

RESEARCH ARTICLE

# Elastic-Net Copula Granger Causality for Inference of Biological Networks

Mohammad Shaheryar Furqan<sup>1,2\*</sup>, Mohammad Yakoob Siyal<sup>1</sup>

**1** School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore, **2** INFINITUS, Infocomm Centre of Excellence, Nanyang Technological University, Singapore, Singapore

\* [mohammad022@e.ntu.edu.sg](mailto:mohammad022@e.ntu.edu.sg)



**OPEN ACCESS**

**Citation:** Furqan MS, Siyal MY (2016) Elastic-Net Copula Granger Causality for Inference of Biological Networks. PLoS ONE 11(10): e0165612. doi:10.1371/journal.pone.0165612

**Editor:** Enrique Hernandez-Lemus, Instituto Nacional de Medicina Genomica, MEXICO

**Received:** April 15, 2016

**Accepted:** October 15, 2016

**Published:** October 28, 2016

**Copyright:** © 2016 Furqan, Siyal. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** HeLa cell Genetic was acquired from a published article of Whitfield et al. and can be accessed from <http://genomewww.stanford.edu/Human-CellCycle/Hela/data.shtml>. Similarly, DREAM 4 dataset was acquired from a freely-accessed server at <http://gnw.sourceforge.net/dreamchallenge.html#dream4challenge>.

**Funding:** The author(s) received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

### Aim

In bioinformatics, the inference of biological networks is one of the most active research areas. It involves decoding various complex biological networks that are responsible for performing diverse functions in human body. Among these networks analysis, most of the research focus is towards understanding effective brain connectivity and gene networks in order to cure and prevent related diseases like Alzheimer and cancer respectively. However, with recent advances in data procurement technology, such as DNA microarray analysis and fMRI that can simultaneously process a large amount of data, it yields high-dimensional data sets. These high dimensional dataset analyses possess challenges for the analyst.

### Background

Traditional methods of Granger causality inference use ordinary least-squares methods for structure estimation, which confront dimensionality issues when applied to high-dimensional data. Apart from dimensionality issues, most existing methods were designed to capture only the linear inferences from time series data.

### Method and Conclusion

In this paper, we address the issues involved in assessing Granger causality for both linear and nonlinear high-dimensional data by proposing an elegant form of the existing LASSO-based method that we call “Elastic-Net Copula Granger causality”. This method provides a more stable way to infer biological networks which has been verified using rigorous experimentation. We have compared the proposed method with the existing method and demonstrated that this new strategy outperforms the existing method on all measures: precision, false detection rate, recall, and F1 score. We have also applied both methods to real HeLa cell data and StarPlus fMRI datasets and presented a comparison of the effectiveness of both methods.

## 1 Introduction

In the modern age of bioinformatics, scientists are endeavoring to find ways to cure diseases at their source, making the recuperation process faster and more efficient. For this reason, researchers from diverse areas are striving to comprehend and replicate complex networks involved in the operation of various functions in human body. Among those networks, most of the research is focused on mapping of effective brain connectivity in the brain for specific task and gene networks in the translation of different biological reactions.

### 1.1 Brain connectivity

The brain connectivity analysis is crucial for exploring the network topology and understanding of the inter- and intra-communications involved during execution of any task as brain function does not involve isolated regions but rather requires a network of various regions to perform any task [1]. This motivated the researchers to develop the means to extract and replicate that network information.

The review by Firston [2] and others [3] divided the brain connectivity studies into three distinctive branches namely structural, functional, and effective connectivity. The Structural connectivity analysis involves the study of the anatomical links of fiber tracks that associate the neuron pools across different brain regions. Functional connectivity maps the region of the brains that are spatially distributed, but functionally connected. These functional maps are generated using statistical concepts that capture the deviation of statistical independence. However, the effective Connectivity represents an amalgamation of structural and functional connectivity showing the directional effects within a network pool.

### 1.2 Gene Networks

Gene is the basic physical and functional unit of heredity that communicates and interacts with each other to make proteins that help in performing various biological functions. Thus motivating researchers to obtain a better understanding of protein's functional interactions which provide exceptionally valuable information for discovering susceptibilities of a disease to its treatment.

In recent times, a wide range of methods for network analysis have been developed to detect the brain connectivity and gene networks that use time series data extracted from fMRI and DNA microarray. These time series data can be analyzed by utilizing a number of techniques from various fields such as econometrics. Among several techniques, Granger causality is ubiquitously used in biological network analysis (gene network analysis [4–7] and mapping of effective brain connectivity [8–12]) because of its simplicity in terms of its implementation and interpretation [13, 14]. However, its use faces limitations when dealing with high dimensional data.

The standard implementation of Granger causality as proposed in [15] was originally developed with the aim of analyzing direct or linear causality by using ordinary least-squares (OLS) methods for causality estimation. However, the use of OLS implementation limits its use for the high-dimensional biological dataset. Therefore, in this paper, we are proposing a new method based on the elastic net and copula approaches for finding the Granger causality for high-dimensional data. The proposed method will not only be able to detect direct causality but will also analyze indirect causality as well.

This paper first reviews the concept of Granger causality, its existing implementations, and their limitations in Section 2. In Section 3, we present our proposed method for addressing these limitations. Section 4 covers the experimental details and results used to evaluate the

performance of the proposed method. Finally, the discussion and conclusion are presented in Section 5.

## 2 Granger Causality

Granger causality analysis was first introduced in econometrics [15] for studying causal relationships using different financial time series data. The classical literature, such as *New Introduction to Multiple Time Series Analysis* by Lütkepohl [16], showed that Granger causality can be applied using many modeling techniques such as vector autoregressive (VAR) models, infinite order VAR, impulsive response functions and more. However, the literature in bioinformatics indicates that the VAR modeling technique is commonly used because of its simplicity and ease of interpretation [17].

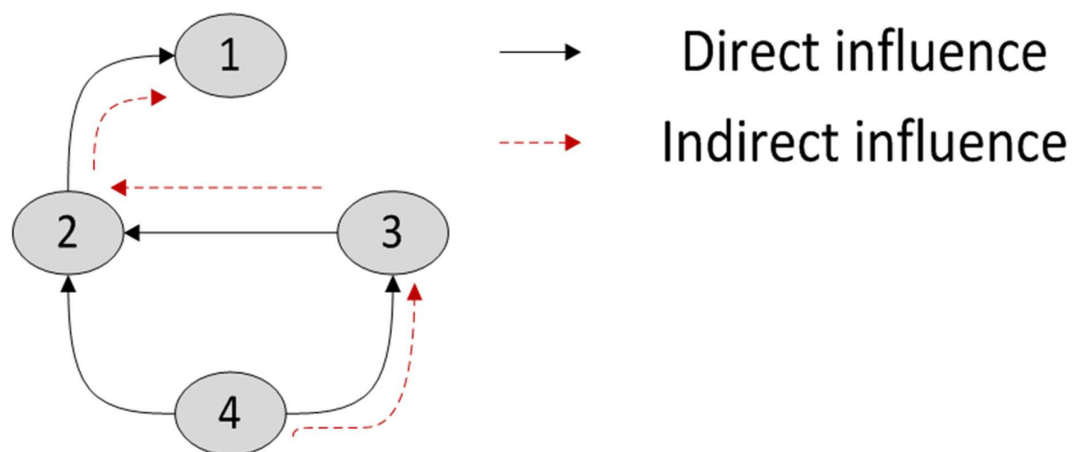
To understand the concept of Granger causality, let us consider two-time series,  $X$ , and  $Y$ . If previous values of both  $X$  and  $Y$  can be used to predict the current value of  $X$ , i.e.,  $X[n]$ , then we can conclude that  $Y$  Granger-causes  $X$ . Thus, including both variables increases the prediction's effectiveness over that of using past values of  $X$  alone [15, 16].

