Optimal sparsity criteria for network inference

Andreas Tjärnberg[*,1,2]; Torbjörn E. M. Nordling[*,3]; Matthew Studham[1,2]; Erik L.L. Sonnhammer[1,2,4]

***Abstract*—Gene regulatory network inference, *i.e.* determination of the regulatory interactions between a set of genes, provides mechanistic insights of central importance to research in systems biology. Most contemporary network inference methods rely on a sparsity/regularization coefficient, which we call ζ (zeta), to determine the degree of sparsity of the network estimates, *i.e.* the total number of links between the nodes. However, they offer little or no advice on how to select this sparsity coefficient, in particular for biological data with few samples. We show that an empty network is more accurate than estimates obtained for a poor choice of ζ. In order to avoid such poor choices, we propose a method for optimisation of ζ which maximizes the accuracy of the inferred network for any sparsity-dependent inference method and data set. Our procedure is based on leave one out cross optimisation and selection of the ζ value that minimizes the prediction error. We also illustrate the adverse effect of noise, few samples, and uninformative experiments on network inference and our method for optimisation of ζ. We demonstrate that our ζ optimisation method for two widely used inference algorithms–Glmnet and NIR–gives accurate and informative estimates of the network structure, given that the data is informative enough.**

I. Introduction

Gene regulatory networks (GRNs) model the mechanistic interactions between genes, giving insight into how signals between genes are propagated through the genetic network and how genes could respond to exogenous and endogenous stimuli. Much work has been done on developing and evaluating algorithms for inference of regulatory networks from perturbations and responses in expression data [1], [15], [13], [21]. The primary goal is to find the structure of the network, *i.e.* the interactions that exist within the set of nodes. Accurate prediction of the observed responses is in general not sufficient for accurate estimation of the network structure. Most contemporary inference methods rely on a sparsity/regularization coefficient, which we call ζ, to control the degree of sparsity of the network estimates, *i.e.* the number of links between the nodes. They however offer little or no advice on how to select this sparsity coefficient, even though it typically is crucial for accurate estimation of the structure. In particular, the low number of samples encountered in biological data sets is problematic and makes it unclear if classical model selection and cross validation strategies work. We therefore investigated if the optimal ζ value can be found based on the prediction error for data with a low number of samples and as a result propose a method for optimisation of ζ which maximizes the accuracy of the inferred network for a given inference method and data set. We apply our method to three *in silico* data sets with varying information content and demonstrate that our method works well when the data is informative enough.

By selecting the sparsity coefficient, one in general determines the trade off between accuracy of data prediction and model complexity. Classical model selection criteria, such as AIC, BIC, and cross validation, have long been studied and shown to under certain conditions perform a good trade off, in particular, when the number of samples is large and the model is used for data prediction [26]. These classical methods have in a few studies been used for selection of the regularization coefficient during the past decade, including applications to biological data [9], [29], [19], [11], [27], [23], [5]. The BIC is in [29], [23], [19] shown to be favorable over the other criteria tested in their study, while [27] found selection based on minimization of the prediction error in cross validation and bootstraping to be favorable over AIC and BIC. Cross validation is in [11], [2] used for selection of the regularization coefficient in LASSO, but they instead chose the value corresponding to the sparsest model within one standard error from the minimum to decrease the number of false positives. Selection of the sparsity coefficient in network inference is only studied in [23]. *In vivo* data is in [2] used to infer the GRN of Halobacterium and they select ζ, but the ”true” network is unknown and the performance can therefore not be properly evaluated. *In silico* data from the DREAM4 challenge is in [23] used for method evaluation, but they do not consider the signed topology that we are interested in. Networks with 100 genes were used to generate their data, simulating 100 multifactorial steady-state perturbation experiments. We consider a similar network inference case but use a small network and systematically vary data properties to gain insight on how to select ζ for three different inference algorithms. We focus on the relation between the prediction error and topological accuracy in network inference from data with few samples and do not therefore consider rules that make a specific trade off between predictive ability and model complexity, such as AIC and BIC.

The use of simulated data from gold standard networks for evaluation of network inference algorithms is today standard. Several *in silico* network and data generating software programs have therefore been created [14], [25], [22]. These tools use varying modeling approaches and biological mechanisms to simulate biological networks and network motifs. Data sets based on simpler directly generated linear ordinary differential equation (ODE) models have also been used in benchmarking [1]. For simplicity and control of network and data properties, we here use a linear ODE model for data generation.

1 Stockholm Bioinformatics Centre, Science for Life Laboratory, Box 1031, SE-17121 Solna, Sweden

2 Department of Biochemistry and Biophysics, Stockholm University

3 Automatic Control lab, School of Electrical Engineering, KTH Royal Institute of Technology, SE-10044 Stockholm, Sweden

4 Swedish eScience Research Center

* These authors contributed equally.

II. Problem description

Throughout this paper we consider a system that can be approximated as a system of first order linear ODEs:

$$\begin{aligned} \dot{x}_i(t) &= \textstyle\sum_{j=1}^{N} a_{ij}x_j(t) + p_i(t) - f_i(t) \\ y_i(t) &= x_i(t) + e_i(t) \end{aligned} \tag{1}$$

where the state vector $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_N(t)]^T$ represent actual mRNA expression changes relative to the initial state of the system. The vector $\boldsymbol{p}(t) = [p_1(t), p_2(t), \ldots, p_N(t)]^T$ is the desired or measured perturbation that differs from the applied one by the noise $\boldsymbol{f}(t)$ and $\boldsymbol{y}(t) = [y_1(t), y_2(t), \ldots, y_N(t)]^T$ is the measured response that differs from the expression changes by the noise $\boldsymbol{e}(t)$. The parameters a_{ij} of the interaction matrix describes the influence of gene j on gene i. A positive value represents an activation, while a negative value represents an inhibition. The relative strength of the interaction is given by the value of the parameter. We make the common assumption that only steady state data is recorded and write the system in matrix notation as

$$\boldsymbol{Y} = -\boldsymbol{A}^{-1}\boldsymbol{P} + \boldsymbol{A}^{-1}\boldsymbol{F} + \boldsymbol{E}. \tag{2}$$

Here $\boldsymbol{Y}$ is our steady state expression matrix after applying the perturbations $\boldsymbol{P}$ and $\boldsymbol{A}$ is the interaction matrix. This type of linear data model has previously been used when inferring gene regulatory networks in *e.g.* [12], [20], [18], [24], [17]. In this case the goal of network inference corresponds to inference of the structure of the interaction matrix $\boldsymbol{A}$.

To obtain sparse network models a variety of methods that either directly or indirectly places a weight or constraint on the model complexity in terms of the number of non-zero parameters a_{ij} have been developed in several fields [16], [10], [7], [3], [26]. The idea is to exclude less influential interactions while keeping the most influential ones in terms of ability to predict experimental observations by penalizing small non-zero parameters. To accomplish this, a method specific sparsity/regularization coefficient, which we here denote $\tilde{\zeta}$ and later standardize such that $\zeta \in [0, 1]$, is used in most methods to set the weight or constraint on the number of links, like in the LASSO method [28]

$$\hat{\boldsymbol{A}}_{\text{reg}}(\tilde{\zeta}) = \arg\min_{\boldsymbol{A}} ||\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{P}||_{l_2} + \tilde{\zeta}||\boldsymbol{A}||_{l_1}, \tag{3}$$

which is one of the most well known ones. Several conditions for near ideal model selection have been established for certain methods, in particular given a suitable choice of ζ [4], [10], [30], but how to in general chose ζ based on data sets with a low number of samples is still an open problem. We therefore here study how the selection of ζ affects the network estimate for data sets of the type common in Systems biology. The idea is to through numerical simulations and analysis investigate if an efficient method for selection of ζ that can be wrapped around implementations of existing inference methods could be developed. In order to be efficient the method must select a ζ value such that no network estimate that is structurally more similar to the "true" network exists for any other choice of ζ for the particular inference method. As a result we propose such a method.

III. Methods

A. Pre-treatment of data

A model is only useful for prediction of experiments similar to the ones it was estimated from, so the validation set should only contain experiments that are linearly dependent on the experiments in the training set.

We therefore measure the degree of linear dependence between an experiment $\boldsymbol{y}_k$ and the ones in the training set $\boldsymbol{Y}_{t\neq k}$ as

$$\eta_{\boldsymbol{y}_k} \triangleq ||\boldsymbol{Y}_{t\neq k}^T \boldsymbol{y}_k||_1 \tag{4}$$

and similarly for the perturbations. All directions of the system are sufficiently excited in the examples that we provide later on and we therefore use the smallest singular value σ_N as our limit on η. The index set of the validation experiments is hence

$$\mathcal{V} \triangleq \big\{k | \eta_{\boldsymbol{y}_k} \geq \sigma_N(\boldsymbol{Y}) \text{ and } \eta_{\boldsymbol{p}_k} \geq \sigma_N(\boldsymbol{P})\big\}. \tag{5}$$

The sparsity of the estimated network for a specific ζ value depends both on the data and the algorithm used. To simplify the comparison between different data sets and algorithms, we therefore standardize ζ, such that an empty network is obtained for $\zeta = 1$ and a full network for $\zeta = 0$. For Glmnet and LSCO the standardized ζ is equal to the actual $\tilde{\zeta}$ divided by the smallest $\tilde{\zeta}$, such that the network estimate $\hat{\boldsymbol{A}}_{\text{reg}} = \boldsymbol{0}$ for all different training data sets $\{\boldsymbol{Y}_{t\neq k}, \boldsymbol{P}_{t\neq k}\}$. For NIR the standardized ζ is $(N - \tilde{\zeta})/N$.

