aligned}
&\frac{1}{2}\sum_{i,j}\sum_{k} W(i,j)\left(S(i,k)-S(j,k)\right)^2 \\
&=\frac{1}{2}\sum_{i,j} W(i,j)\left\|S(i,:)-S(j,:)\right\|_2^2 \\
&=\frac{1}{2}\left(\sum_{i,j} W(i,j)S(i,:)S(i,:)^T+\sum_{i,j} W(i,j)S(j,:)S(j,:)^T-2\sum_{i,j} W(i,j)S(i,:)S(j,:)^T\right) \\
&=\frac{1}{2}\left(\sum_{i} D(i,i)S(i,:)S(i,:)^T+\sum_{j} D(j,j)S(j,:)S(j,:)^T-2\sum_{i,j} W(i,j)S(i,:)S(j,:)^T\right) \\
&=\frac{1}{2}\left(2\mathrm{tr}(S^TDS)-2\mathrm{tr}(S^TWS)\right) \\
&=\mathrm{tr}(S^TLS),
\end{aligned}
\tag{13}
$$

where tr() is the trace of a matrix. W is the adjacency matrix for the gene co-expression network $\mathcal{G}^E$ and D is a diagonal matrix with the degree of every gene on its diagonal. L is the Laplacian matrix of $\mathcal{G}^E$.

A.2 Derivation from (3) to (5)

The key of the derivation is the following transformation.

$$
\begin{aligned}
&\lambda\left(\|S_0\|_0-\|S\odot S_0\|_0+\left\|S\odot\bar{S}_0\right\|_0\right)+\eta\|S\|_0 \\
=&\lambda\left(\|S_0\|_0-\|S\odot S_0\|_0+\left\|S\odot\bar{S}_0\right\|_0\right)+\eta\|S\odot\mathbf{1}\|_0 \\
=&\lambda\left(\|S_0\|_0-\|S\odot S_0\|_0+\left\|S\odot\bar{S}_0\right\|_0\right)+\eta\left\|S\odot(S_0+\bar{S}_0)\right\|_0 \\
=&\lambda\left(\|S_0\|_0-\|S\odot S_0\|_0+\left\|S\odot\bar{S}_0\right\|_0\right)+\eta\left\|S\odot S_0+S\odot\bar{S}_0\right\|_0 \\
=&\lambda\left(\|S_0\|_0-\|S\odot S_0\|_0+\left\|S\odot\bar{S}_0\right\|_0\right)+\eta\left(\|S\odot S_0\|_0+\left\|S\odot\bar{S}_0\right\|_0\right) \\
=&\lambda\|S_0\|_0+(\eta-\lambda)\|S\odot S_0\|_0+(\eta+\lambda)\left\|S\odot\bar{S}_0\right\|_0.
\end{aligned}
\tag{14}
$$

In the above derivation, we use the fact that $\mathbf{1}=S_0+\bar{S}_0$ and the property that $\left\|S\odot S_0+S\odot\bar{S}_0\right\|_0=\|S\odot S_0\|_0+\left\|S\odot\bar{S}_0\right\|_0$ because S_0 and $\bar{S}_0$ are complementary to each other. Replacing the corresponding term in (3) with the last line of (14), we have

$$
\begin{aligned}
&\min_{S,A}\ \frac{1}{2}\|E-SA\|_F^2+\lambda\|S_0\|_0+(\eta-\lambda)\|S\odot S_0\|_0+(\eta+\lambda)\left\|S\odot\bar{S}_0\right\|_0+\kappa\mathrm{tr}(S^TLS)+\xi\|S\|_F^2+\mu\|A\|_F^2 \\
&s.t.\ \ \|S\|_\infty\le a,\ \ \|A\|_\infty\le b.
\end{aligned}
\tag{15}
$$

Because $\|S_0\|_0$ is a constant number, we disregard it and our formulation becomes to (16) that is exactly the same to (5).