Granger causality can be used to explain direct or indirect influences. To understand the concept of direct (linear) influence and indirect (nonlinear) influence, consider a simple four-variable scenario as shown in Fig 1. Direct influence is represented by direct links between nodes, as exemplified by the edge between Node 2 and Node 1, whereas indirect influence can be traced using the edge that is mediated by one or more nodes. In Fig 1, an indirect influence can be observed by tracing the edge from Node 4 to Node 1, mediated by Node 3 and Node 2. If either Node 2 or Node 3 is blocked, then Node 4 will have no effect on Node 1.

Inference of biological networks using high-dimensional data extracted via various techniques confronts two critical challenges in finding the temporal causal relationships: 1) not all significant confounders in the data sets are known and 2) a large number of high-dimensional time series need to be analyzed.

The first challenge centers on the fact that most data sets do not measure all confounders, which makes surreptitious effects due to unobserved confounders unavoidable. In this circumstance, prior knowledge of unobserved confounders can be utilized to overshadow their effect.

The second challenge requires us to develop adaptive algorithms that can reveal temporal dependencies by utilizing a substantial number of time series that have few observation values.



**Fig 1. Direct and indirect influence.**

doi:10.1371/journal.pone.0165612.g001

In the past, several methods were used to handle such high-dimensional data, such as pairwise analysis [18], kernel-based algorithms [11, 19], and other regularization-based methods [20, 21]. Another viable alternative, proposed by Nelsen [22], uses the copula to discover dependencies between random variables. The use of probability theory in the form of copulas counters the spurious effect of confounders by utilizing marginal probabilities to incorporate prior information about them.

Recently, Liu, Lafferty, and Wasserman [23] introduced the Gaussian copula with nonparametric marginals, which can be used to estimate high-dimensional undirected graphs for inferring influences between high-dimensional time series. This idea was taken forward by Bahadori and Liu [24], who proposed a new method for finding Granger causality, called the Gaussian non-paranormal (GNPN) model.

They defined the GNPN model by considering a time series  $X = (X_1, \dots, X_n)$  having a GNPN distribution of  $GNPN(X, B, F)$  as long as there are functions  $\{F_j\}_{j=1}^n$  such that  $F_j(X_j)$  for  $j = 1, \dots, n$  are jointly Gaussian and can be factored according to the VAR model through coefficient  $B = \{\beta_{i,j}\}$ . To be more precise, variables  $Z_i \hat{=} F_j(X_j)$  can be factored as shown below:

$$p_z(z) = \mathcal{N}(z(1, \dots, L)) \times \prod_{j=1}^n \prod_{t=L+1}^T p_{\mathcal{N}} \left( z_j(t) \left| \sum_{ij} \beta_{i,j}^T z_i^{t,Lagged}, \sigma_j \right. \right)$$

where  $p_{\mathcal{N}}(z|\mu, \sigma)$  is the Gaussian density function having variance  $\sigma^2$  and mean  $\mu$ ,  $L$  is the maximal time delay,  $z_i^{t,Lagged} = [z_i(t-L), \dots, z_i(t-1)]$  is the history of  $z_i$  till time  $t$ , and  $\beta_{i,j} = [\beta_{i,j}(1), \dots, \beta_{i,j}(L)]$  is the vector of coefficients modeling the effect of time series  $z_i$  on the target time series.

Based on this information, the time series  $z_j$  Granger-causes  $z_i$  if at least one value in the coefficient vector  $\beta_j$  is statistically nonzero.

In order to show the convergence of Copula Granger is consistent like LASSO Granger, Bahadori and Liu [24] use the mathematical concept discussed in [23, 25] which uses the linear model:

$$y = \beta^T \mathbf{x} + \varepsilon$$

where  $\mathbf{x}$  is a  $p \times 1$  is a random vector having zero mean and unit variance,  $\beta$  is the coefficient vector and  $\varepsilon$  is noise having zero mean and unit variance.

The real  $n$  observation samples  $\tilde{x}_i$  for  $i = 1, \dots, n$ , yields the covariance that follows the following rate as suggested by [23]

$$\max_{j,k} |\tilde{S}_{jk}^n - \hat{S}_{jk}^n| = O_p \left( \sqrt{\frac{\log p \log^2 p}{n^{\frac{1}{2}}}} \right)$$

Where  $\hat{S}_{jk}^n = (X^T X)_{jk}$  and  $\tilde{S}_{jk}^n = (\tilde{X}^T \tilde{X})_{jk}$  is our estimate of covariance using the actual and noisy samples  $x_i$  and  $\tilde{x}_i$  and assuming that the matrix  $\Delta = \tilde{C} - C$  is positive semi-definite. This assumption is bound by following equation which is modified version of equation (22) of [25]:

$$\gamma^T \tilde{C} \gamma \leq \lambda \sqrt{s} \|\gamma\|_2 + \varphi_{max}(\Delta)$$

Provided  $\varphi_{max}(\Delta) \leq K_2 \max |\tilde{S}_{jk}^n - \hat{S}_{jk}^n|$  is bounded for some constant  $K_2$  and deriving the lower bound in Eq 26 using the fact that  $\varphi_{min}(\Delta) \geq 0$  yields the following equation:

$$K\vartheta_{min} \|\gamma\|_2^2 \leq \frac{\frac{\lambda}{n} \sqrt{s}}{K\vartheta_{min} + \varphi_{max}}$$

Since  $\varphi_{max}(\Delta)$  diminishes with respect to  $\vartheta_{min}(\tilde{C})$  according to results from [23] and having the incoherent design assumption [25] for lower bound of  $\vartheta_{min}(\tilde{C})$ , the proof is established following the steps in [25].

Further mathematical details and proof of the concepts are explicitly not discussed here and can be referred to in the original article [24]. However, the gist of their technique is to isolate the marginal properties of the data from their dependency structures. In order to implement this idea, they used  $\ell_1$  (LASSO) regularization to estimate dependency structures for high-dimensional data. In this way, they used the advantage of LASSO scalability for higher dimensions and at the same time used copulas to handle the nonlinearity of the data.

### 2.1 Limitations of Existing Work

In the standard LASSO estimator [26], the  $\ell_1$  penalty is used to obtain the sparse solution to the following optimization problem:

$$\beta(LASSO) = \min_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

where  $\|\beta\|_1 = \sum_{i=1}^p |\beta_i|$  is the  $\ell_1$ -norm penalty on  $\beta$  and induces sparsity in the solution, and  $\lambda \geq 0$  is a tuning parameter.

The use of the  $\ell_1$ -norm penalty helps in the simultaneous operations of regularization and shrinkage and thus makes LASSO an appealing variable selection method. However, despite these advantages, LASSO faces some limitations, as discussed by Zou [27], that makes it unstable when used for high-dimensional data and limits the variable selection before saturation when the number of variables is greater than the number of observation points.