B. Leave one out cross optimisation

When few experiments exist, use of the whole data set to predict the network might be an attractive option but it would in general lead to over fitting of the parameters to noise. As the number of observations tend to be small, we employ a leave one out cross optimisation (LOOCO) scheme, where each experiment k that fulfills (5) in turn is left out and the remaining $t \neq k$ experiments are used to estimate the model $\hat{\boldsymbol{A}}$. This model is then after re-estimation of the parameters used to predict the response in kth experiment.

Introduction of a term penalizing model complexity unfortunately also affects the estimate of most non-zero parameters, biasing them away from the value minimizing the prediction error. Though a number of approaches to reduce the bias have been constructed, see *e.g.* [9], [8], [19], none of them removes it completely in all cases and many inference methods do not utilize these approaches. We therefore for all models re-estimate the non-zero parameters using a constrained least squares (CLS) method solving the following optimization problem

$$\hat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} \sum \operatorname{diag}(\boldsymbol{\Delta}^T \boldsymbol{R} \boldsymbol{\Delta}) \tag{6a}$$

$$\text{s.t. } \boldsymbol{\Delta} = \boldsymbol{A}\boldsymbol{Y} + \boldsymbol{P}, \tag{6b}$$

$$\boldsymbol{R} = \left(\hat{\boldsymbol{A}}_{\text{init}} \operatorname{Cov}[\boldsymbol{y}] \hat{\boldsymbol{A}}_{\text{init}}^T + \operatorname{Cov}[\boldsymbol{p}]\right)^{-1}, \tag{6c}$$

$$\operatorname{sign} \boldsymbol{A} = \operatorname{sign} \hat{\boldsymbol{A}}_{\text{reg}}. \tag{6d}$$

Here $\hat{\boldsymbol{A}}_{\text{init}}$ is equal to the estimate given by the network inference method $\hat{\boldsymbol{A}}_{\text{reg}}$ if the network inference method gives

an estimate that can be used for prediction, otherwise the ordinary least squares estimate $\hat{\boldsymbol{A}}_{\text{ols}} = -\boldsymbol{P}\boldsymbol{Y}^{\dagger}$ is used. Here † denotes the Moore-Penrose generalized inverse, $\mathrm{Cov}[\boldsymbol{y}]$ the covariance matrix of the response in an experiment or an estimate of it, $\mathrm{Cov}[\boldsymbol{p}]$ the covariance matrix of the perturbation in an experiment or an estimate of it, and sign the signum function. This is a simplification of the method presented in [18], where the covariance is assumed to be identical in all experiments and their l_1 constraint is replaced by constraining the structure to be identical to that of the estimate given by the network inference method for the investigated $\tilde{\zeta}$. It is based on the relation between weighted least squares and maximum likelihood estimators clear that the solution of this optimization problem is close to the maximum likelihood estimate under the structural constraints, given that the errors are normally distributed with the assumed covariance matrices and $\hat{\boldsymbol{A}}_{\text{init}}$ is sufficiently close to the final estimate. Consequently, the estimate is close to the the best linear unbiased estimate under the structural constraints, which is equal to the minimum variance unbiased estimate for normally distributed errors. This re-estimation therefore enables us to obtain unbiased estimates of the non-zero parameters that minimizes the prediction error. We solve this convex optimisation problem using CVX (cvxr.com/cvx) in Matlab (www.mathworks.com).

C. Selection of optimal ζ

The prediction error of all network estimates $\hat{\boldsymbol{A}}$ obtained by the LOOCO for a specific ζ is evaluated by the mean residual sum of squares (RSS)

$$\mathrm{RSS}(\zeta) \triangleq \frac{1}{\#\mathcal{V}} \sum_{k \in \mathcal{V}} ||\hat{\boldsymbol{y}}_k - \boldsymbol{y}_k||^2. \tag{7}$$

Here $\hat{\boldsymbol{y}}_k = -\hat{\boldsymbol{A}}^{-1}\boldsymbol{p}_k$ is the predicted response and $\#\mathcal{V}$ is the cardinality of the index set of the validation experiments determined in (5).

We select the largest ζ value that minimizes the mean RSS

$$\zeta^* \triangleq \max \arg \min_{\zeta \in [0,1]} \mathrm{RSS}(\zeta). \tag{8}$$

We later demonstrate that this selection gives efficient estimates for sufficiently informative data.

D. Performance evaluation

We asses the accuracy of the network estimates by comparing their structure to the "true" network $\breve{\boldsymbol{A}}$ used for generation of the *in silico* data. Both the existence and sign of each link is equally important for us, so for each estimated $\hat{\boldsymbol{A}}$ we measure the similarity of signed topology (SST)

$$\mathrm{SST} \triangleq \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\mathrm{sign}(\hat{a}_{ij}) == \mathrm{sign}(\breve{a}_{ij})\right). \tag{9}$$

The estimate corresponding to the ζ value at which the SST is maximized is most similar to the "true" network for the inference method, *i.e.* the efficient estimate.

E. Network inference algorithms

To demonstrate the proposed workflow we choose two commonly used network inference algorithms, Glmnet[11] and NIR[6]. Glmnet is a fast linear regression method that uses LASSO, ridge regression or a combination of them, but we only utilize LASSO. NIR is a linear regression method that uses a discrete regularisation parameter, k, representing the number of regulators per gene and does an exhaustive search of all k combinations. For comparison, we also use ordinary least squares with a cutoff (LSCO) to set all links that are weaker than the threshold $\tilde{\zeta}$ to zero

$$\hat{a}_{ij} \triangleq \begin{cases} a_{ij}^{ols} & \text{if } a_{ij}^{ols} \geq \tilde{\zeta} \\ 0 & \text{otherwise} \end{cases} \quad \text{with } \boldsymbol{A}_{ols} = -\boldsymbol{P}\boldsymbol{Y}^{\dagger}. \tag{10}$$

F. Data sets

To be able to control and systematically vary the properties of the data, as well as evaluate the performance by comparing to the "true" network, we constructed a network with 10 nodes and 25 links (Figure 5 and Table 1) and used it to generate *in silico* data. We want to simulate steady-state perturbation experiments of the type previously performed *in vivo* for inference of a ten gene network of the Snf1 signalling pathway in *S. cerevisiae* [20] and have therefore tuned the properties such that the network is biologically plausible. It is sparse, each gene has a negative self-loop representing mRNA degradation, the digraph forms one strong component, the degree of interampatteness is 145, it is stable, *i.e.* all eigenvalues are negative, and the time constants of the system are in the range 0.089 to 12 [24]. We used the network to generate data sets with different information content by varying the perturbations and noise that is added to the simulated response. We here present results for three data sets with 20, 15, and 10 experiments with SNR 35, 7, 1.5, respectively. The same white Gaussian noise realization obtained by `randn` in Matlab is after scaling of the variance added to the response in all three data sets. The perturbations were designed to counteract the intrinsic signal attenuation such that all directions of the system are excited and the singular values of the response matrices in all cases are in the range 0.77 to 1.2.

IV. Results and discussion

We have investigated the effect of the sparsity coefficient ζ on network inference using three different algorithms: Glmnet, NIR, and LSCO, as well as the performance of the proposed method to identify an optimal ζ value. The inference was done for 1000 ζ values covering the range from a full network to an empty one and we present the fraction of links (FOL), similarity of signed topology (SST), and residual sum of squares (RSS) of the estimated networks in Figure 1-4. The three data sets differ only in terms of the information content. We therefore call the most informative data set with 20 experiments and SNR 35 the "informative data set", the one with 15 experiments and SNR 7 the "partly informative data set", and the one with 10 experiments and SNR 1.5 the "uninformative data set". Using Glmnet and LSCO it is possible to recover the "true" network for certain ζ values with

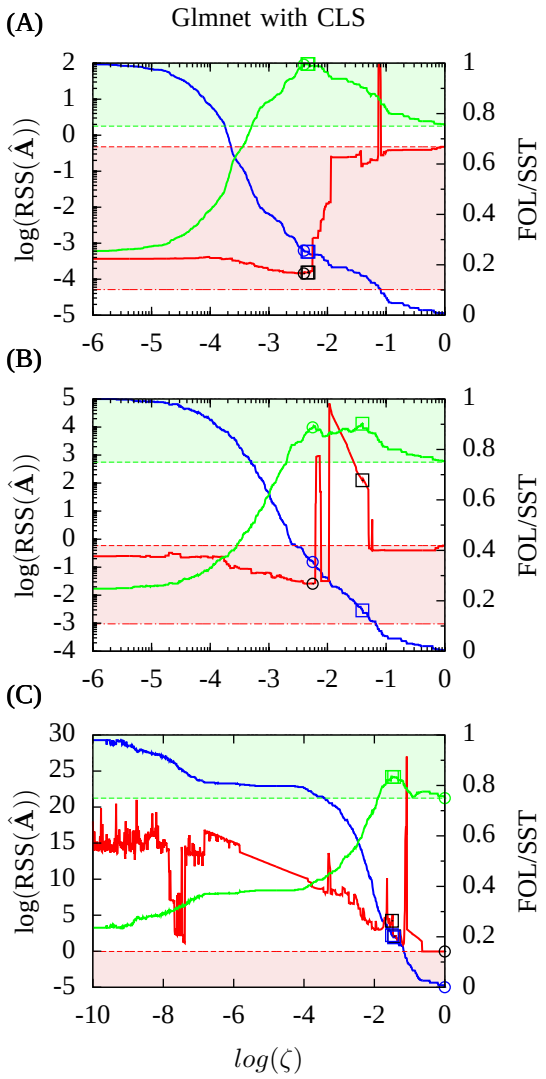


Fig. 1: Performance of Glmnet with constrained least squares (CLS) as a function of ζ. Blue curve is the mean fraction of links (FOL), green curve the mean similarity of signed topology (SST) and red curve the mean residual sum of squares (RSS). Informative data with 20 experiments (A), partly-informative data with 15 experiments (B) and uninformative data with 10 experiments (C). The best estimates marked with □ and minimum RSS marked with ◯.