$$\begin{aligned}
&\min_{S,A} \frac{1}{2}\|E-SA\|_F^2 + (\eta-\lambda)\|S\odot S_0\|_0 + (\eta+\lambda)\left\|S\odot \bar{S}_0\right\|_0 + \kappa \mathrm{tr}(S^T L S) + \xi\|S\|_F^2 + \mu\|A\|_F^2 \\
&s.t.\ \|S\|_\infty \le a,\ \|A\|_\infty \le b.
\end{aligned} \tag{16}$$

A.3 Derivations from (7a) and (7b) to (8a) and (8b)

The derivation is based on completing the square technique. We simply show the derivation from (7a) to (8a) and the derivation from (7b) to (8b) is the same.

$$\begin{aligned}
S^{k+1} &\in \arg\min_{S\in\Upsilon}\left\{\left\langle S-S^k,\ \nabla_S F(S^k,A^k)\right\rangle + \frac{c^k}{2}\left\|S-S^k\right\|_F^2 + \Phi(S)\right\} \\
&\in \arg\min_{S\in\Upsilon}\left\{\left\langle S-S^k,\ \nabla_S F(S^k,A^k)\right\rangle + \frac{c^k}{2}\left\langle S-S^k, S-S^k\right\rangle + \Phi(S)\right\} \\
&\in \arg\min_{S\in\Upsilon}\left\{\mathrm{tr}(S^T\nabla_S F(S^k,A^k)) + \frac{c^k}{2}\mathrm{tr}(S^TS - 2S^TS^k + (S^k)^TS^k) + \Phi(S)\right\} \\
&\in \arg\min_{S\in\Upsilon}\left\{\Phi(S) + \frac{c^k}{2}\left\|S-(S^k-\frac{1}{c^k}\nabla_S F(S^k,A^k))\right\|_F^2\right\} \\
&\in \arg\min_{\|S\|_\infty\le a}\left\{\Phi(S) + \frac{c^k}{2}\left\|S-(S^k-\frac{1}{c^k}\nabla_S F(S^k,A^k))\right\|_F^2\right\}.
\end{aligned} \tag{17}$$

The second last line is obtained based on completing the square technique and disregarding a term that is consist of the constants S^k and $\nabla_S F(S^k, A^k)$. The last line is obtained by putting in the specific constraint for S.

A.4 Derivations of the Proximal Mappings used in the NetREX Algorithm

The derivation for (11a) using Corollary 1 is shown below.

$$\begin{aligned}
S^{k+1} &\in \arg\min_{\|S\|_\infty\le a}\left\{\Phi(S) + \frac{c^k}{2}\left\|S-U^k\right\|_F^2\right\} \\
&= \arg\min_{\|S\|_\infty\le a}\left\{\frac{2(\eta+\lambda)}{c^k}\left\|\bar{\mathbb{S}}_0\odot S\right\|_0 + \frac{2(\eta-\lambda)}{c^k}\|\mathbb{S}_0\odot S\|_0 + \frac{2\xi}{c^k}\|S\|_F^2 + \left\|S-U^k\right\|_F^2\right\} \\
&= \mathrm{prox}_{\|\cdot\|_\infty\le a}\left(U^k, \frac{2(\eta+\lambda)}{c^k}, \frac{2(\eta-\lambda)}{c^k}, \frac{2\xi}{c^k}\right).
\end{aligned} \tag{18}$$

The derivation for (11b) is shown below.

$$\begin{aligned}
A^{k+1} &= \arg\min_{\|A\|_\infty\le b}\left\{\Psi(A) + \frac{d^k}{2}\left\|A-V^k\right\|_F^2\right\} \\
&= \arg\min_{\|A\|_\infty\le b}\left\{\frac{2\mu}{d^k}\|A\|_F^2 + \left\|A-V^k\right\|_F^2\right\} \\
&= \mathbb{P}_{\|\cdot\|_\infty\le b}\left(\frac{1}{1+\frac{2\mu}{d^k}}V^k\right).
\end{aligned} \tag{19}$$