These LASSO problems mainly arise when dealing with either highly correlated predictors, which usually results in the random selection of predictors, or when all predictors are identical, as discussed in detail in Friedman, Hastie, and Tibshirani [28].

### 3 Elastic-Net Copula Granger Causality

The instabilities of LASSO can be circumvented using an extension of LASSO called elastic net. It is robust to high correlations among predictors [27] and can select more than  $p$  variables when  $n \gg p$  (that is, the number of variables  $\gg$  the number of observation points). It uses a mixture of the  $\ell_1$  (LASSO) and  $\ell_2$  (ridge regression) penalties and can be formulated as:

$$\beta(enet) = \min_{\beta} \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$$

where  $\|\beta\|_1 = \sum_{i=1}^p |\beta_i|$ ,  $\|\beta\|_2^2 = \sum_{i=1}^p \beta_i^2$ ,  $\lambda_1$  is the tuning parameter for  $\ell_1$  (LASSO), and  $\lambda_2$  is the tuning parameter for  $\ell_2$  (ridge regression).

As we know,  $\ell_2$  (ridge regression) [29] works well with a large number of predictors that either have nonzero coefficients, or are drawn from a normal distribution and are highly correlated [28]. Therefore, its presence in elastic net helps improve variable selection, whereas the  $\ell_1$  (LASSO) penalty induces the grouping effect and stabilizes the solution paths with respect to

random sampling [30]. Thus, using the combination of both penalties should greatly improve the predictions.

Based on these findings, we are proposing a method called elastic-net copula Granger causality (ECGC) that we expect to be more precise than the existing GNPN method. In this new method, we employ elastic-net regularization instead of  $\ell_1$  (LASSO) regularization to estimate dependency structures. The use of elastic net will overcome the shortcomings of the existing LASSO methods and will also exploit the advantages of copula to handle the nonlinearity of data.

Therefore, instead of solving the optimization problem proposed by Bahadori and Liu [24], i.e.,

$$\min_{\beta} \sum_{t=L+1}^T \left\| X_i(t) - \sum_{i=1}^p \beta_{i,j}^T X_i^{t,Lagged} \right\|_2^2 + \lambda \|\beta\|_1$$

where  $\lambda$  is the tuning parameter, we are proposing to use the following optimization problem:

$$\min_{\beta} \sum_{t=L+1}^T \left\| X_i(t) - \sum_{i=1}^p \beta_{i,j}^T X_i^{t,Lagged} \right\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$$

where  $\lambda_1$  and  $\lambda_2$  are tuning parameters for  $\ell_1$  (LASSO) and  $\ell_2$  (ridge regression) penalties, respectively and are calculated using as follow:  $\lambda_1 = \alpha$  and  $\lambda_2 = (1-\alpha)/2$ . The pseudocode for implementing elastic-net copula Granger causality is summarized below:

1. Find the marginal distribution for each time series:

$$\hat{F}_n(t) = \frac{\text{number of elements in the sample} \leq t}{n}$$

2. Map the observations into copula space:

$$\hat{f}_i(X_i^t) = \hat{\mu}_i + \hat{\sigma}_i \Phi^{-1}(\hat{F}_i(X_i^t))$$

In practice, as proposed by Bahadori and Liu [17], to avoid large numbers  $\Phi^{-1}(0^+)$  and  $\Phi^{-1}(1^-)$  we use the Winsorized estimator of the distribution function:

$$\tilde{F}_j = \begin{cases} \delta_n, & \text{if } \hat{F}_j(X_j) < \delta_n \\ \hat{F}_j(X_j), & \text{if } \delta_n \leq \hat{F}_j(X_j) \leq 1 - \delta_n \\ (1 - \delta_n), & \text{if } \hat{F}_j(X_j) > 1 - \delta_n \end{cases}$$

3. Find Granger causality using elastic net and copula for different values of tuning parameters among  $\hat{F}_i(X_i^t)$ .
4. Select Granger causality based on the minimum Akaike information criteria [31].

To evaluate our proposed method, we performed extensive experiments, which are discussed in the next section.

## 4 Experimentation and Performance Evaluation

### 4.1 Experimentation

For the implementation of our proposed method, we used MATLAB together with the Sparse Learning with Efficient Projections (SLEP) toolbox [32]. SLEP is a well-known toolbox that has functions related to regularization and has been used in the past by many researchers [33, 34].

For the comparison exercise, we used the code provided in [35], which implements LASSO copula Granger causality, and the Glmnet toolbox [36] in MATLAB.

In order to remain unbiased in our comparison, instead of using self-created data sets, we used the same data that have already been used by other authors to test similar kinds of algorithms.

**4.1.1 Simulated Data.** The first set of data was used in [37] and [38]; it simulates the scenario of three variables (genes) and uses the following set of mathematical equations:

$$x_1(t) = 0.8 x_1(t - 1) - 0.5 x_1(t - 2) + 0.4 x_3(t - 1) + \varepsilon_1(t)$$

$$x_2(t) = 0.9 x_2(t - 1) - 0.8 x_3(t - 2) + \varepsilon_3(t)$$

$$x_3(t) = 0.5 x_3(t - 1) - 0.2 x_3(t - 2) + 0.5 x_2(t - 1) + \varepsilon_3(t)$$

The second set of simulated data was first used by Schelter et al. [39] and later in [18, 40, 41] and [42]. It simulates the scenario of five variables and uses the following set of equations:

$$x_1(t) = 0.6 x_1(t - 1) + 0.65 x_2(t - 2) + \varepsilon_1(t)$$

$$x_2(t) = 0.5 x_2(t - 1) - 0.3 x_2(t - 2) - 0.3 x_3(t - 4) + 0.6 x_4(t - 1) + \varepsilon_2(t)$$

$$x_3(t) = 0.8 x_3(t - 1) - 0.7 x_3(t - 2) - 0.1 x_5(t - 3) + \varepsilon_3(t)$$

$$x_4(t) = 0.5 x_4(t - 1) + 0.9 x_3(t - 2) + 0.4 x_5(t - 2) + \varepsilon_4(t)$$

$$x_5(t) = 0.7 x_5(t - 1) - 0.5 x_5(t - 2) - 0.2 x_3(t - 1) + \varepsilon_5(t)$$

We generated the desired quantities of data from these sets of equations, as is done by other authors [18, 37, 38, 40, 41] and [42], by first initializing the equations with white Gaussian noise having zero mean and unit variance; we then used the iterative process to generate the remainder of the data.

The direct and indirect influence structure of the simulated autoregressive models 1 and 2 are depicted by the directed graphs in Fig 2 and Fig 3, respectively. In both figures, a directed edge represents the Granger causality between two nodes or variables.

The third simulated dataset used in this paper is from the Dialogue for Reverse Engineering Assessments and Methods (DREAM). DREAM is one of the major platforms for genetic research, in which series of challenges are given to researchers. The aim of these challenges is to provide researchers, the means to find and implement new and existing theories in the areas of cellular network inference and quantitative model building in systems biology.