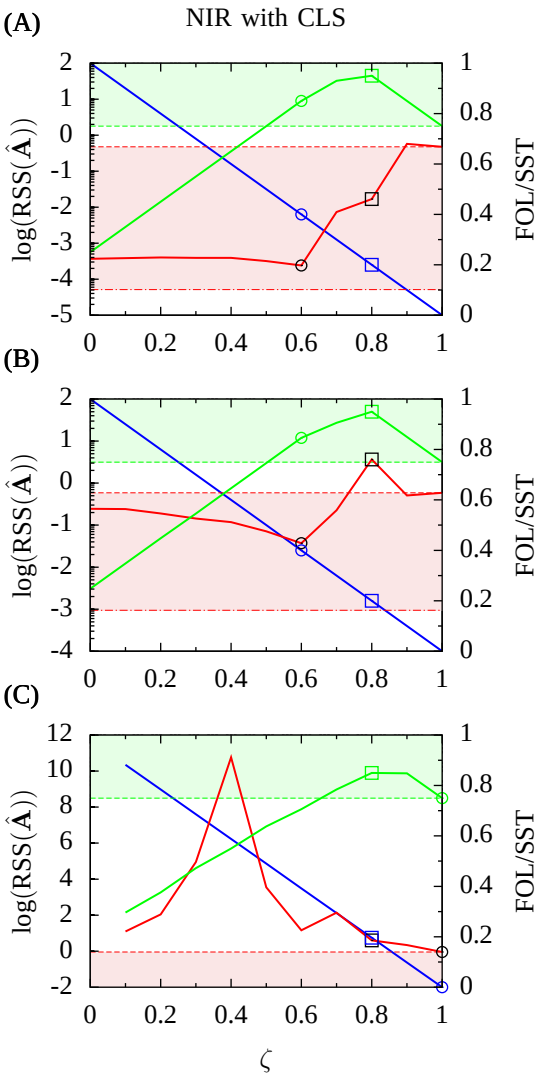


Fig. 2: Performance of NIR with CLS as a function of ζ. For explanations see Figure 1.

the informative data set (Figure 1 A and Figure 3 A), while none of the three algorithms can recover the ”true” network for any ζ value with the other two data sets (Figure 1-3 B & C). It is based on the RSS possible to select the ζ value that gives the best network estimate for the informative and partly informative data sets (Figure 1-3 A & B), but not for the uninformative data set (Figure 1-3 C).

All curves represents mean values over all the network estimates obtained by the leave one out cross optimisation (LOOCO) for each ζ value. An SST of one implies that all the network estimates contain the same links with same signs as the ”true” network, and zero that they are completely different. The shaded green area marks the region in which the network estimates are more similar to the ”true” network than an empty network. The lower boundary of this region is in other words defined by the SST of the empty network, and any network estimates with lower SST provide essentially no mechanistic understanding of the system. Previous works have typically used random networks with the correct number of links to define a lower boundary on when an inference method is useful, see *e.g.* [23], but we use the empty network, since even random networks with both the correct number of links and fraction of positive and negative ones on average have a lower SST for sufficiently sparse networks, like ours (see supplemental). The shaded red area is bounded by the mean RSS for prediction of the responses in the validation set by the empty network $\boldsymbol{A} = \boldsymbol{0}$ and the least squares estimate $\boldsymbol{A} = -\boldsymbol{P}\boldsymbol{Y}^{\dagger}$. The lower boundary is merely included for visual comparision to the least squares estimate, while any

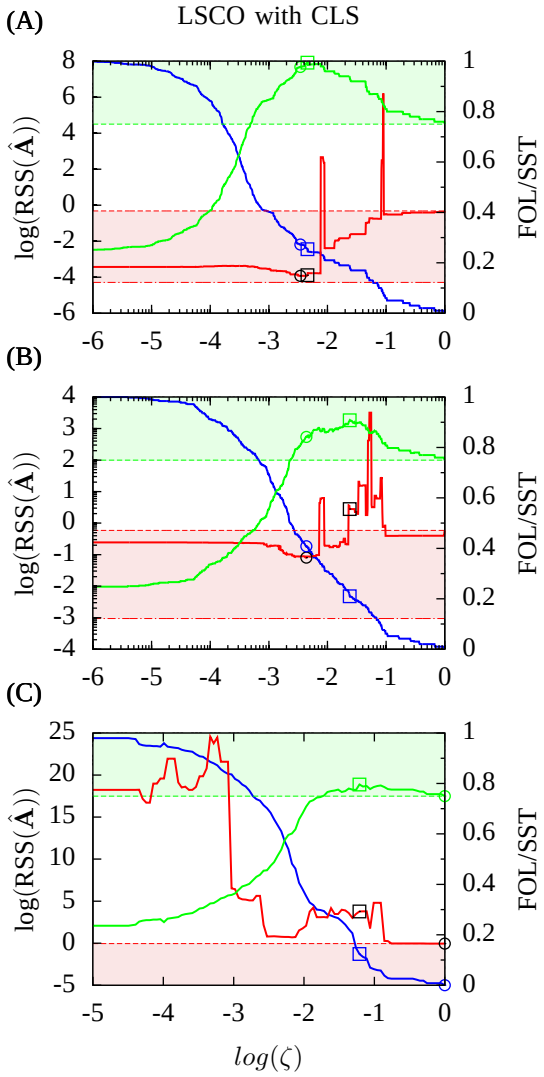


Fig. 3: Performance of LSCO with CLS as a function of ζ. For explanations see Figure 1.

network estimate with a larger RSS than the upper bound essentially is useless for data prediction.

By studying the green SST curve for the three methods in Figure 1-3 it is obvious that an efficient network estimate only is obtained for a narrow range of ζ values. To get a network estimate with an SST within 5% of the best one for the informative data, ζ must be selected within a range of 0.0078 (Glmnet), 0.1 (NIR), and 0.013 (LSCO). This confirms previous reports stating that careful selection of ζ is essential to obtain a network estimate with a structure similar to that of the ”true” network [23].

By comparing the results obtained for the three data sets, the adverse effect of noise and few samples is seen. The SST of the best estimate decreases with decreasing information content for all three methods as expected (Figure 1-3). Note that each gene will have the same indegree in the estimates given by NIR, while one gene has indegree 4, 3 genes indegree 3 and the remaining ones indegree 2 in our network (Figure 5), so NIR cannot for any ζ or data set give an estimate with SST one. This explains why the best NIR estimate for the informative data differs from the ”true” network and why the SST of the best estimate is identical for the informative and partly informative data sets. For all three methods the RSS of the estimates increases with decreasing information content for most ζ values. The RSS is for all ζ value larger than the RSS of an empty network for the uninformative data set, illustrating that a ζ value giving efficient estimates for mechanistic inference cannot be found based on the RSS when the number of experiments or SNR is too low.

It is well known that regularization introduces a bias in the parameter estimates, and several methods intended to decrease it have been published [9], [8], [19]. The effect of this bias on the RSS is in our figures best seen by comparing the red RSS curve in Figure 1 A and Figure 4 A for the informative data set. The RSS curve generated after re-estimation of the non-zero parameters using constrained least squares (CLS) in Figure 1 A has its minimum near the optimal ζ, while the minimum in Figure 4 A is located at a smaller ζ value, corresponding to an estimate that has a lower SST than an empty network. In order to find the optimal ζ based on the minimum RSS, this bias must be removed. We have chosen to do such a re-estimation of the non-zero parameters, since many network inference methods with this bias exist and we want to find a method for selection of ζ that can be wrapped around any network inference method that depends on a sparsity coefficient. In [11], [2] the previouly mentioned one standard deviation rule is used to in part compensate for the effect of bias.

Only linearly dependent experiments should be included in the validation data set to avoid predicting the outcome of experiments that the training data set lacks information about. Inclusion of such experiments in general increases the RSS for almost all ζ values and moves the minimum away from the efficient estimate. This is for the informative data set seen by comparing the RSS when all experiments are included in the validation data set used during LOOCO (Figure 4 B) and when only the experiments in $\mathcal{V}$ (5) are included (Figure 1 A). The RSS of the former is at least three orders of magnitudes larger in the interesting ζ range.

By combining selection of linearly dependent experiments and unbiased re-estimation of parameters it is for sufficiently informative data based on the RSS of the validation data set possible to chose a ζ value close to the optimal one, at least for the three inference algorithms that we have investigated here (Figure 1-3). The proposed ζ selection method hence optimizes the accuracy of the inferred network, with a tendency towards having more false positives than negatives when the ”true” network cannot be found. This tendency is best seen for NIR. The ”true” network has four links on one row, hence all NIR estimates for a ζ larger than 0.6 lack at least one true link, which increases the RSS and prevents us from finding the optimal ζ. Instead our method identifies the largest ζ value at which the estimate contains all links present in the ”true” network and some false positives.