B Propositions, Corollary, and Their Proofs

B.1 The Proposition for the Proximal Mapping of ℓ_0 Elastic Net and Its Proof

Proposition 1 (**Proximal Mapping of ℓ_0 Elastic Net Under $\|\|_\infty$ Constrain Proposition**). *For given $Y \in \mathbb{R}^{m\times n}$, the proximal mapping of ℓ_0 elastic net under $\|\|_\infty$ norm Constrain is*

$$\arg\min_{\|X\|_\infty \le C} \left\{ \|Y - X\|_F^2 + b\,\|X\|_F^2 + c^2\,\|X\|_0 \right\} = T_{\frac{c}{\sqrt{b+1}}}\left(\mathbb{P}_{\|\cdot\|_\infty \le C}\left(\frac{Y}{b+1} \right) \right), \tag{20}$$

where the projection operator $\mathbb{P}_{\|\cdot\|_\infty \le C}(\cdot)$ is defined in (21)

$$\mathbb{P}_{\|\cdot\|_\infty \le C}(Y) := \arg\min \left\{ \|Y - X\|_F^2 : \|X\|_\infty \le C \right\} = \operatorname{sign}(Y) \odot \max\{|Y|, C\}, \tag{21}$$

where the $||$, $\operatorname{sign}()$ and $\max\{\}$ operations are taken component-wise. And the hard-thresholding operator $T_c(\cdot)$ is defined as

$$T_c(Y) := \arg\min_X \left\{ \|Y - X\|_F^2 + c^2\,\|X\|_0 \right\}, \tag{22}$$

where $Y \in \mathbb{R}^{m\times n}$ is any given matrix and $T_c : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is a component-wise mapping that can be explicitly wrote out

$$(T_c(Y))\,(i,j) = \begin{cases} Y(i,j), \text{ if } & |Y(i,j)| > c; \\ \{0, c\}, \text{ if } & |Y(i,j)| = c; \\ 0, & o.w.. \end{cases} \tag{23}$$

Proof.

$$\begin{aligned} &\arg\min_{\|X\|_\infty \le C} \left\{ \|Y - X\|_F^2 + b\,\|X\|_F^2 + c^2\,\|X\|_0 \right\} \\ =&\arg\min_{\|X\|_\infty \le C} \left\{ (b+1) \left\| \frac{Y}{b+1} - X \right\|_F^2 + c^2\,\|X\|_0 \right\} \\ =&\arg\min_{\|X\|_\infty \le C} \left\{ \left\| \frac{Y}{b+1} - X \right\|_F^2 + (\frac{c}{\sqrt{b+1}})^2\,\|X\|_0 \right\} \\ =&T_{\frac{c}{\sqrt{b+1}}}\left(\mathbb{P}_{\|\cdot\|_\infty \le C}\left(\frac{Y}{b+1} \right) \right). \end{aligned} \tag{24}$$

The last equality is derived based on Lemma 1. □

Lemma 1. *Let $U \in \mathbb{R}^{m\times n}$, then*

$$\arg\min \left\{ \|U - X\|_F^2 + c^2\,\|X\|_0 : \|X\|_\infty \le C \right\} = T_c\left(\mathbb{P}_{\|\cdot\|_\infty \le C}(U) \right). \tag{25}$$

Proof. For given $U \in \mathbb{R}^{m\times n}$, let us introduce the following notations

$$\|X\|_+^2 = \sum_{(i,j)\in\mathcal{I}^+} X(i,j)^2 \text{ and } \|X\|_-^2 = \sum_{(i,j)\in\mathcal{I}^-} X(i,j)^2, \tag{26}$$

where

$$\mathcal{I}^+ = \{(i,j) \in \{1,...,m\} \times \{1,...,n\} : |U(i,j)| \le C\} \tag{27}$$

and

$$\mathcal{I}^- = \{(i,j) \in \{1,...,m\} \times \{1,...,n\} : |U(i,j)| > C\} \tag{28}$$