The “In Silico Network Challenge,” one of the DREAM4 challenges, has the objective of inferring gene regulation network structures via reverse engineering from the given in silico gene expression data sets using any possible means. For this paper, we used the dataset for the InSilico\_Size10 sub-challenge, entitled “In Silico Network of Size 10” [43–45].





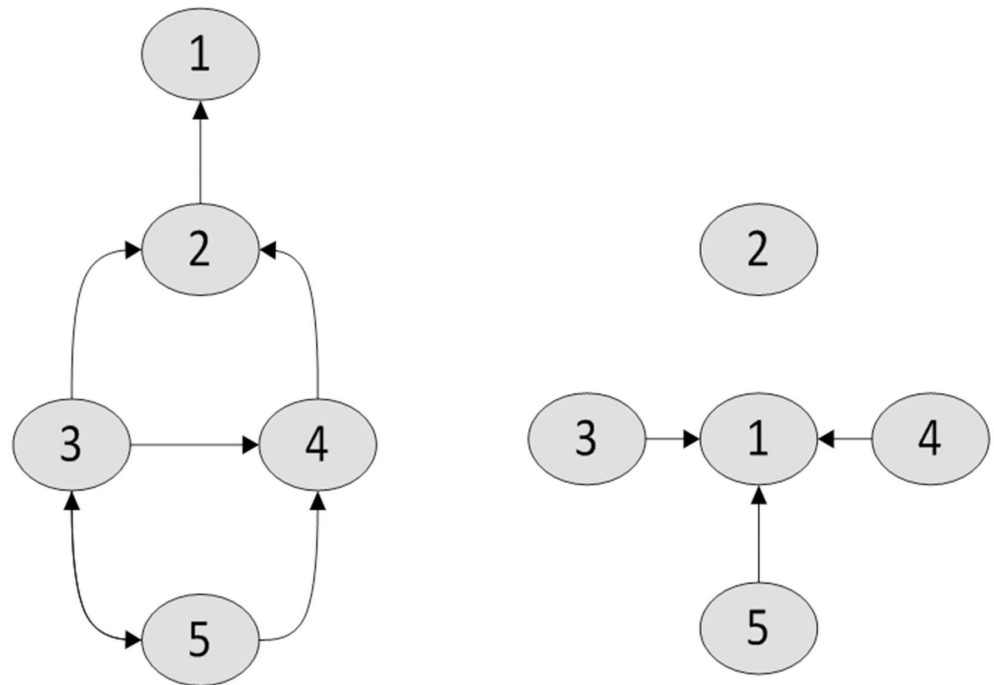
(a) Direct Influence (b) Indirect Influence

Fig 2. Influence graph for simulated data set 1.

doi:10.1371/journal.pone.0165612.g002

**4.1.2 Real Data.** HeLa Cancer Cell Dataset: The real data used in this paper is from the HeLa human cancer cell line that is collected by Whitfield et al. [46] by performing multiple experiments using DNA microarray analysis of the HeLa cell line. In this paper, we are using their experiment 3 data set, which has been used by other researchers as well [6, 47].

The experiment 3 data set identified more than 1100 genes that are periodically expressed during the cancer cell cycle. From these genes, we used 19 preselected genes that are regarded as highly influential and have been investigated by other researchers [6, 47, 48]. The 19 preselected genes that we considered are: “PCNA, NPAT, E2F1, CCNE1, CDC25A, CDKN1A, BRCA1, CCNE, CCNA2, CDC20, STK15, BUB1B, CKS2, CDC25C, PLK1, CCNB1, CDC25B, TYMS, and DHFR”.



(a) Direct Influence (b) Indirect Influence

Fig 3. Influence graph for simulated data set 2.

doi:10.1371/journal.pone.0165612.g003



As the data was not collected at homogeneous intervals, it was interpolated by cubic smoothing splines [49] before being used as advised in [6, 47].

fMRI StarPlus Dataset: The Second real dataset used in this study is called StarPlus dataset collected by [50] that can be freely accessed from [51]. This dataset had been used in past by several researchers [8, 48, 52, 53] and contained the data that was acquired to study the brain activity involved during comprehension of the relationship between sentence and pictures.

During the study, they performed series of experiments on 13 normal subjects and divided these experiments into 40 trials. In each trial, every subject had to relate a sentence with a picture and then decide the relation between sentence and picture. These 40 trials were further divided into two equal parts. In one part, they first introduced the sentence and asked to relate it to a picture whereas for the next part, they showed picture followed by a sentence.

In either setting, both stimuli were provided only for 4-second exposure and a 4-second blank screen in between. Then after the second stimulus, the subjects were asked to answer the question then rest for 15 seconds before the start of next trial. More details about experimental settings, sentences, and pictures are explicitly not discussed here and can be referred to [50].

The images that were acquired during the experiments were pre-processed using FIASCO program [54] to reduce some artifacts (signal drift and head motions) introduced during image acquisition process. Then pre-processed images were analyzed, and 25 distinct anatomical regions of interest (ROIs) were selected for further study. These regions includes: “left dorsolateral prefrontal cortex (LDLPFC) and right dorsolateral prefrontal cortex (RDLPFC), calcarine sulcus (CALC), left frontal eye fields (LFEF), right frontal eye fields (RFEF), left inferior parietal lobule (LIPL), right inferior parietal lobule (RIPL), left intraparietal sulcus (LIPS), right intraparietal sulcus (RIPS), left inferior frontal gyrus (LIFG), left opercularis (LOPER), right opercularis (ROPER), supplementary motor areas (SMA), left and right inferior temporal lobule (LIT, RIT), left and right posterior precentral sulcus (LPPREC, RPPREC), left and right supramarginal gyrus (LSGA, RSGA), left temporal lobe (LT), right temporal lobe (RT), left and right triangularis (LTRIA, RTRIA), left superior parietal lobule (LSPL) and right superior parietal lobule (RSPL)”.

## 4.2 Performance Evaluation

The performance evaluation of both methods was done by comparing our proposed method with the existing method using the following measures: precision, false detection rate (FDR), recall, and F1 score. These measures were calculated more than 5000 times for each scenario, and then their average values were used as results, to minimize impulsive errors.

Results for simulated data sets 1, and 2 are shown in Tables 1 and 2, respectively, whereas the results for the DREAM4 data are summarized in Fig 4. Moreover, the effective brain map involved in human deductive reasoning is shown in Fig 5.