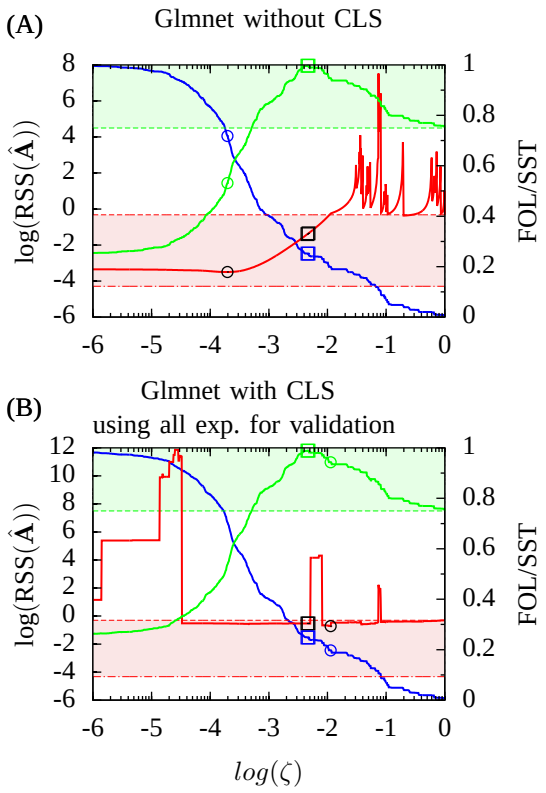


Fig. 4: Performance of Glmnet without CLS as a function of ζ for the informative data (A) and with CLS when all experiments are included in the validation set (B). For explanations see Figure 1.

V. Conclusion

Network estimates that are structurally close to the "true" network are only obtained within a narrow range of ζ values and estimates that are more different than the empty network are obtained for poor choices. The range of good estimates depends on both the algorithm and data set, which makes data based selection of ζ imperative. We demonstrate that the ζ value yielding highly accurate network estimates can be determined by minimization of the prediction error in a leave one out cross optimization scheme for data with only 1.5 to 2 times as many experiments as variables.

We also demonstrate the adverse effect of noise and too few samples, and how these problems can be detected based on the residual sum of squares becoming larger than for an empty network. Correct selection of ζ requires that non-informative data samples are excluded from the validation set, as well as re-estimation of biased parameters. We therefore include these two steps in the proposed method for ζ optimization. The proposed method can be wrapped around almost any regulatory network inference method and works well for sufficiently informative data, thus improving the workflow from data to network models that provide a better understanding of the biological mechanism.

References

[1] Mukesh Bansal, Vincenzo Belcastro, Alberto Ambesi-Impiombato, and Diego Di Bernardo, *How to infer gene networks from expression profiles*, Mol Syst Biol **3** (2007), no. 1, 78.

[2] Richard Bonneau, David J Reiss, Paul Shannon, Marc Facciotti, Leroy Hood, Nitin S Baliga, and Vesteinn Thorsson, *The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo.*, Genome biology **7** (2006), no. 5, R36 (en).

[3] E.J. Candes and M.B. Wakin, *An Introduction To Compressive Sampling*, IEEE Signal Processing Magazine **25** (2008), no. 2, 21–30.

[4] Emmanuel J. Candès and Yaniv Plan, *Near-ideal model selection by ℓ_1 minimization*, The Annals of Statistics **37** (2009), no. 5A, 2145–2177.

[5] Ozgur Demir-Kavuk, Mayumi Kamada, Tatsuya Akutsu, and Ernst Walter Knapp, *Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features.*, BMC bioinformatics **12** (2011), no. 1, 412 (en).

[6] D Di Bernardo, T S Gardner, and J J Collins, *Robust identification of large genetic networks.*, Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing (2004), 486–97.

[7] J. Fan and R. Li, *Statistical challenges with high dimensionality: feature selection in knowledge discovery*, Proceedings of the International Congress of Mathematicians (Madrid, Spain), European Mathematical Society, 2006.

[8] Jianqing Fan, Yang Feng, and Yichao Wu, *NETWORK EXPLORATION VIA THE ADAPTIVE LASSO AND SCAD PENALTIES.*, The annals of applied statistics **3** (2009), no. 2, 521–541.

[9] Jianqing Fan and Runze Li, *Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties*, Journal of the American Statistical Association **96** (2001), no. 456, 1348–1360.

[10] Jianqing Fan and Jinchi Lv, *A Selective Overview of Variable Selection in High Dimensional Feature Space (Invited Review Article)*, Statistica Sinica **20** (2010), no. 1, 101–148.

[11] Jerome Friedman, Trevor Hastie, and Rob Tibshirani, *Regularization Paths for Generalized Linear Models via Coordinate Descent.*, Journal of statistical software **33** (2010), no. 1, 1–22.

[12] Timothy S Gardner, Diego Bernardo, David Lorenz, and James J Collins, *Inferring Genetic Networks and Identifying Compound Mode of Action via Expression Profiling*, Science **301** (2003), no. July, 102–105.

[13] Hendrik Hache, Hans Lehrach, and Ralf Herwig, *Reverse engineering of gene regulatory networks: a comparative study.*, EURASIP journal on bioinformatics & systems biology (2009), 617281.

[14] Hendrik Hache, Christoph Wierling, Hans Lehrach, and Ralf Herwig, *GeNGe: systematic generation of gene regulatory networks*, Bioinformatics **25** (2009), no. 9, 1205–1207.

[15] Feng He, Rudi Balling, and An-Ping Zeng, *Reverse engineering and verification of gene networks: principles, assumptions, and limitations of present methods and future perspectives.*, Journal of biotechnology **144** (2009), no. 3, 190–203.

[16] Michael Hecker, Sandro Lambeck, Susanne Toepfer, Eugene van Someren, and Reinhard Guthke, *Gene regulatory network inference: data integration in dynamic models-a review.*, Bio Systems **96** (2009), no. 1, 86–103.

[17] Rebecka Jörnsten, Tobias Abenius, Teresia Kling, Linnéa Schmidt, Erik Johansson, Torbjörn E M Nordling, Bodil Nordlander, Chris Sander, Peter Gennemark, Keiko Funa, Björn Nilsson, Linda Lindahl, and Sven Nelander, *Network modeling of the transcriptional effects of copy number aberrations in glioblastoma.*, Molecular systems biology **7** (2011), no. 486, 486.

[18] A Julius, M Zavlanos, S Boyd, and G J Pappas, *Genetic network identification using convex programming.*, IET systems biology **3** (2009), no. 3, 155–166.

[19] Heng Lian, *Shrinkage Tuning Parameter Selection in Precision Matrices Estimation*, Journal of Statistical Planning and Inference **141** (2009), no. 8, 2839–2848.

[20] David R Lorenz, Charles R Cantor, and James J Collins, *A network biology approach to aging in yeast.*, Proceedings of the National Academy of Sciences of the United States of America **106** (2009), no. 4, 1145–50.

[21] Daniel Marbach, Robert J Prill, Thomas Schaffter, Claudio Mattiussi, Dario Floreano, and Gustavo Stolovitzky, *Revealing strengths and weaknesses of methods for gene network inference.*, Proceedings of the National Academy of Sciences of the United States of America **107** (2010), no. 14, 6286–91.

[22] P. Mendes, W. Sha, and K. Ye, *Artificial gene networks for objective comparison of analysis algorithms*, Bioinformatics **19** (2003), no. Suppl 2, ii122–ii129.

[23] Patricia Menéndez, Yiannis a I Kourmpetis, Cajo J F ter Braak, and Fred a van Eeuwijk, *Gene regulatory networks from multifactorial perturbations using Graphical Lasso: application to the DREAM4 challenge.*, PloS one **5** (2010), no. 12, e14147.

[24] T E M Nordling and E W Jacobsen, *Interampatteness - a generic property of biochemical networks.*, IET systems biology **3** (2009), no. 5, 388.

[25] Thomas Schaffter, Daniel Marbach, and Dario Floreano, *GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods.*, Bioinformatics (Oxford, England) (2011), btr373–.

[26] P Stoica and Y Selen, *Model-order selection: a review of information criterion rules*, IEEE Signal Processing Magazine **21** (2004), no. 4, 36–47.

[27] Vesteinn Thorsson and H Michael, *Statistical Applications in Genetics and Molecular Biology Reverse Engineering Galactose Regulation in Yeast through Model Selection Reverse Engineering Galactose Regulation in Yeast through Model Selection* , Statistical Applications in Genetics and Molecular Biology **4** (2005), no. 1.

[28] R Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society Series B Methodological **58** (1996), no. 1, 267–288.

[29] Hansheng Wang, *Tuning parameter selectors for the smoothly clipped absolute deviation method*, **94** (2009), no. 3, 553–568.

[30] Hansheng Wang, Runze Li, and Chih-Ling Tsai, *Tuning parameter selectors for the smoothly clipped absolute deviation method.*, Biometrika **94** (2007), no. 3, 553–568.

S.6. SUPPLEMENTARY MATERIAL

We here provide additional details, in particular, about the network used for data generation and simulated data. This information makes it possible to reproduce our figures. We also give additional comments on the results for interested readers.

A. In silico data sets and network properties

The gene regulatory network that we use to generate the *in silico* data is shown in Figure 5 and the strength of all interactions in Table 1. The true network $\check{\boldsymbol{A}}$ has 10 states, each representing the mRNA abundance of a gene. The states form a stable system with self-degradation of each mRNA. One path through all states exists, which means that it consists of one strong component.

We generated the responses in Matlab using (S2), with $\boldsymbol{F} = \boldsymbol{0}$ and $\boldsymbol{E}$ being white Gaussian noise with the variance given in Table 2. The perturbation matrices $\boldsymbol{P}$ and obtained response with noise $\boldsymbol{Y}$ are given in Table 5a-tab:noneinf-expr. Further properties of the data sets are shown in Table 2.

B. Network inference and selection of ζ

The range of the actual $\tilde{\zeta}$ values used in the three algorithms is given in Table 3. The RSS, FOL, and SST for all three algorithms and data sets without re-estimation of the model parameters using CLS is shown in Figure 6-8. The RSS, FOL, and SST is for Glmnet both with and without CLS when all experiments are included in the validation set shown in Figure 9. By comparing the RSS in this figure to that for inclusion of linearly dependent experiments one sees the importance of selection of validation experiments. To make it possible to asses the benefits of using cross optimisation we also include Figure 10 and 11 where all experiments were used for training and then the mean RSS over all experiments calculated. The minimum RSS occurs in this case for full or almost full network and the parameters are over fitted to the noise in the data.