The following observations hold

$$\begin{aligned}
&(i)\ \|X\|_F^2 = \|X\|_+^2 + \|X\|_-^2 \\
&(ii)\ \|X-U\|_+^2 + \|X-C\|_-^2 = \left\|X - \mathbb{P}_{\|\cdot\|_\infty \le C}(U)\right\|_F^2 \\
&(iii)\ \|X-C\|_-^2 = 0 \Leftrightarrow X(i,j) = C\ \forall (i,j) \in \mathcal{I}^-
\end{aligned} \tag{29}$$

where the second observation follows from observation (i) and the fact that $\left(\mathbb{P}_{\|\cdot\|_\infty \le C}(U)\right)(i,j) = U(i,j)$ for $(i,j) \in \mathcal{I}^+$ and $\left(\mathbb{P}_{\|\cdot\|_\infty \le C}(U)\right)(i,j) = C$ for $(i,j) \in \mathcal{I}^-$.

Based on the above facts, we have that $\bar{X} \in \text{prox}_{\|\cdot\|_0}^{\|\cdot\|_\infty \le C}(U,c)$ if and only if

$$\bar{X} \in \arg\min \left\{\|U-X\|_F^2 + c^2 \|X\|_0 : \|X\|_\infty \le C\right\} \tag{30a}$$

$$= \arg\min \left\{\|U-X\|_+^2 + \|U-X\|_-^2 + c^2 \|X\|_0 : \|X\|_\infty \le C\right\} \tag{30b}$$

$$= \arg\min \left\{\|U-X\|_+^2 + c^2 \|X\|_0 : X(i,j) = C\ \forall (i,j) \in \mathcal{I}^-,\ \|X\|_\infty \le C\right\}, \tag{30c}$$

where the last equality follows that fact that the solution of (30c) is also the solutions of (30b), while the converse follows by a simple contradiction argument. Furthermore, one can find that the constraint $\|X\|_\infty \le C$ can be removed without affecting the optimal solution of the problem. Therefore, applying observation (ii) and (iii), we obtain

$$\begin{aligned}
\bar{X} &\in \arg\min \left\{\|U-X\|_+^2 + c^2 \|X\|_0 : \|X-C\|_-^2 = 0\right\} \\
&= \arg\min \left\{\|U-X\|_+^2 + \|X-C\|_-^2 + c^2 \|X\|_0\right\} \\
&= \arg\min \left\{\left\|X - \mathbb{P}_{\|\cdot\|_\infty \le C}(U)\right\|_F^2 + c^2 \|X\|_0\right\} = T_c\left(\mathbb{P}_{\|\cdot\|_\infty \le C}(U)\right),
\end{aligned} \tag{31}$$

where the last equality is the definition of T_c in Eq. (24). □

B.2 The Corollary of the Proximal Mapping of ℓ_0 Elastic Net Proposition and Its Proof

Corollary 1. *For given $U \in \mathbb{R}^{m\times n}$, the proximal mapping of $\Phi(S) = \alpha \left\|\bar{S}_0 \odot S\right\|_0 + \beta \|S_0 \odot S\|_0 + \gamma \|S\|_F^2$ on $\|S\|_\infty \le C$ is*

$$\begin{aligned}
\text{prox}_{\|\cdot\|_\infty \le C}(U,\alpha,\beta,\gamma) &\in \arg\min_{\|S\|_\infty \le C} \left\{\alpha \left\|\bar{S}_0 \odot S\right\|_0 + \beta \|S_0 \odot S\|_0 + \gamma \|S\|_F^2 + \|U-S\|_F^2\right\} \\
&= T_{\sqrt{\frac{\beta}{\gamma+1}}}\left(\mathbb{P}_{\|\cdot\|_\infty \le C}\left(\frac{\mathbb{U}}{\gamma+1}\right)\right) + T_{\sqrt{\frac{\alpha}{\gamma+1}}}\left(\mathbb{P}_{\|\cdot\|_\infty \le C}\left(\frac{\bar{\mathbb{U}}}{\gamma+1}\right)\right),
\end{aligned} \tag{32}$$

where $\mathbb{U} = S_0 \odot U$ and $\bar{\mathbb{U}} = \bar{S}_0 \odot U$.