**Table 1. Results for Simulated Data Set 1.**

<b>Number of Time Points</b>		<b>15</b>	<b>20</b>	<b>35</b>	<b>75</b>	<b>150</b>	<b>500</b>	<b>1000</b>
False Discovery Rate	<i>LASSO Copula GC</i>	0.52	0.59	0.7	0.76	0.76	0.81	0.8
	<i>Elastic-Net Copula GC</i>	0.41	0.45	0.44	0.35	0.45	0.53	0.52
Recall	<i>LASSO Copula GC</i>	0.41	0.41	0.41	0.41	0.41	0.4	0.41
	<i>Elastic-Net Copula GC</i>	0.51	0.57	0.69	0.84	0.79	0.86	0.91
F1 Score	<i>LASSO Copula GC</i>	0.42	0.39	0.33	0.3	0.3	0.26	0.27
	<i>Elastic-Net Copula GC</i>	0.51	0.53	0.58	0.70	0.64	0.61	0.62
Precision	<i>LASSO Copula GC</i>	0.48	0.51	0.30	0.24	0.24	0.19	0.20
	<i>Elastic-Net Copula GC</i>	0.59	0.56	0.56	0.65	0.55	0.48	0.48

doi:10.1371/journal.pone.0165612.t001

**Table 2. Results for Simulated Data Set 2.**

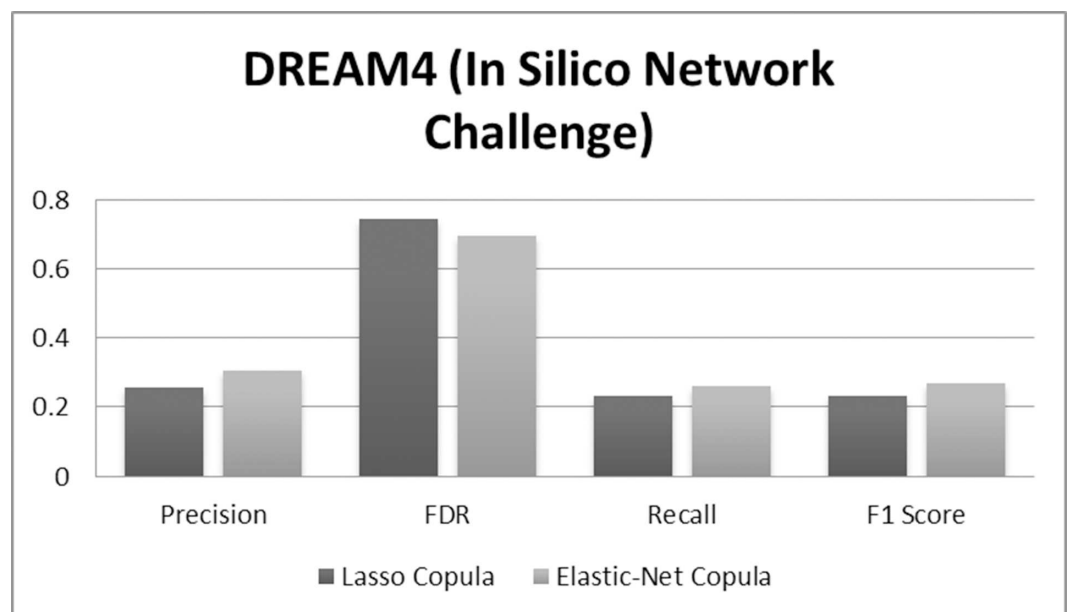
Number of Time Points		15	20	35	75	150	500	1000
False Discovery Rate	LASSO Copula GC	0.7	0.67	0.61	0.59	0.6	0.55	0.59
	Elastic-Net Copula GC	0.59	0.56	0.51	0.55	0.54	0.54	0.55
Recall	LASSO Copula GC	0.15	0.36	0.58	0.58	0.53	0.91	0.80
	Elastic-Net Copula GC	0.35	0.44	0.67	0.78	0.8	0.91	0.95
F1 Score	LASSO Copula GC	0.19	0.33	0.46	0.48	0.46	0.6	0.55
	Elastic-Net Copula GC	0.36	0.42	0.54	0.57	0.58	0.61	0.61
Precision	LASSO Copula GC	0.31	0.33	0.39	0.41	0.4	0.45	0.42
	Elastic-Net Copula GC	0.41	0.44	0.49	0.45	0.46	0.46	0.45

doi:10.1371/journal.pone.0165612.t002

For the real HeLa cell data, as there are no means to compare the effectiveness against other methods, we extracted the top 20 significant interactions. We looked up these 20 interactions in the BioGRID database [55] to see whether they had been reported earlier. Once all 20 interactions were analyzed, we compared the number of matches found by each method. This approach for evaluation has been used and suggested by other researchers [56]. The results of our proposed method and the LASSO copula method are summarized in Table 3.

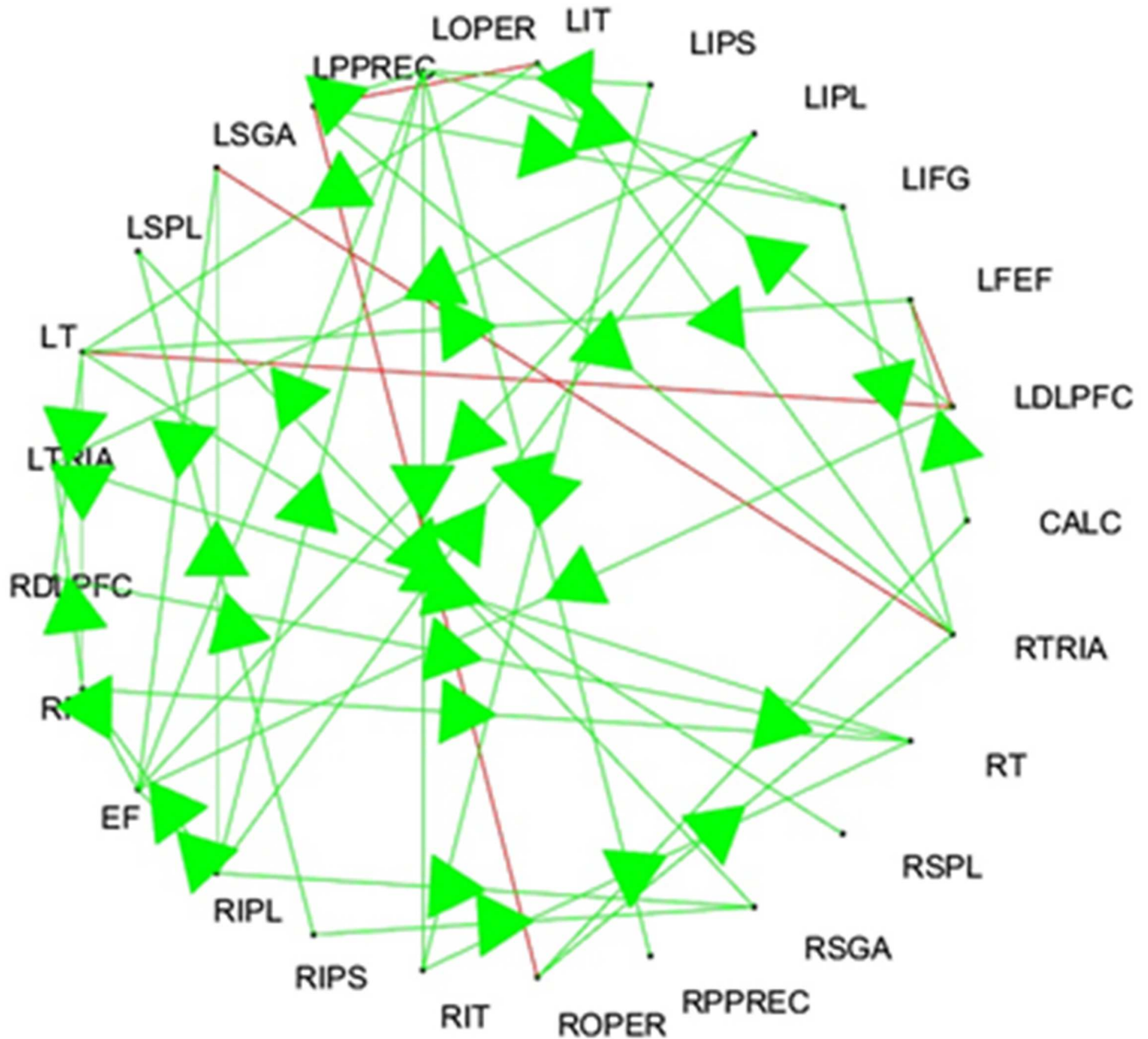
### 5 Discussion and Conclusion

For the evaluation of performance, we divided the simulated data into two groups: high-dimensional data and low-dimensional data. High-dimensional cases were defined as those having less than 100-time points, whereas low-dimensional cases are those having more than 100-time points. These definitions were determined based on the fact that techniques similar to DNA microarray analysis or fMRI, which can analyze multiple genes simultaneously, usually generate 15–75-time points because of their data procurement procedures.