C. Similarity of signed topology of random networks

Previous works have typically used random networks with the correct number of links to define a lower boundary on when an inference method is useful, see *e.g.* [23], but we use the empty network, since even random networks with both the correct number of links and fraction of positive and negative ones on average have a lower SST for sufficiently sparse networks, like ours. Based on Monte Carlo simulations in Matlab we found that the mean and median of SST is 0.6. And if we make all all diagonal elements negative in the random networks then it increases to 0.74, while it is 0.75 for an empty network.

D. Constrained least squares

To account for the fact that that our estimated network does not in all cases have a full rank and the equation to the minimisation problem presented for CLS can not be solved we added in these cases a small regularisation term of $\ell = 10^{-15}$ to the correction term.

E. minimum RSS for informative data

It can be observed in the Figure 1 that the minimum RSS is slightly to the left of the optimal point. This can be explained by the fact that in the informative data (Figure 5b) validation set, node G is not perturbed. Therefore the algorithm can remove links in the "true" network with small effect on the prediction error.

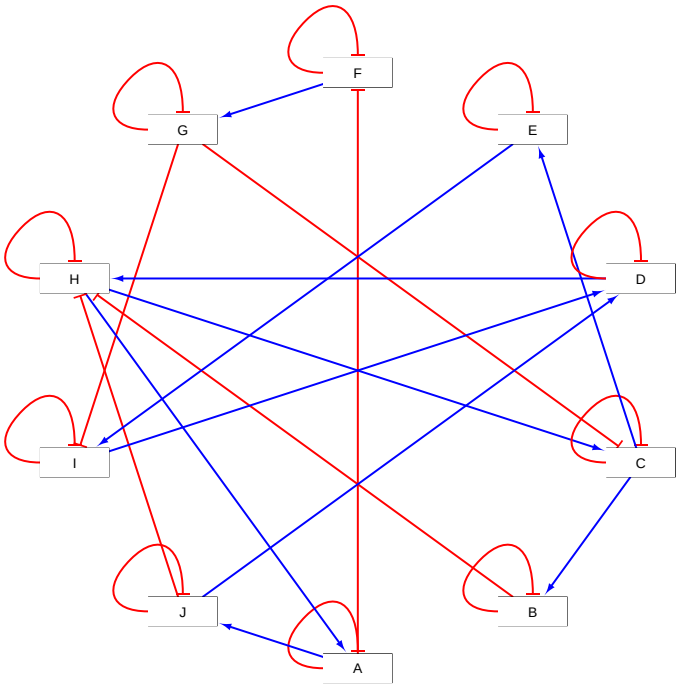


Fig. 5: Network structure

TABLE 1: Network interactions

Source Node	Interaction Type	Target Node	Connectivity	Interaction Strength
A	Inhibition	A	-0.507	0.507
H	Activation	A	0.67	0.67
B	Inhibition	B	-0.943	0.943
C	Activation	B	1.157	1.157
C	Inhibition	C	-7.535	7.535
G	Inhibition	C	-0.5	0.5
H	Activation	C	2.876	2.876
D	Inhibition	D	-11.22	11.22
I	Activation	D	0.472	0.472
J	Activation	D	0.995	0.995
C	Activation	E	1.087	1.087
E	Inhibition	E	-0.99	0.99
A	Inhibition	F	-0.1	0.1
F	Inhibition	F	-0.842	0.842
F	Activation	G	0.1	0.1
G	Inhibition	G	-0.085	0.085
B	Inhibition	H	-0.279	0.279
D	Activation	H	4.171	4.171
H	Inhibition	H	-2.176	2.176
J	Inhibition	H	-0.25	0.25
E	Activation	I	0.148	0.148
G	Inhibition	I	-0.05	0.05
I	Inhibition	I	-0.584	0.584
A	Activation	J	0.181	0.181
J	Inhibition	J	-0.808	0.808

TABLE 2: Data set properties: Signal to noise ratio, SNR, noise variance, λ and smallest singular values $\sigma_N(\boldsymbol{Y})$ and $sigma_N\boldsymbol{P}$, and largest singular values $\sigma_1(\boldsymbol{Y})$ and $sigma_1\boldsymbol{P}$

	Informative	Partly informative	Uninformative
# experiments	20	15	10
SNR	35	7	1.5
λ	8.7224e-06	2.8399e-04	0.0049
$\sigma_N(\boldsymbol{Y})$	7.8284e-01	7.7531e-01	6.5951e-01
$\Sigma_N(\boldsymbol{P})$	8.4024e-02	8.3036e-02	8.4392e-02
$\sigma_1(\boldsymbol{Y})$	1.2357e+00	1.2283e+00	1.3784e+00
$\Sigma_1(\boldsymbol{P})$	1.1998e+01	1.1986e+01	1.2041e+01

TABLE 3: ζ conversion

	Informative (min/max)	Partly informative (min/max)	Uninformative (min/max)
Glmnet	2.3144e-06/2.3144	2.9644e-06/2.9644	3.4851e-06/3.4851
LSCO	1.1292e-05/11.2923	1.1631e-05/11.6308	1.3812e-05/13.8121
NIR	10/0	10/0	9/0

TABLE 4: Optimal SST and 95% of optimal SST and corresponding ζ values, RSS, FOL, SST and Matthews correlation coefficient (MCC). -5% corresponds to smallest ζ closest to 5% lower than optimal SST. +5% corresponds to largest ζ closest to 5% lower than optimal SST.

	ζ	RSS	FOL	SST	MCC
Glmnet info					
-5%	0.00280187	0.0128184	0.301667	0.948333	0.878854
optimal	0.00473888	0.0445398	0.251667	0.996667	0.991283
+5%	0.0105688	0.436417	0.199167	0.9475	0.859023
Glmnet partly					
-5%	0.0102804	92.6268	0.286	0.86	0.646725
optimal	0.0393183	1953.2	0.162	0.904	0.736548
+5%	0.0579112	1.18043e+06	0.12	0.866	0.625866
Glmnet uni					
-5%	0.0177112	472.361	0.262	0.784	0.490044
optimal	0.0337698	2128.13	0.2	0.826	0.556872
+5%	0.0849042	1268.1	0.104	0.782	0.319335
LSCO info					
-5%	0.002688	0.00014431	0.303333	0.946667	0.875584
optimal	0.0045463	0.000128145	0.254167	0.994167	0.984788
+5%	0.0157833	0.0127847	0.193333	0.943333	0.847922
LSCO partly					
-5%	0.0102804	0.192458	0.308	0.862	0.668126
optimal	0.0238989	2.78229	0.21	0.908	0.749516
+5%	0.0563314	3.77737	0.144	0.86	0.61867
LSCO uni					
-5%	0.0305386	2991.37	0.266	0.776	0.431465
optimal	0.0613591	5835.49	0.124	0.796	0.389913
+5%	0.559081	0.894621	0.008	0.758	0.139262
NIR info					
-5%	0.7	0.00735917	0.3	0.93	0.831522
optimal	0.8	0.0168469	0.2	0.95	0.866025
+5%	0.9	0.572672	0.1	0.85	0.57735
NIR partly					
-5%	0.7	0.225786	0.3	0.906	0.771048
optimal	0.8	3.64408	0.2	0.95	0.866025
+5%	0.9	0.507342	0.1	0.85	0.57735
NIR uni					
-5%	0.7	138.251	0.294	0.784	0.469483
optimal	0.8	3.95052	0.196	0.85	0.576567
+5%	0.9	2.18126	0.098	0.848	0.570821

TABLE 5: Informative data set. Bold numbers indicate samples in validation set.

(a) Expression matrix, $\boldsymbol{Y}$

	1	**2**	**3**	**4**	**5**	6	**7**	8	**9**	**10**
η	1.30	1	0.89	1.24	1.65	0.40	1.38	0.04	1.40	1.28
A	0.0823	0.0497	0.0391	-0.2440	0.6598	-0.1020	-0.2752	-0.0047	-0.5339	-0.0663
B	-0.5522	0.0176	0.0159	-0.0959	0.0124	0.6028	-0.0991	-0.0036	-0.0048	-0.0264
C	-0.4447	0.0153	0.0152	-0.0756	0.0119	-0.7857	-0.0815	-0.0071	-0.0038	-0.0232
D	0.0003	-0.6874	-0.5500	-0.1014	0.0167	-0.0024	-0.1193	-0.0018	-0.0113	-0.0354
E	-0.4907	0.0191	0.0156	-0.0756	0.0048	0.0027	-0.0853	-0.0034	-0.0097	-0.0233
F	-0.0084	-0.0069	-0.0038	0.0290	-0.0736	0.0171	0.0343	-0.0022	0.0647	0.0020
G	-0.0113	-0.0036	-0.0076	0.0360	-0.0962	0.0154	0.0410	0.0023	0.0734	0.0098
H	0.0602	0.0326	0.0289	-0.1864	0.0025	-0.0802	-0.2073	-0.0009	-0.0021	-0.0510
I	-0.1226	0.0023	-0.0019	-0.0288	0.0139	0.0076	-0.0296	-0.0007	-0.0054	-0.6854
J	0.0243	0.0066	0.0063	-0.0499	0.1478	-0.0155	-0.0588	0.0016	-0.1163	-0.0101
	11	12	13	14	15	16	17	**18**	**19**	**20**
η	1.00	1.08	0.89	0.29	0.03	0.76	0.82	1.32	0.86	1.25
A	0.1028	-0.0050	-0.0036	-0.0014	0.0018	0.0341	0.0690	0.0710	-0.1401	-0.0203
B	-0.5205	0.0686	0.0422	0.0010	-0.0002	0.0120	0.0272	0.0210	-0.0012	-0.0128
C	-0.4211	0.0561	0.0337	-0.0023	-0.0006	0.0164	0.0177	0.0242	-0.0050	-0.0061
D	0.0055	0.0051	-0.0013	-0.0032	0.0055	-0.4400	0.0902	0.0325	-0.0011	-0.0078
E	0.5455	0.0589	0.0433	0.0029	0.0005	0.0063	0.0181	0.0215	-0.0055	-0.6095
F	-0.0038	-0.0060	-0.0024	-1.0014	-0.0018	-0.0036	-0.0070	-0.0108	0.0159	-0.0008
G	-0.0162	-0.8036	-0.5762	0.0122	-0.0058	-0.0048	-0.0116	-0.0124	0.0192	0.0045
H	0.0812	-0.0090	-0.0043	0.0008	0.0042	0.0270	0.0475	0.0557	0.9575	-0.0168
I	0.1383	0.0888	0.0576	0.0025	-0.0016	-0.0003	0.0096	0.6985	-0.0059	-0.1575
J	0.0193	-0.0018	-0.0048	-0.0042	-0.0029	0.0065	0.9823	0.0162	-0.0293	-0.0098

