JOURNAL OF TEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015

PALLAS: Penalized mAximum LikeLihood and pArticle Swarms for Inference of Gene Regulatory Networks from Time Series Data

Yukun Tan, Fernando B. Lima Neto, and Ulisses Braga Neto

Abstract—We present PALLAS, a practical method for gene regulatory network (GRN) inference from time series data, whic employs penalized maximum likelihood and particle swarms for optimization. PALLAS is based on the Partially-Observable I Dynamical System (POBDS) model and thus does not require ad-hoc binarization of the data. The penalty in the likelihood is LASSO regularization term, which encourages the resulting network to be sparse. PALLAS is able to scale to networks of re under no prior knowledge, by virtue of a novel continuous-discrete Fish School Search particle swarm algorithm for efficien simultaneous maximization of the penalized likelihood over the discrete space of networks and the continuous space of obparameters. The performance of PALLAS is demonstrated by a comprehensive set of experiments using synthetic data gene from real and artificial networks, as well as real time series microarray and RNA-seq data, where it is compared to several well-known methods for gene regulatory network inference. The results show that PALLAS can infer GRNs more accurately other methods, while being capable of working directly on gene expression data, without need of ad-hoc binarization. PALL fully-fledged program, written in python, and available on GitHub (https://github.com/yukuntan92/PALLAS).

| Index Terms—Gene regulatory network, partially-observable | e Boolean dynamical system | , penalized maximum lik | elihood, par |
|---|----------------------------|-------------------------|--------------|
| swarms, PALLAS. | | | |

INTRODUCTION

NFERENCE of gene regulatory networks (GRN) fromogrameA-seq datas modeled by the observation process, expression time-series data is a probl**eritical**m- while the Boolean states are hidden. This allows the optir portance in Bioinformatic Maly mathematical models ference of the sequence of Boolean states, as well as sy have been proposed in the literature to address this auauhiete;s, from the time series data. including lineamodels[2], [3], Bayesian networks], In this paperwe present PALLAS practicate thod [5],neuralnetworks [6],ifferential quations [3][7] and for parametric gene network inference based on the POBI information theory based approach [2] [B] Booleanmodel using penalized maximum likelihood and particle network (BN) model 0] is an effective model GRNs swarms for optimizatiBALLAS is a sophisticated statedue to its ability to describe temporatterns ofene space method that n detectionality and acactivation and inactivation and its comparatively stimal tibat/anhibition status; thoutany prior knowledge, requirement for inference [12][13][14][15]. Several in addition to being capable working directly on gene extensions of the BN model have been proposed, inexpediaseion dataithout the need for ad-hoc binarization. Random Boolean Networks [86]plean Networks with penalty in the likelihood score/is-aorm LASSO perturbation (BNp)16], and Probabilist Boolean Net-regularization term [24]hich encourages the resulting works (PBN[17],and Boolean ContrNetworks (BCN) network to be sparse, i.e., contain a small number of edge [18],[19].Howeverall of those models assume that between geness value can be adjusted by the user to system Boolean states are completely obscilivishise. obtain a desired level sparsityThe likelihood itself a significant drawbasince appracticanethods for thecalculated efficiently by an auxiliary particle filter (APF) inference of Boolean networks must include a stepi**nfplei**mentation (the Boolean Kalman Filt[20],[22]. hoc binarization of the gene expression data. The Plantitally movel feature of PALLAS is the application to observed Boolean dynamisyatem (POBDS)nodel[20] Boolean models of a particle swarm method mixed addresses this problem in a principlebly ypage tulating continuous-discrete version of the Fish School Search alg separate Boolean state and geotestalvation processetthm [23][24] for efficient simultaneous maximization of The time-series gene expressionwhatther microarrathe penalized likelihood over the discrete space of network

- ing, Texas A&M University, College Station, TX, 77843, USA.
- Fernando B.Lima Neto is with Department of Computer Engineering, Polytechnic University of Pernambuco, Recife, 50720, Brazil.
- Ulisses Braga Neto is with the Department o**E**lectrical and Computer Engineering, Texas A&M University, College Station, TX, 77843, USA. (RNN) in [26] and S-systems in [27]. E-mail: ulisses@ece.tamu.edu (corresponding author.)