**Fig 4. Results for DREAM4 In Silico Network Challenge.**

doi:10.1371/journal.pone.0165612.g004



**Fig 5. Brain Connectivity map involves in deductive reasoning.**

doi:10.1371/journal.pone.0165612.g005

Based on this division, the results for simulated data sets 1 and 2 show that for high-dimensional data, on the average, elastic-net copula has 14.8% better precision, 15.92% lower FDR, 19.18% higher recall, and a 16.33% higher F1 score. For the low-dimensional cases, the elastic-net copula has 16.5% better precision, 16% lower FDR, 29.3% higher recall, and a 20.75% higher F1 score.

Based on these findings on simulated data, we observe that our method is consistent with both low- and high-dimensional data with respect to precision, FDR, and F1 score. For recall, we believe that the drastic increase of performance for low-dimensional cases is due to the greater availability of information as more time point values are considered.

**Table 3. Top 20 Significant Gene Interactions using Elastic-Net Copula Granger Causality and LASSO Copula Granger Causality.**

<i>Elastic-Net</i>			<i>LASSO</i>		
<i>Copula Granger</i>			<i>Copula Granger</i>		
CCNB1	↔	CDC25B	CCNB1	↔	CDC25B
E2F1	↔	CCNE1	E2F1	↔	CCNE1
CCNE1	↔	CDC25A	CCNE1	↔	CDC25A
PLK1	↔	CCNB1	PLK1	↔	CCNB1
CDKN1A	↔	BRCA1	PCNA	↔	NPAT
PCNA	↔	NPAT	PCNA	↔	PCNA
CDC25A	↔	CDKN1A	CDC25A	↔	CDKN1A
PCNA	↔	PCNA	CDKN1A	↔	BRCA1
<b>CCNB1</b>	↔	<b>CCNF</b>	BRCA1	↔	CDC25B
CDC25C	↔	PLK1	CDC25C	↔	PLK1
CDC25B	↔	TYMS	CDC25B	↔	TYMS
CCNB1	↔	STK15	NPAT	↔	E2F1
BUB1B	↔	CKS2	BUB1B	↔	CKS2
NPAT	↔	E2F1	DHFR	↔	DHFR
DHFR	↔	DHFR	CDC20	↔	CDC25B
STK15	↔	BUB1B	CCNA2	↔	CDC20
CCNA2	↔	CDC20	STK15	↔	BUB1B
CKS2	↔	CDC25C	CKS2	↔	CDC25C
PCNA	↔	E2F1	NPAT	↔	NPAT
<b>CCNB1</b>	↔	<b>CKS2</b>	BUB1B	↔	CDC25B

doi:10.1371/journal.pone.0165612.t003

Similar trends for precision, FDR, recall, and F1 score can be seen from the DREAM4 data (Fig 4), where we observe improved results with the use of elastic-net copula Granger causality.

For brain connectivity map, there is no standard way to verify the resultant network other than performing some clinical trials. However, clinical trials and their results are out of scope of this paper.

However, for the results on real HeLa data (among those top 20 significant interactions listed in Table 3), we were able to detect 7 reported interactions using the elastic-net method, whereas only 5 interactions were found using the LASSO-based method. These matched interactions are shown in bold in Table 3. Of those seven interactions found by the elastic-net method, we note two interesting interactions (highlighted in Table 3) that were not detected at all by the LASSO copula method. These interactions are related to different cancer cell cycles, as reported in [57–59].

In this paper, we have proposed a new method called elastic-net copula Granger causality, which can use high-dimensional data for assessing both linear and nonlinear gene and brain networks. We have compared the performance of the new method with its predecessor, LASSO copula Granger causality. Based on the evidence from extensive experimentation, it is clear that elastic-net copula outperforms the existing LASSO copula. Moreover, when applied to real cancer cell data, it shows the capacity to detect some significant interactions that the other method is not able to detect, further reinforcing the effectiveness of our approach. Therefore, in our view, our proposed method provides a more stable regularization based technique to study gene and brain networks thus helping the researcher to manage and treat disease more meritoriously and proficiently.

## Consent and Sources of Real Data

Although this study involves human participants, formal consents or ethical committee approval is not required as experimental data used in this research is not collected by current authors and is freely accessible. HeLa cell Genetic is acquired from published article of Michael et al. and can be accessed from <http://genome-www.stanford.edu/Human-CellCycle/Hela/data.shtml>. Similarly, StarPlus fMRI data is acquired from published work of Mitchell et al. and can be accessed freely from <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>.

## Author Contributions

**Conceptualization:** MSF.

**Formal analysis:** MSF.

**Investigation:** MSF MYS.

**Methodology:** MSF MYS.

**Project administration:** MYS.

**Resources:** MSF.

**Software:** MSF.

**Supervision:** MYS.

**Validation:** MYS.

**Visualization:** MSF MYS.

**Writing – original draft:** MSF.

**Writing – review & editing:** MSF MYS.

## References

1. Heuvel MPvd, Sporns O. Rich-club organization of the human connectome. *The Journal of Neuroscience*. 2011; 31(44):15775–86. doi: [10.1523/JNEUROSCI.3539-11.2011](https://doi.org/10.1523/JNEUROSCI.3539-11.2011) PMID: [22049421](https://pubmed.ncbi.nlm.nih.gov/22049421/).
2. Friston KJ. Functional and effective connectivity: a review. *Brain Connectivity*. 2011; 1(1):13–36. doi: [10.1089/brain.2011.0008](https://doi.org/10.1089/brain.2011.0008) PMID: [22432952](https://pubmed.ncbi.nlm.nih.gov/22432952/).
3. Horwitz B. The elusive concept of brain connectivity. *NeuroImage*. 2003; 19(2):466–70. doi: [10.1016/s1053-8119\(03\)00112-5](https://doi.org/10.1016/s1053-8119(03)00112-5) PMID: [12814595](https://pubmed.ncbi.nlm.nih.gov/12814595/)
4. Tam GHF, Hung YS, Chang C, editors. Meta-Analysis on Gene Regulatory Networks Discovered by Pairwise Granger Causality. The 7th International Conference on Systems Biology (ISB) 2013 23–25 August; Huangshan, China: IEEE.
5. Michailidis G, d'Alche-Buc F. Autoregressive models for gene regulatory network inference: sparsity, stability and causality issues. *Math Biosci*. 2013; 246(2):326–34. doi: [10.1016/j.mbs.2013.10.003](https://doi.org/10.1016/j.mbs.2013.10.003) PMID: [24176667](https://pubmed.ncbi.nlm.nih.gov/24176667/).
6. Hlaváčková-Schindler K, Bouzari H, editors. Granger Lasso Causal Models in Higher Dimensions-Application to Gene Expression Regulatory Networks. ECML/PKDD 2013 Workshop Scalable Decision Making: Uncertainty, Imperfection, Deliberation (SCALE); 2013; Prague, Czech Republic.
7. Zhang ZG, Hung YS, Chan SC, Xu WC, Hu Y, editors. Modelling and Identification of gene regulatory networks: A Granger causality approach. Ninth International Conference on Machine Learning and Cybernetics; 2010 11–14 July Qingdao: IEEE.
8. Hutchinson RA, Niculescu RS, Keller TA, Rustandi I, Mitchell TM. Modeling fMRI data generated by overlapping cognitive processes with unknown onsets using Hidden Process Models. *NeuroImage*. 2009; 46(1):87–104. doi: [10.1016/j.neuroimage.2009.01.025](https://doi.org/10.1016/j.neuroimage.2009.01.025) PMID: [19457397](https://pubmed.ncbi.nlm.nih.gov/19457397/).