Proof. We know that the U can be decomposed into $U = \mathbf{1} \odot U = (S + \bar{S}_0) \odot U = S \odot U + \bar{S}_0 \odot U = \mathbb{U} + \bar{\mathbb{U}}$. Similarly, $S = S \odot S + \bar{S}_0 \odot S = \mathbb{S} + \bar{\mathbb{S}}$. Applying those decomposition into Eq. (32), we can decompose the proximal mapping into two parts.

$$\begin{aligned}
&\arg \min_{\|S\|_\infty \le C} \left\{ \alpha \left\| \bar{S}_0 \odot S \right\|_0 + \beta \left\| S_0 \odot S \right\|_0 + \gamma \left\| S \right\|_F^2 + \left\| U - S \right\|_F^2 \right\} \\
&= \arg \min_{\|\mathbb{S}\|_\infty \le C} \left\{ \beta \left\| \mathbb{S} \right\|_0 + \gamma \left\| \mathbb{S} \right\|_F^2 + \left\| \mathbb{U} - \mathbb{S} \right\|_F^2 \right\} + \arg \min_{\|\bar{\mathbb{S}}\|_\infty \le C} \left\{ \alpha \left\| \bar{\mathbb{S}} \right\|_0 + \gamma \left\| \bar{\mathbb{S}} \right\|_F^2 + \left\| \bar{\mathbb{U}} - \bar{\mathbb{S}} \right\|_F^2 \right\}.
\end{aligned} \tag{33}$$

Based on Proximal Mapping of ℓ_0 Elastic Net Proposition 1, we know

$$\arg \min_{\|\mathbb{S}\|_\infty \le a} \left\{ \beta \left\| \mathbb{S} \right\|_0 + \gamma \left\| \mathbb{S} \right\|_F^2 + \left\| \mathbb{U} - \mathbb{S} \right\|_F^2 \right\} = T_{\sqrt{\frac{\beta}{\gamma+1}}} \left(\mathbb{P}_{\|\cdot\|_\infty \le C} \left(\frac{\mathbb{U}}{\gamma + 1} \right) \right). \tag{34}$$

Similarly,

$$\arg \min_{\|\bar{\mathbb{S}}\|_\infty \le a} \left\{ \alpha \left\| \bar{\mathbb{S}} \right\|_0 + \gamma \left\| \bar{\mathbb{S}} \right\|_F^2 + \left\| \bar{\mathbb{U}} - \bar{\mathbb{S}} \right\|_F^2 \right\} = T_{\sqrt{\frac{\alpha}{\gamma+1}}} \left(\mathbb{P}_{\|\cdot\|_\infty \le C} \left(\frac{\bar{\mathbb{U}}}{\gamma + 1} \right) \right). \tag{35}$$

Combining Eq. (34) and Eq. (35) proves the proposition.

$$\begin{aligned}
&\arg \min_{\|S\|_\infty \le C} \left\{ \alpha \left\| \bar{\mathbb{S}}_0 \odot S \right\|_0 + \beta \left\| \mathbb{S}_0 \odot S \right\|_0 + \gamma \left\| S \right\|_F^2 + \left\| U - S \right\|_F^2 \right\} \\
&= T_{\sqrt{\frac{\beta}{\gamma+1}}} \left(\mathbb{P}_{\|\cdot\|_\infty \le C} \left(\frac{\mathbb{U}}{\gamma + 1} \right) \right) + T_{\sqrt{\frac{\alpha}{\gamma+1}}} \left(\mathbb{P}_{\|\cdot\|_\infty \le C} \left(\frac{\bar{\mathbb{U}}}{\gamma + 1} \right) \right).
\end{aligned} \tag{36}$$