Manuscript received XXXX; revised XXXX.

and the continuous space of observational parameters. Yukun Tan is with the Department of Electrical and Computer Enginee@arly version of this work appeared previously in a short communication [25]. We mention that particle swarm met ods have been employed before for GRN inference using non-Boolean modelsamely, recursive neuraletworks

PALLAS is an extension of the adaptive filtering method proposed in [22] he latter performs maximization of the

likelihood function by exhaustive search over the siscessofmed here to be ectoral with a smallrobability networks and expectation maximization over the spake of independently of the others. The user can select parameters of the observationdelfor each candidate fixed value for or allow the algorithm to treats network is well suited if there is prior knowledge about ameter to be estimated from the tast in the the network,g.,most of the edges are known and or symplementary Materiadicate that here is not a big few putative edges are being s**quiytal**, the prohibitive difference between using a small fixe d ₹âlû€ and computationabst of exhaustively searching the spacestimating in an interval 0.01, 0.1] for the p53-MDM2 all networks shown in Section 2i8, the absence of Negative-Feedback Loop Gene Regulatory Network. prior knowledge abothte gene interactions, number We assume a specific mddelthe network function of networks is given3by× 2^d, where is the number of that is motivated by gene pathway diagrams commonly en genes. With only 4 genes, there are a total, of 7,536 countered in biomedical research. Let a sample state vec Boolean network models to be seamble with d=10 $\mathbf{x} \in \{0, 1\}^{-d}$ and the network functible expressed in genes this number is larger than, rendering exhaus component form $\mathbf{x} \in (x_1, \dots, \delta)$ and $\mathbf{f} = (f_1, \dots, \delta)$, tive search completely unfeasible. PALLAS differs **trespentively.** Each completely unfeasible. PALLAS differs trespentively. Each completely unfeasible. PALLAS differs trespentively. method in [22]n using penalized maximum likelihods/ and particle swarms for optimizatidnich allows ito handle networks of realistic size in the absence of any prior $f_i(\mathbf{x}) = \begin{pmatrix} 1 & P_{j=1}^d & a_{ij} \times_j + b_i > 0 \\ 0 & \text{otherwise,} \end{pmatrix}$ method in [22]n using penalized maximum likelihoda/ (2) knowledge.

The performance of PALLAS is demonstrated by Wherm = +1 if there is positive regulation (activation) prehensive set experiments sing synthetic data generom generol to generol; $a_{ij} = -1$ if there is negative erated from both reahd artificia GRNs, which allows regulation (inhibition) from gettegene; and $a_{ij} = 0$ computation of performance metrics, we compare FAGENRS is not an input to gene, whereas = +1/2 to regression-based metheds,GENIE3 [28][29],TI- if gene is positively biased in the sense that qual GRESS [30],29];Bayesian Networks metheds,Banjo number ofactivation and inhibition inputs wilduce [31]and Boolean network methægsβest-Fit algorithnactivationthe reverse being the case if −1/2 . The [32],REVEAL [33]GABNI [34],and FBNNe [35].Using network model is depicted in Figure 1, where the threshol real time series microarray data from the SOS DNAURelsaare step functions that output 1 if the input is positive System in Ecoli,we compare PALLAS to the method8114 0, otherwise. This model constraint reduces the number [36], [37], [27]. We also illustrate the performance of parameters needed to specify 2^a to $a^a + b$. in recovering known regulatory links in the E. Coli Biofilm

 x_1

 x_2

:

 a_{1d}

 a_{d1}

Formation Pathway using time series RNA-Seq data.

2 **METHODS**

2.1 Partially-Observable Boolean Dynamical Systems

The Partially-Observab Boolean dynamical system (POBDS) model [20] allows for uncertainty in Boolean s transitionand partialobservation of the Boolean state variables through noise.

2.1.1 State model

Consider a state process; $k = 0, 1, ... \}$, where $K_k \in$ {0, 1}^a is a Boolean vector of, size ch evolves according to

the process;
$$k = 0, 1, ...\}$$
, where $\mathbf{X}_k \in \mathbf{X}_k \in \mathbf{X}_k$ blean vector of, size the evolves according $\mathbf{X}_k = \mathbf{f}\left(\mathbf{X}_{k-1}\right) \oplus \mathbf{n}_k$, (1) Fig. 1: Schematic representation of the network function

for k = 1, 2, ... where $\{0, 1\}^d \to \{0, 1\}^d$ is called the network function $\mathbf{n}_k \in \{0, 1\}^d$ is additive noise at timed 2.1.2 Observation Model

"" indicates component-wise modulo-2 addition. The state the sequence states is observed indirectly through the and noise processes are assumed to be indep**Ehdent** process k; k = 0, 1, ..., where the measurement vector state model (1) can be suitably modified to include external nonlinear function of the state and observables. inputs, if desired.

The noise random vectormodels uncertainty in the $\mathbf{Y}_k = \mathbf{h} (\mathbf{X}_k, \mathbf{v}_k)$ (3)state transition: if a compoments of, the corresponding component $\mathbf{b}(\mathbf{X}_{k-1})$ is flipped. As long as all components = 1, 2, ..., where the noise vector assumed to of n_k have a nonzero probability of being 1, the state n_k and n_k are transition noise. is an ergodic Markov Chain, with a steady state dispribationable describe nextee two observationabdels But if the noise is too intense, i.e., the probability of drisidered in this paper, responding to two common is too large, state evolution becomes chaotic. Howgene, etiasression modalities: RNA-Seq count data or micro well known that important biological pathways arertightly prescence deltaservational models for other data regulated. Accordingly, each component of the noiseodedities can be introduced, if desired.