9. Zhenyu Z, Yun J, Tianyu T, Zuhong L, Yijun L, Yonghong C, et al., editors. Detecting Effective Connectivity in Human Brain using Granger Causality. *BioMedical Engineering and Informatics, 2008 BMEI 2008 International Conference on*; 2008 27–30 May 2008.
10. Liao W, Ding J, Marinazzo D, Xu Q, Wang Z, Yuan C, et al. Small-world directed networks in the human brain: multivariate Granger causality analysis of resting-state fMRI. *NeuroImage*. 2011; 54(4):2683–94. doi: [10.1016/j.neuroimage.2010.11.007](https://doi.org/10.1016/j.neuroimage.2010.11.007) PMID: [21073960](https://pubmed.ncbi.nlm.nih.gov/21073960/).
11. Amblard PO, Vincent R, Michel OJJ, Richard C, editors. Kernelizing Geweke's measures of granger causality. *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*; 2012 23–26 Sept. 2012.
12. Barnett L, Seth AK. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *Journal of Neuroscience Methods*. 2014; 223:50–68. doi: [10.1016/j.jneumeth.2013.10.018](https://doi.org/10.1016/j.jneumeth.2013.10.018) PMID: [24200508](https://pubmed.ncbi.nlm.nih.gov/24200508/).
13. Roy S, Das D, Choudhury D, Gohain G, Sharma R, Bhattacharyya D. Causality Inference Techniques for In-Silico Gene Regulatory Network. In: Prasath R, Kathirvalvakumar T, editors. *Mining Intelligence and Knowledge Exploration. Lecture Notes in Computer Science*. 8284: Springer International Publishing; 2013. p. 432–43.
14. Tam GHF, Chunqi C, Yeung Sam H, editors. Application of Granger causality to gene regulatory network discovery. *Systems Biology (ISB), 2012 IEEE 6th International Conference on*; 2012 18–20 Aug. 2012.
15. Granger CWJ. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*. 1969; 37(3):424–38.
16. Lütkepohl H. *New Introduction to Multiple Time Series Analysis*: Springer Berlin Heidelberg; 2005.
17. Fujita A, Sato JR, Kojima K, Gomes LR, Nagasaki M, Sogayar MC, et al. Identification of Granger Causality between Gene Sets. *Journal of Bioinformatics and Computational Biology*. 2010; 8(4):679–701.
18. Tam GHF, Chunqi C, Hung YS. Gene regulatory network discovery using pairwise Granger causality. *IET Systems Biology*. 2013; 7(5):195–204. doi: [10.1049/iet-syb.2012.0063](https://doi.org/10.1049/iet-syb.2012.0063) PMID: [24067420](https://pubmed.ncbi.nlm.nih.gov/24067420/)
19. Marinazzo D, Pellicoro M, Stramaglia S. Kernel Granger Causality and the Analysis of Dynamical Networks. *Physical Review E*. 2008; 77(5).
20. Shimamura T, Imoto S, Yamaguchi R, Fujita A, Nagasaki M, Miyano S. Recursive regularization for inferring gene networks from time-course gene expression profiles. *BMC Systems Biology*. 2009; 3(1):41.
21. Fujita A, Sato J, Garay-Malpartida H, Yamaguchi R, Miyano S, Sogayar M, et al. Modeling gene expression regulatory networks with the sparse vector autoregressive model. *BMC Systems Biology*. 2007; 1:39. doi: [10.1186/1752-0509-1-39](https://doi.org/10.1186/1752-0509-1-39) PMID: [17761000](https://pubmed.ncbi.nlm.nih.gov/17761000/)
22. Nelsen RB. *An Introduction to Copulas*. 2nd ed. New York: Springer-Verlag; 2006.
23. Liu H, Lafferty J, Wasserman L. The Nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. *Journal of Machine Learning Research*. 2009; 10:2295–328.
24. Bahadori MT, Liu Y, editors. *An Examination of Practical Granger Causality Inference*. 2013 SIAM International Conference on Data Mining; 2013; Austin, Texas, USA.
25. Meinshausen N, Yu B. Lasso-type recovery of sparse representations for high-dimensional data. 2009:246–70. doi: [10.1214/07-AOS582](https://doi.org/10.1214/07-AOS582)
26. Tibshirani R. Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society, Series B (Methodological)*. 1996; 58(1):267–88.
27. Zou H. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*. 2006; 101(476):1418–29. doi: [10.1198/016214506000000735](https://doi.org/10.1198/016214506000000735)
28. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*. 2010; 33(1):1–22. PMID: [20808728](https://pubmed.ncbi.nlm.nih.gov/20808728/)
29. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*. 1970; 12(1):55–67.
30. Ogutu JO, Schulz-Streeck T, Piepho HP. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proc*. 2012; 6 Suppl 2:S10. doi: [10.1186/1753-6561-6-S2-S10](https://doi.org/10.1186/1753-6561-6-S2-S10) PMID: [22640436](https://pubmed.ncbi.nlm.nih.gov/22640436/); PubMed Central PMCID: [PMC3363152](https://pubmed.ncbi.nlm.nih.gov/PMC3363152/).
31. Akaike H. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*. 1974; 19(6):716–23. doi: [10.1109/TAC.1974.1100705](https://doi.org/10.1109/TAC.1974.1100705)
32. Liu J, Ji S, Ye J. *SLEP: Sparse Learning with Efficient Projections*. 4.1 ed: Arizona State University; 2009.
33. Suk H-I, Wee C-Y, Shen D. Discriminative Group Sparse Representation for Mild Cognitive Impairment Classification. In: Wu G, Zhang D, Shen D, Yan P, Suzuki K, Wang F, editors. *Machine Learning in*