□

B.3 The convergence of NetREX algorithm

Proposition 2 (**Convergnece Proposition**). *Let $\left\{ (S^k, A^k) \right\}_{k \in \mathbb{N}}$ be a sequence generated by NetREX algorithm. Then,*
(i) *The sequence $\left\{ (S^k, A^k) \right\}_{k \in \mathbb{N}}$ has finite length, that is*

$$\sum_{k=1}^{\infty} \left\| S^{k+1} - S^k \right\|_F + \left\| A^{k+1} - A^k \right\|_F < \infty. \tag{37}$$

(ii) *The sequence $\left\{ (S^k, A^k) \right\}_{k \in \mathbb{N}}$ converges to a critical point (S^*, A^*) of the NetREX problem.*

Proof. We apply Theorem 3.1 in [11] to guarantee that the sequence generated by NetREX is globally convergent to critical points of (5). □

Supplementary Materials II:
Implementation Details

C The NetREX Algorithm

C.1 The Details of the NetREX Algorithm

Algorithm 1: The NetREX algorithm.

Input : S_0, E, L, η, λ, κ, ξ, $v > 0$ and K;
Output: S and A.

begin
  (S^0, A^0) =Initialization(S_0). // Algorithm 2.
  for $k = 0, 1, 2, ..., K$ **do**
    $c^k = \max\{v,\ L(A^k)\}$.
    $U^k = S^k - \frac{1}{c^k}\left(S^k A^k (A^k)^T + 2\kappa L S^k - E(A^k)^T\right)$. // put (10) into (9).
    $S^{k+1} \in \text{prox}_{\|\cdot\|_\infty \le a}\left(U^k, \frac{2\eta}{c^k}, \frac{2(\eta-\lambda)}{c^k}, \frac{2\xi}{c^k}\right)$. //as shown in (11a).
    $d^k = \{v,\ L(S^{k+1})\}$.
    $V^k = A^k - \frac{1}{d^k}\left((S^{k+1})^T (S^{k+1}) A^k - (S^{k+1})^T E\right)$. // put (10) into (9).
    $A^{k+1} = \mathbb{P}_{\|\cdot\|_\infty \le b}\left(\frac{1}{1+\frac{2\mu}{d^k}} V^k\right)$. //as shown in (11b).
  end
  $S = S^K$ and $A = A^K$
end

C.2 The Initialization Algorithm for NetREX

To ensure that the starting point is consistent with the prior network, (S^0, A^0) have to be inferred from our prior network $\mathcal{G}_0$. To do this, we compute (S^0, A^0) by solving the following problem, which is obtained from the original NetREX formulation by dropping the constraints and disregarding the non-smooth regularization term $(\eta - \lambda)\|S \odot S_0\|_0 + (\eta + \lambda)\left\|S \odot \bar{S}_0\right\|$ of S.

$$\min_{S,A} : J(S, A) = \frac{1}{2}\|E - SA\|_F^2 + \kappa \text{tr}(S^T L S) + \xi \|S\|_F^2 + \mu \|A\|_F^2 \tag{38}$$

The problem (38) can be solved by the standard Gauss-Seidel scheme [39] that alternatively solves the multi-variable optimization problem with respect to one variable while fixing the rest of the variables. Specifically, we can fix $S = S_k^0$ and solve (38) with respect to A in closed form shown in Line 4 of Algorithm 2. Then, we fix $A = A_k^0$ and solve (38) with respect to S, whose solution is the solution of the Sylvester equation $SA_k^0(A_k^0)^T + 2(\kappa L + \xi I)S = E(A_k^0)^T$ (derived by setting $\nabla H(S, A_k^0) = 0$). The Sylvester equation is solved by standard Bartels-Stewart algorithm. We alternatively ran lines 4 and 5 K times. In the end, we project the solutions A_K^0 and S_K^0 into feasible space of Eq. (5) by the projection operator (21) shown in lines 7 and 8. Algorithm 2 elaborates the details of obtaining (S^0, A^0).