(b) Perturbation matrix, $\boldsymbol{P}$

	1	**2**	**3**	**4**	**5**	6	**7**	8	**9**	**10**
η	42.14	123.76	109.34	22.62	0.32	56.64	25.09	0.07	0.28	0.16
A	0	0	0	0	0.3300	0	0	0	-0.2660	0
B	0	0	0	0	0	1.4770	0	0	0	0
C	-3.5550	0	0	0	0	-5.6840	0	0	0	0
D	0	-7.7710	-6.1450	-1.1230	0	0	-1.2540	0	0	0
E	0	0	0	0	0	0.8610	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	0
H	0	2.9740	2.3550	0	0	0	0	-0.0080	0	0
I	0	0	0	0	0	0	0	0	0	-0.3960
J	0	0	0	0	0	0	0	0	0	0
	11	12	13	14	15	16	17	**18**	**19**	**20**
η	42.29	0.01	0.01	0.01	0.00	95.12	0	0.16	50.58	1.11
A	0	0	0	0	0	0	0	0	-0.7150	0
B	0	0	0	0	0	0	0	0	0	0
C	-3.4010	0	0	0	0	0	0	0	-2.7680	0
D	0	0	0	0	0	-4.9700	0	0	0	0
E	0.9960	0	0	0	0	0	0	0	0	-0.5980
F	0	0	0	-0.8440	-0.0030	0	0	0	0	0
G	0	-0.0690	-0.0490	0.1010	0	0	0	0	0	0
H	0	0	0	0	0	1.9040	0	0	2.0940	0
I	0	0	0	0	0	0	0	0.4040	0	0
J	0	0	0	0	0	0	0.7810	0	0	0

TABLE 6: Partly informative data set. Bold numbers indicate samples in validation set.

(a) Expression matrix, $\boldsymbol{Y}$

	1	2	**3**	4	**5**	6	7	8	9	10
η	0.85	0.59	1.07	0.55	1.25	0.53	0.22	0.33	0.32	0.89
A	-0.0979	0.0536	-0.1061	-0.1752	0.9245	0.0349	0.0483	-0.0076	-0.0830	-0.0199
B	-0.0117	-0.0356	-0.0447	0.9847	0.0194	0.0275	0.0111	0.0718	-0.0196	0.0905
C	-0.0243	-0.0470	-0.0448	-0.0599	0.0152	-0.0100	0.0068	0.9589	-0.0206	0.0352
D	-0.3855	0.0618	0.6442	-0.0018	0.0203	0.0847	0.0432	0.0172	-0.0101	-0.0393
E	-0.0314	-0.0356	-0.0273	-0.0655	0.0107	0.0257	-0.0114	-0.0156	-0.0215	0.0558
F	0.0294	-0.0268	0.0170	0.0423	-0.1093	-0.0173	-0.0224	-0.0108	0.0235	-0.0043
G	0.0110	-0.0013	0.0284	0.0435	-0.1399	-0.0071	0.0228	0.0034	0.0072	-0.7398
H	-0.0594	-0.9901	-0.0851	-0.1473	0.0375	0.0260	0.0076	-0.0033	-0.0599	-0.0224
I	0.0162	-0.0121	-0.0146	-0.0139	0.0009	0.0096	-0.0027	-0.0372	0.0192	0.0563
J	0.0094	-0.0078	-0.0050	-0.0213	0.1935	0.9792	0.0084	0.0041	-0.0332	-0.0020

	11	**12**	13	14	**15**
η	0.87	0.79	0.36	0.84	1.03
A	0.0153	-0.0417	0.0339	0.1303	-0.0320
B	0.0433	0.0025	-0.0048	0.0602	0.0206
C	0.0425	-0.0198	-0.0364	0.0487	-0.0237
D	0.0216	-0.0140	-0.0096	0.0478	-0.6244
E	0.0435	-0.9767	0.0197	0.0406	-0.0032
F	-0.0102	0.0157	0.9530	-0.0462	0.0297
G	-0.6193	0.0074	-0.0150	0.0225	-0.0012
H	0.0127	-0.0342	-0.0093	0.0594	-0.0516
I	0.0993	-0.2368	-0.0346	0.9632	-0.0174
J	-0.0144	-0.0002	0.0213	0.0135	-0.0119

(b) Perturbation matrix, $\boldsymbol{P}$

	1	2	**3**	4	**5**	6	7	8	9	10
η	79.64	35.61	106.09	1.08	0.32	0	2.61	21.86	1.26	0.01
A	0	0.6830	0	0	0.4640	0	0	0	0	0
B	0	0	0	0.9700	0	0	0	-1.1070	0	0
C	0	2.6300	0	0	0	0	0	7.5150	0	0
D	-4.6250	0	7.4050	0	0	0	0.1370	0	0	0
E	0	0	0	0	0	0	0	-1.0940	0	0
F	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	-0.0630
H	1.5830	-2.1560	-2.9410	0	0	0	0	0	-0.1370	0
I	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0.7800	0	0	0	0

	11	**12**	13	14	**15**
η	0.01	1.03	0.01	0	102.69
A	0	0	0	0	0
B	0	0	0	0	0
C	0	0	0	0	0
D	0	0	0	0	-7.0150
E	0	-0.9380	0	0	0
F	0	0	0.8300	0	0
G	-0.0550	0	-0.0990	0	0
H	0	0	0	0	2.5430
I	0	0	0	0.5640	0
J	0	0	0	0	0

TABLE 7: Uninformative data. Bold numbers indicate samples in validation set.

(a) Expression matrix, $\boldsymbol{Y}$

	1	**2**	**3**	**4**	**5**	6	7	8	9	**10**
η	1.03	0.95	0.92	0.72	0.80	0.94	1.51	0.74	0.66	0.93
A	0.0378	0.0734	0.9885	0.0433	-0.0320	0.0070	0.0231	-0.0459	0.1084	-0.0635
B	0.1283	-0.7536	-0.0692	-0.0408	-0.0260	0.0302	0.0388	0.1331	-0.0277	0.0137
C	-0.1594	0.1012	0.0633	-0.0427	0.0151	-0.0652	0.0686	0.0520	-0.9847	0.0560
D	0.0606	-0.0004	0.1340	-0.0570	0.0113	0.0088	0.1206	0.0668	0.1120	-0.8205
E	0.0218	0.1054	0.0485	-0.1714	-1.0195	0.0217	0.1413	0.0182	-0.0470	-0.0413
F	-1.0849	-0.0344	-0.0390	0.1035	0.0008	0.0998	-0.0058	-0.0983	0.0480	0.0111
G	-0.0221	-0.0322	-0.0804	0.0256	-0.0081	-0.0636	-0.1189	-1.0997	0.0575	-0.0082
H	0.0241	0.2321	-0.0106	0.9413	0.0254	0.0774	0.0233	0.0324	-0.0107	-0.1229
I	0.2508	0.1151	0.0355	0.1052	-0.1663	-0.0094	0.9040	0.0914	0.0177	-0.0292
J	0.1948	0.1374	0.1555	-0.1247	0.0725	1.0579	0.1877	-0.0144	-0.0801	-0.1224

(b) Perturbation matrix, $\boldsymbol{P}$

	1	**2**	**3**	**4**	**5**	6	7	8	9	**10**
η	0.01	1.08	0.32	29.19	1.03	0	0	0.01	21.83	9.16
A	0	0	0.4700	-0.6760	0	0	0	0	0	0
B	0	-0.9700	0	0	0	0	0	0	1.1100	0
C	0	0	0	-2.6140	0	0	0	0	-7.5440	0
D	0	0	0	0	0	0	0	0	0	-11.2510
E	0	0	0	0	-0.9410	0	0	0	1.0950	0
F	-0.8360	0	0	0	0	0	0	0	0	0
G	0.1000	0	0	0	0	0	0	-0.0850	0	0
H	0	0	0	2.1740	0	0	0	0	0	4.2120
I	0	0	0	0	0	0	0.5660	0	0	0
J	0	0	0	0	0	0.7790	0	0	0	0