- Medical Imaging. Lecture Notes in Computer Science. 8184: Springer International Publishing; 2013. p. 131–8.
34. Liu J, Xu Y, Cheng J, Zhang Z, Wong D, Yin F, et al. Multiple Modality Fusion for Glaucoma Diagnosis. In: Zhang Y-T, editor. The International Conference on Health Informatics. IFMBE Proceedings. 42: Springer International Publishing; 2014. p. 5–8.
  35. Liu Y, Bahadori MT, Li H. Copula-Granger 2012. Available: [www-scf.usc.edu/~mohammab/codes/codes.html](http://www-scf.usc.edu/~mohammab/codes/codes.html).
  36. Qian J, Hastie T, Friedman J, Tibshirani R, Simon N. Glnet for Matlab 2013. Available: [http://www.stanford.edu/~hastie/glnet\\_matlab/](http://www.stanford.edu/~hastie/glnet_matlab/).
  37. Ding M, Chen Y, Bressler SL. Granger Causality: Basic Theory and Application to Neuroscience. Handbook of Time Series Analysis: Wiley; 2006. p. 438–60.
  38. Siyal MY, Furqan MS, Monir SMG, editors. Granger Causality: Comparative Analysis of Implementations for Gene Regulatory Networks 13th International Conference on Control, Automation, Robotics and Vision; 2014; Singapore: IEEE.
  39. Schelter B, Winterhalder M, Eichler M, Peifer M, Hellwig B, Guschlbauer B, et al. Testing for directed influences among neural signals using partial directed coherence. *Journal of Neuroscience Methods*. 2006; 152(1–2):210–9. doi: [10.1016/j.jneumeth.2005.09.001](https://doi.org/10.1016/j.jneumeth.2005.09.001) PMID: [16269188](https://pubmed.ncbi.nlm.nih.gov/16269188/).
  40. Furqan MS, Siyal MY. Random forest Granger causality for detection of effective brain connectivity using high-dimensional data. *Journal of Integrative Neuroscience*. 2015; 15(01):55–66. doi: [10.1142/S0219635216500035](https://doi.org/10.1142/S0219635216500035) PMID: [26620192](https://pubmed.ncbi.nlm.nih.gov/26620192/)
  41. Baccala LA, Sameshima K. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*. 2001; 84:463–74. doi: [10.1007/PL00007990](https://doi.org/10.1007/PL00007990) PMID: [11417058](https://pubmed.ncbi.nlm.nih.gov/11417058/)
  42. Seth AK. A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*. 2010; 186(2):262–73. doi: [10.1016/j.jneumeth.2009.11.020](https://doi.org/10.1016/j.jneumeth.2009.11.020) PMID: [19961876](https://pubmed.ncbi.nlm.nih.gov/19961876/).
  43. Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G. Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences*. 2010; 107(14):6286–91. doi: [10.1073/pnas.0913357107](https://doi.org/10.1073/pnas.0913357107) PMID: [20308593](https://pubmed.ncbi.nlm.nih.gov/20308593/)
  44. Marbach D, Schaffter T, Mattiussi C, Floreano D. Generating Realistic In Silico Gene Networks for Performance Assessment of Reverse Engineering Methods. *Journal of Computational Biology*. 2009; 16(2):229–39. doi: [10.1089/cmb.2008.09TT](https://doi.org/10.1089/cmb.2008.09TT) PMID: [19183003](https://pubmed.ncbi.nlm.nih.gov/19183003/)
  45. Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, Xue X, et al. Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges. *PLoS ONE*. 2010; 5(2):e9202. doi: [10.1371/journal.pone.0009202](https://doi.org/10.1371/journal.pone.0009202) PMID: [20186320](https://pubmed.ncbi.nlm.nih.gov/20186320/)
  46. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, et al. Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. *Molecular Biology of the Cell*. 2002; 13(6):1977–2000. doi: [10.1091/mbc.02-02-0030](https://doi.org/10.1091/mbc.02-02-0030) PMC117619. PMID: [12058064](https://pubmed.ncbi.nlm.nih.gov/12058064/)
  47. Lozano AC, Abe N, Liu Y, Rosset S. Grouped graphical Granger modeling for gene expression regulatory networks discovery. *Bioinformatics*. 2009; 25(12):i110–8. doi: [10.1093/bioinformatics/btp199](https://doi.org/10.1093/bioinformatics/btp199) PMID: [19477976](https://pubmed.ncbi.nlm.nih.gov/19477976/); PubMed Central PMCID: [PMC2687953](https://pubmed.ncbi.nlm.nih.gov/PMC2687953/).
  48. Furqan MS, Siyal MY. Inference of biological networks using Bi-directional Random Forest Granger causality. *SpringerPlus*. 2016; 5(1):1–10. doi: [10.1186/s40064-016-2156-y](https://doi.org/10.1186/s40064-016-2156-y) PMID: [27186478](https://pubmed.ncbi.nlm.nih.gov/27186478/)
  49. Green PJ, Silverman BW. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. New York: Chapman and Hall; 1994.
  50. Keller TA, Just MA, Stenger VA. Reading span and the time-course of cortical activation in sentence-picture verification. Annual Convention of the Psychonomic Society; Orlando, FL2001.
  51. Mitchell T, Wang W. StarPlus fMRI data Carnegie Mellon University's CCBI2001. Available: <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>.
  52. Mitchell T, Hutchinson R, Niculescu RS, Pereira F, Wang X. Learning to Decode Cognitive States from Brain Images. *Machine Learning*. 2004; 57:145–75.
  53. Wang X, Mitchell T. Detecting cognitive states using machine learning. Iterim working paper. 2002.
  54. Eddy WF, Fitzgerald M, Genovese C, Lazar N, AudrisMockus, Welling J. The Challenge of Functional Magnetic Resonance Imaging. *Journal of Computational and Graphical Statistics*. 1999; 8(3):545–58.
  55. Chatr-aryamontri A, Breitkreutz B-J, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. *Nucleic Acids Research*2014.
  56. Cheng D, Bahadori MT, Liu Y. FBLG: a simple and effective approach for temporal dependence discovery from time series data. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining; New York, New York, USA. 2623709: ACM; 2014. p. 382–91.



57. Wolthuis R, Clay-Farrace L, van Zon W, Yekezare M, Koop L, Ogink J, et al. Cdc20 and Cks Direct the Spindle Checkpoint-Independent Destruction of Cyclin A. *Molecular Cell*. 2008; 30(3):290–302. doi: [10.1016/j.molcel.2008.02.027](https://doi.org/10.1016/j.molcel.2008.02.027) PMID: [18471975](https://pubmed.ncbi.nlm.nih.gov/18471975/)
58. Kong M, Barnes EA, Ollendorff V, Donoghue DJ. Cyclin F regulates the nuclear localization of cyclin B1 through a cyclin–cyclin interaction. *The EMBO Journal*. 2000; 19(6):1378–88. doi: [10.1093/emboj/19.6.1378](https://doi.org/10.1093/emboj/19.6.1378) PMID: [PMC305678](https://pubmed.ncbi.nlm.nih.gov/PMC305678/).
59. Radulovic M, Crane E, Crawford M, Godovac-Zimmermann J, Yu VPCC. CKS Proteins Protect Mitochondrial Genome Integrity by Interacting with Mitochondrial Single-stranded DNA-binding Protein. *Molecular & Cellular Proteomics*. 2010; 9(1):145–52. doi: [10.1074/mcp.M900078-MCP200](https://doi.org/10.1074/mcp.M900078-MCP200) PMID: [19786724](https://pubmed.ncbi.nlm.nih.gov/19786724/)