Algorithm 2: The initialization for NetREX

Function: Initialization(S_0);
Input : S_0;
Output : S^0 and A^0.

```
begin
    S_0^0 = S_0.
    for k = 0, 1, 2, ..., K do
        A_k^0 = ((S_k^0)^T S_k^0 + μI)^{-1} (S_k^0)^T E.
        S_{k+1}^0 := { Ŝ | Ŝ A_k^0 (A_k^0)^T + 2(κL + ξI)Ŝ = E(A_k^0)^T }.
    end
    A^0 = P_{||·||_∞ ≤ b}(A_K^0).
    S^0 = P_{||·||_∞ ≤ a}(S_K^0).
end
```

D Evaluation Metrics

D.1 F-measure

F-measure is defined as

$$\text{F-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{39}$$

where

$$\text{Precision} = \frac{|\mathcal{E}^p \cap \mathcal{E}|}{|\mathcal{E}^p|}, \ \text{Recall} = \frac{|\mathcal{E}^p \cap \mathcal{E}|}{|\mathcal{E}|}. \tag{40}$$

$\mathcal{E}$ and $\mathcal{E}^p$ are edge sets of the underling regulatory network $\mathcal{G}$ and the predicted regulatory network, respectively. F-measure ranges from 0 to 1, where 1 presents that the underlining $\mathcal{G}$ is fully recovered and 0 means the opposite.

D.2 Fold Enrichment for GO annotations

We consider two genes are co-regulated if the Jaccard similarity coefficient between the TF set regulating the first gene and the TF set regulating the second gene is larger than 0.5. The Jaccard similarity coefficient between two sets is the ratio of the size of the intersection of the given two sets to the size of the union of these two sets. Then for each co-regulated gene pair, we again use the Jaccard similarity coefficient to measure the similarity between the GO annotation set corresponding to the first gene and the GO annotation set corresponding to the second gene. In the end, we compute the average of this coefficient overall co-regulated gene pairs. The same procedure was done for 100 randomized networks, and the enrichment is the ratio of the average coefficient of the original network to the average of the randomized networks. The randomized networks are generated by permuting the node labels of the original network. Hence, all randomized networks have the same topology to the original network but with different node labels. The statistical significance is accessed at a level of 0.05 using a one-side unpaired T-test for comparing the Jaccard coefficients from the original network with coefficients from 100 randomized networks.

D.3 Fold Enrichment for PPIs

Enrichment of co-regulated genes for PPIs was computed analogously to enrichment for GO annotations. Specifically, we computed the ratio of the number of PPIs for co-regulated gene pairs to the average number of such PPIs in 100 randomized networks, using the same definition for co-regulation and network randomization.

E Parameters

E.1 The NetREX_NP and NetREX_ℓ_1 Algorithms

The NetREX_NP algorithm is same to Algorithm 1 with $\lambda = 0$. The formulation of NetREX_NP is

$$\begin{aligned} &\min_{S,A} \frac{1}{2} \|E - SA\|_F^2 + \eta \|S\|_0 + \kappa \mathrm{tr}(S^T L S) + \xi \|S\|_F^2 + \mu \|A\|_F^2 \\ &s.t.\ \|S\|_\infty \le a,\ \|A\|_\infty \le b. \end{aligned} \tag{41}$$

The formulation is similar to sparse coding [19] if we remove the graph embedding term.
The NetREX_ℓ_1 formulation is as following.