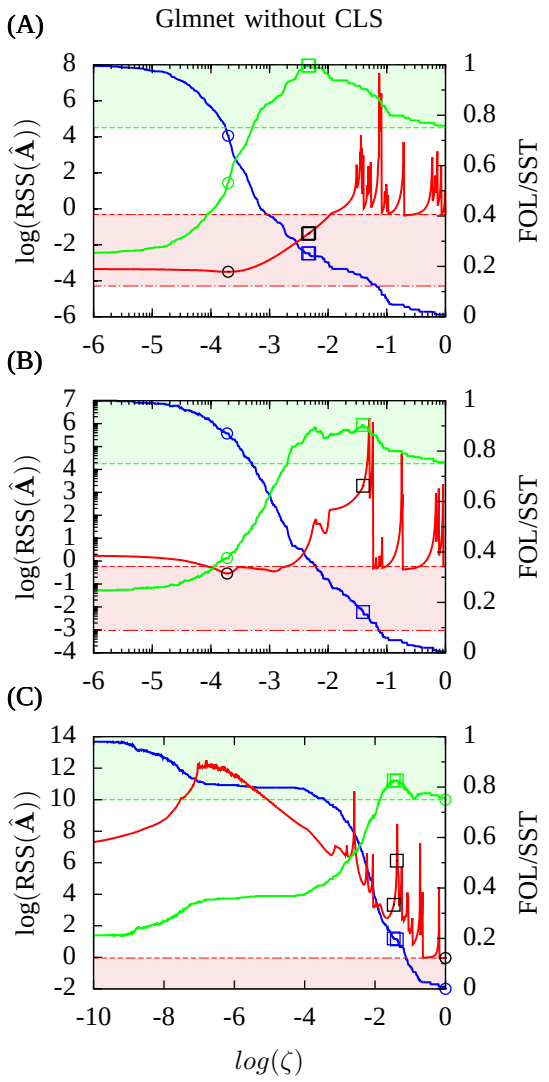


Fig. 6: Performance of Glmnet without CLS as a function of ζ. Blue curve is the mean fraction of links (FOL), green curve the mean similarity of signed topology (SST) and red curve the mean residual sum of squares (RSS). Informative data with 20 experiments (A), partly-informative data with 15 experiments (B) and uninformative data with 10 experiments (C). The best estimates marked with □ and minimum RSS marked with ◯.

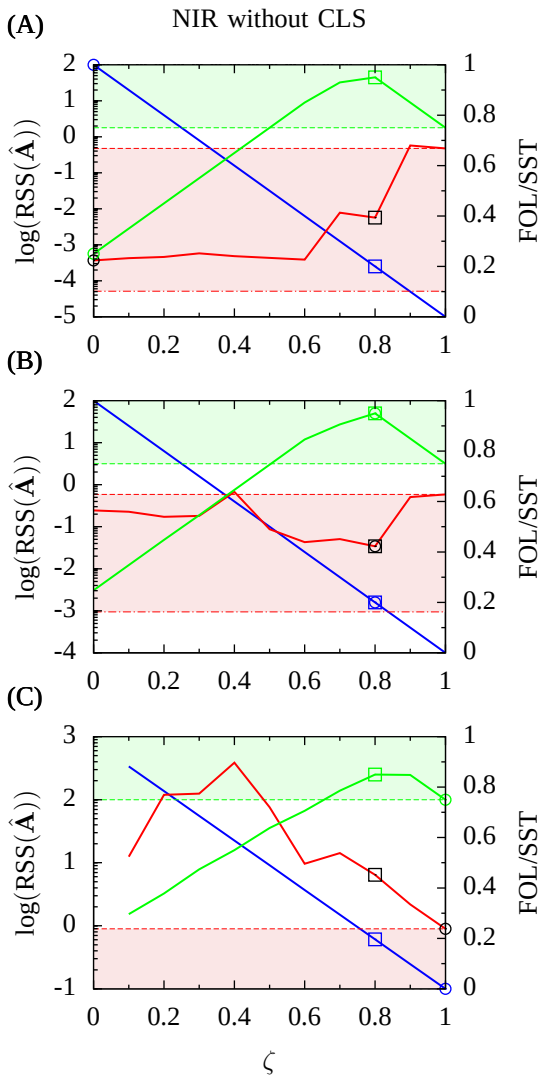


Fig. 7: Performance of NIR without CLS as a function of ζ. Blue curve is the mean fraction of links (FOL), green curve the mean similarity of signed topology (SST) and red curve the mean residual sum of squares (RSS). Informative data with 20 experiments (A), partly-informative data with 15 experiments (B) and uninformative data with 10 experiments (C). The best estimates marked with □ and minimum RSS marked with ◯.

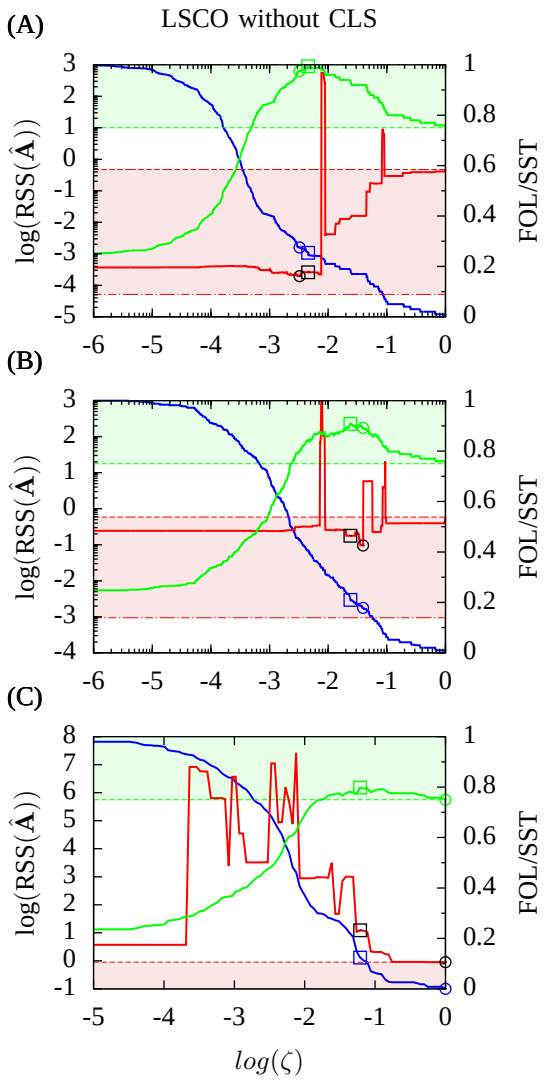
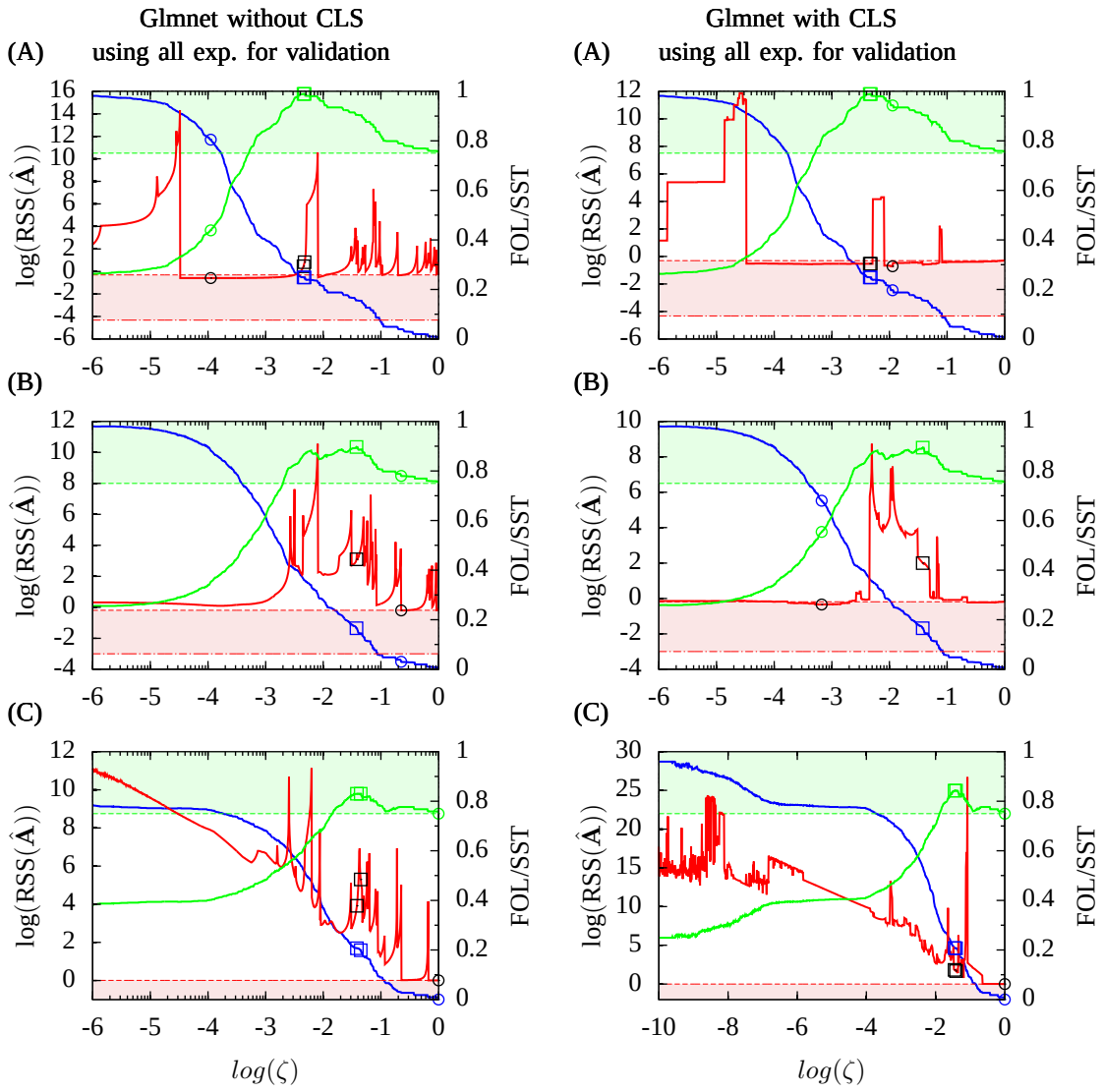