RNA-Seq observation modeNA-Seq data can be mod-2 PALLAS Algorithm eled with the Poisson distribution [38]the negative In this sectionye describe in detSALLAS (Penalized binomiablistribution [39]40].Here,we employ the lat-mAximum LikeLihood and pArticle Swarms), an algorithm distributionWe assume that the transcript cVunts (Y_{k1}, \dots, k_d) are related to the s**X**ate (X_{k1}, \dots, k_d)

$$P(\mathbf{Y}_k = \mathbf{y} \mid \mathbf{X}_k = \mathbf{x}) = \sum_{i=1}^{\mathbf{Y}^d} P(Y_{ki} = y_i \mid X_{ki} = x_i),$$
 (4)

and adopt the negative binomial model for each countprior knowledge is available tonscatelparameters.

$$(Y_{ki} = y_i | X_{ki} = x_i) = \frac{\Gamma(y_i + \varphi_i)}{y_i! \Gamma(\varphi_i)} \frac{\lambda_i}{\lambda_i + \varphi_i} \frac{y_i}{\lambda_i + \varphi_i} \frac{\varphi_i}{\lambda_i + \varphi_i},$$
 (5)

where denotes the Gamma function φ_i , $\lambda_i > 0$ are countof transcript, respectively $qr^{-1} = 1, \ldots, d$. The inverse dispersion paramétespecifies the amount is present. We model the parameteog-space as:

$$\log \lambda_i = \log s + \mu_i + \delta_i X_i, \qquad (6)$$

depends on the instrument,≥ 0 is the baseline level increases extremely weight the number of enest. For of expression in the inactivated transcri**stiateal**nd $\delta > 0$ is the difference between read count aggreence the other hand contains the observational parameters: from the inactivated=(0) to the activated=(1) state, the baseline expression levels and the differential

Microarray observation mod Alreasonable mod 60 r ear model:

$$\mathbf{y} = \boldsymbol{\mu} + D \mathbf{x} + \mathbf{v} \,, \tag{7}$$

sion levels corresponding to the "zero" or inactive state for nality θ_{fnt} is Q = 3d in both cases. each gene, = $\operatorname{diag}\{\delta_{1},\ldots,\delta\}>0$ is a diagonal matrix. The mixed discrete-continuous fish section prostill satisfied here.

2.1.3 Boolean Kalman Filter

Given a time series of observations $\{Y_1, \ldots, Y\}$, the Boolean Kalman Filter [20]scribed in detail the mean-square error state estimator:

$$\hat{\mathbf{X}}_{k}^{\text{MS}} = \underset{\hat{\mathbf{X}}_{k} \in \{0,1\}}{\text{argmin}} E[||\hat{\mathbf{X}}_{k} - \mathbf{X}_{k}||^{2} ||\mathbf{Y}_{1:k}|].$$
 (8)

The BKF also computes the probabilities needed to deter logs prior to computing the needed to deter mine the likelihood function, detailed in Section 2.2 mean values. Then the following intervals are assumed: when the network is large, wever, computation the $\mu_i \in [\min, \max]$, BKF is intractable since each transition matrix contain elements which requires large computation and memory $[\min\{\max - \max, \max - \min\}/3, \max - \min]$, In this case, pproximate methods must be usual, as $\sigma_i \in [0.1]$, max{max - mean, mean - min}/3], the Sequential Monte Carlo (SMC) metisod (nown as $\varphi_i \in [0.5, 7]$, particle filter [4H]ere we use the auxiliary particle filter (9) implementation the Boolean Kalman Filter (APF-BKh); $i=1,\ldots,d$. Some ofthe parameters an often be described in [42] (please see that reference for theadstaile) to be the same across differently hich re-

ter,since it is able to address overdispersion in the count for inference oboolean gene regulatory networks from noisy time series géne expression dathe algorithm has two main components efficient omputation of penalized log-likelihood cost fun@iomaximization of

> the previous costinction using a novelarticle swarm methodnamelya mixed discrete-continuous fish school search procedure. We describe here the general case, wh

The algorithm can be easily modified to allow some of the parameters to be specified by they semply reducing (5) the size of the parameter space and using the known para eter values in the likelihood computation.