$$\begin{aligned} &\min_{S,A} \frac{1}{2} \|E - SA\|_F^2 + (\eta - \lambda) \|S \odot S_0\|_1 + (\eta + \lambda) \left\|S \odot \bar{S}_0\right\|_1 + \kappa \mathrm{tr}(S^T L S) + \xi \|S\|_F^2 + \mu \|A\|_F^2 \\ &s.t.\ \|S\|_\infty \le a,\ \|A\|_\infty \le b. \end{aligned} \tag{42}$$

To do a fair comparison, we also use the PALM algorithm to solve it which is analogous to Algorithm 1. The only difference is that in line 6 of Algorithm 1, we use proximal mapping of ℓ_1 elastic net that is given in [40] instead of proximal mapping for ℓ_0 elastic net.

E.2 Parameter Settings for Simulated Data

The parameters used to generate simulated data is $L = 60, N = 500, M = 100$. The density of the ground truth GRN is 0.1. The noise level in simulated expression data E is controlled by $\sigma = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. The percentage of true edges in $\mathcal{G}_0$ is controlled by $\theta = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

There are seven parameters for NetREX algorithm, which are λ, η, κ, ξ, μ, a and b. We applied grid search to find the optimal parameters. The settings are as following. We set $\eta - \lambda \in [0.2, 5]$ with interval 0.2, $\eta + \lambda \in [1, 50]$ with interval 1, $\kappa \in [0.1, 0.5]$ with interval 0.1, $\xi = \{0.1, 1\}$, $\mu = \{0.1, 1\}$ and $a = b = \max_{i,j}(\text{abs } E(i,j))$. We used the same parameter setting for NetREX_NP except $\lambda = 0$ and $\eta \in [1, 50]$ with interval 1. For NetREX_ℓ_1 algorithm, we use exactly the same parameters to the NetREX algorithm.

To test the potential of the competing algorithms, for certain noise level, we first applied grid search for all algorithms to find their optimal parameters on only one simulated data set based on F-measure. Then we use the optimal parameters to other 50 simulated data set under the same noise level. We compared the performance of different algorithms based on the F-measures.

E.3 Parameter Settings for DrosDel Data

We set $\eta - \lambda \in [0.01, 0.2]$ with interval 0.01, $\eta + \lambda \in [0.5, 10]$ with interval 0.5, $\kappa = \{0.05, 0.1\}$, $\xi = \{0.1, 1, 5\}$, $\mu = \{0.1, 1\}$ and $a = b = \max_{i,j}(\text{abs } E(i,j))$. Because we do not know the ground truth regulatory networks, we proposed a heuristic score to rank our predicted networks. The score can be computed as

$$R = \|E - SA\|_F^4 \times \|S\|_0 . \tag{43}$$

We reasoned that the promising networks should be able to describe the underling regulatory system (making $\|E - SA\|_F$ small) as well as have only the contributing regulations (the number of edges in the network $\|S\|_0$ is small). We used power of 4 on fitting error ($\|E - SA\|_F$) because, to build a condition specific RGN, fitting the condition specific expression data E is more important. The

smaller R implies that we can fit the expression data using the network with smaller number of edges. We ranked all predicted networks under different parameters in terms of R in ascending order. We showed the performance of the top 20 networks for the female flies.

For constructing the TF-gene correlation networks, we used Pearson coefficient cutoffs ($\{0.6, 0.7, 0.8, 0.9\}$) and show the performance of the networks under different cutoffs.

For GENIE3, there is only a parameter K used by it. [16] suggests two settings for K, which are $K = M - 1$ and $K = \sqrt{M}$. We compared the results of these two Ks. $K = M - 1$ is better than $K = \sqrt{M}$. Therefore, we use $K = M - 1$ in comparison. We also need a cutoff to get the final GRN. We ranked the weighted predicted by GENIE3 and used the top 100,000, 200,000, 300,000, 400,000 and 500,000 as output, respectively.

The co-expressed gene pairs used in the comparison shown in Fig. 3 is the same to the one we inputted in our formulation as the graph embedding term.