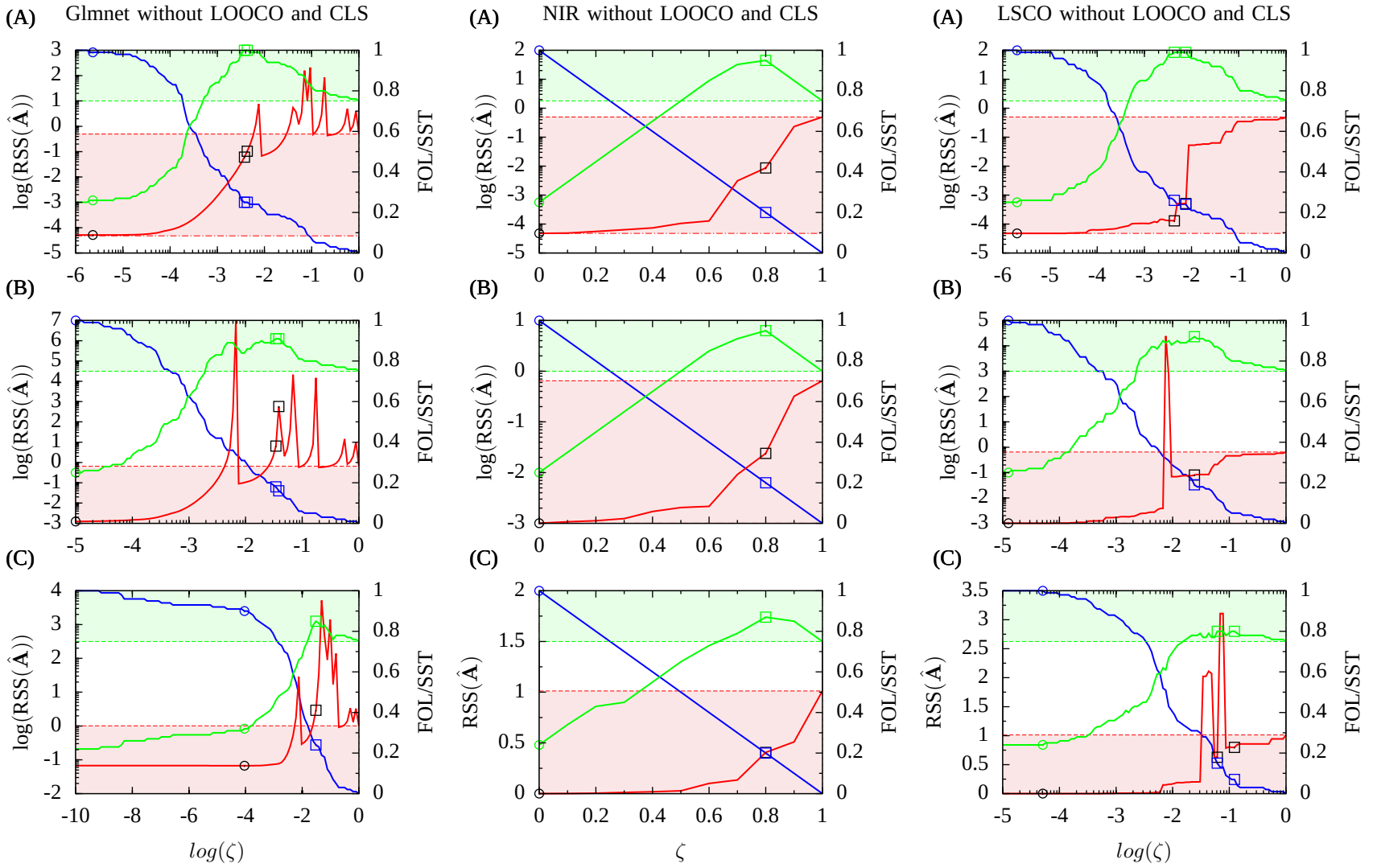
Fig. 8: Performance of LSCO without CLS as a function of ζ. Blue curve is the mean fraction of links (FOL), green curve the mean similarity of signed topology (SST) and red curve the mean residual sum of squares (RSS). Informative data with 20 experiments (A), partly-informative data with 15 experiments (B) and uninformative data with 10 experiments (C). The best estimates marked with □ and minimum RSS marked with ◯.



(a) Glmnet without CLS

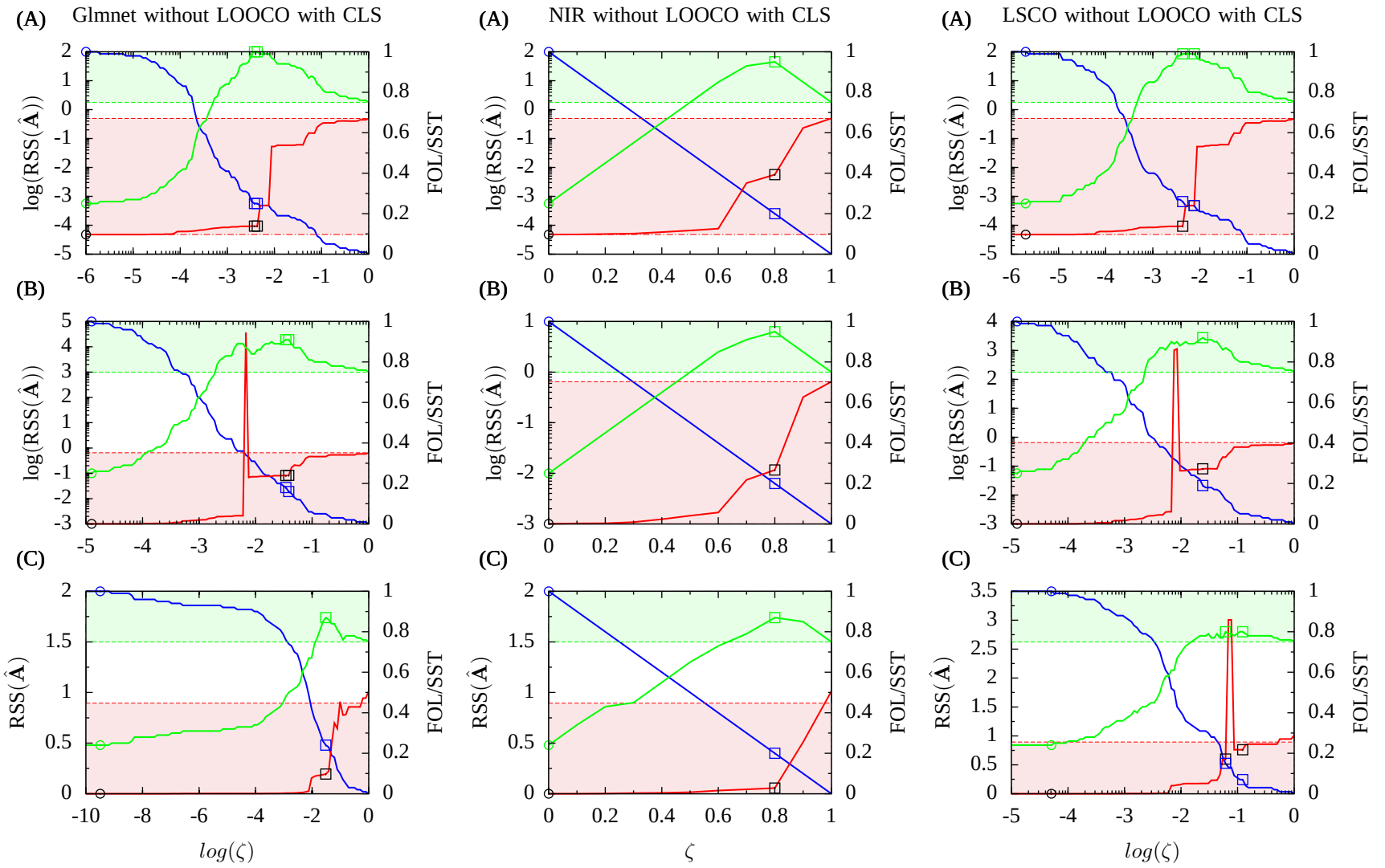
(b) Glmnet with CLS

Fig. 9: Performance of Glmnet when all samples are used in validation set, without and with constrained least squares (CLS) as a function of ζ. Blue curve is the mean fraction of links (FOL), green curve the mean similarity of signed topology (SST) and red curve the mean residual sum of squares (RSS). Informative data with 20 experiments (A), partly-informative data with 15 experiments (B) and uninformative data with 10 experiments (C). The best estimates marked with □ and minimum RSS marked with ◯.



(a) Glmnet (b) NIR (c) LSCO

Fig. 10: Glmnet, NIR and LSCOs' performance without LOOCO and constrained least squares as a function of ζ. Blue curve is the mean fraction of links (FOL), green curve the mean similarity of signed topology (SST) and red curve the mean residual sum of squares (RSS). Informative data with 20 experiments (A), partly-informative data with 15 experiments (B) and uninformative data with 10 experiments (C). The best estimates marked with □ and minimum RSS marked with ◯.



(a) Glmnet (b) NIR (c) LSCO

Fig. 11: Glmnet, NIR and LSCOs' performance without LOOCO and with constrained least squares as a function of ζ. Blue curve is the mean fraction of links (FOL), green curve the mean similarity of signed topology (SST) and red curve the mean residual sum of squares (RSS). Informative data with 20 experiments (A), partly-informative data with 15 experiments (B) and uninformative data with 10 experiments (C). The best estimates marked with □ and minimum RSS marked with ◯.