Let θ = $(\theta_{\rm disc}, \theta_{\rm cont}) \in \Theta$, with $\theta_{\rm disc} \in \Theta_{\rm disc}$ and $\theta_{\rm cont} \in \Theta_{\rm disc}$ the real-valued inverse dispersion parameter and mean read. rametersyhere Θ , Θ_{disc} and Θ_{cont} are the corresponding parameter spaces, with $\Theta_{\rm disc} \times \Theta_{\rm cont}$. Here $\theta_{
m disc}$ conobservation noise: the larger it is, the less observation noise the network function in (2), namely is present. We model the parameters as: and the regulation bias parameters 1/2, 1/2 , for **(6)** i = 1, ..., d. Hence $\mathcal{P}_{disc} = \{-1, 0, 1\}^{d^2} \times \{-1/2, 1/2\}^{d}$

This is a finite spake, its cardinal $|\mathbf{G}_{disc}| = 3^{d^2} \times 2^d$ exampleor a network with only= 4 genes $\Theta_{
m disc}$ = 688747536 while ifd = 8 , then $|\Theta_{
m disc}| \approx 8.8 imes 10^{-32}$. On

expression levels> 0, for i = 1, ..., d for both RNA-Seq and microarray data, the inverse dispersion parameter continuous microarray fluorescence data is a Gaussian line \vec{q} for RNA-Seq data, and the standard deviation $\mathbf{S}_i > 0$, for i = 1, ..., d, for microarray data (7) (the sequencing depth parameterssumed known for

where $=(\mu_1,\dots,\mu)\geq 0$ is the vector of baseline expression RNA-seq assay, so it is not θ part. Hence, the

containing differential expression values for each general amployed by PALLAS assumes that the parameter ${f v}\sim N\left(0,\,\Sigma\right)$ is an uncorrelated zero-mean Gaussian space is a closed and bounded region with an absorbing vector, where $\Sigma = \text{diag}\{\sigma_1^2, \dots, \theta\} > 0$. Notice that (4) is decision boundary (if the current best estimate exceeds the state of the current best estimated by the current best estimated by the state of the current best estimated by the current by the current best est boundary, it remains at the boundary). This is not a limitin requirement practices ince sensible lowerd upper boundscan be sefor all the observationparameters. These intervals can be set by the user, or the following da driven procedure to obtain defautervals is employed. Supplementary Materize) mputes exactly the minimum, max, and mean be respectively the minimum, maximumand mean value o**f**he observed data for all genes across alme points and available time selies. the case of RNA-Seq dtta,data must be normalized by dividing the measurements by the sequencing depth and

duces the data requirencenthe estimation problem. The maximum-likelihood estimator of parameteme the simplest case, there are single parameters or k is then given by φ for all genes, so that 3. In any case, is a closed and bounded rectangular region in

 $\hat{\theta}_k^{\text{ML}} = \arg \max_{k} L_k(\theta)$.

Nextwe describe the two main steps comprising the fish-school search method.

PALLAS algorithm computation of the penalized log. A state estimat $\hat{\mathbf{X}}_k^{\mathrm{ML}} = \hat{\mathbf{X}}_k(\hat{\boldsymbol{\theta}}_k^{\mathrm{ML}})$ can be obtained for the computation of the penalized log. likelihood function and the novel mixed discrete-collective here $\hat{\mathbf{X}}_k(\theta)$ denotes the optimate estimator produced by a BKF tuned to the parameter

2.2.1 Penalized Maximum-Likelihood Computation

serie $\mathbf{Y}_{1:k}^{I} = \{\mathbf{Y}_{1}^{I}, \ldots, \mathbf{Y}\}$ up to time, for $j = 1, \ldots, n$. The penalized log-likelihood of madelme is defined searchcalled the mixed fish scheelarch (MFSS)go-

$$L_{k}(\theta) = \frac{1}{kn} \log p_{\theta}(\mathbf{Y}_{1:k}^{(1)}, \dots, \mathbf{Y}_{:k}^{(p)}) - \eta \sum_{i,j=1}^{k^{d}} |a_{ij}|$$

$$= \frac{1}{kn} \sum_{j=1}^{k^{d}} \log p_{\theta}(\mathbf{Y}_{1:k}^{j}) - \eta \sum_{i,j=1}^{k^{d}} |a_{ij}|,$$
(10)

where $\eta > 0$ is a regularization parametric has a default value of 0.01 in our implementation ce, the penalized log-likelihood in (10) is the sum of the average contain a smallumber of edges between gewleis.h is **Notice that**

$$\log p_{\theta}(\mathbf{Y}_{1:k}^{j}) = \log \frac{h}{p_{\theta}(\mathbf{Y}_{k}^{j} \mid \mathbf{Y}_{1:k-1}^{j})p_{\theta}(\mathbf{Y}_{k-1}^{j} \mid \mathbf{Y}_{1:k-2}^{j})}$$

$$p(\mathbf{Y}_{2}^{j} \mid \mathbf{Y}_{1}^{j})p(\mathbf{Y}_{1}^{j})$$

$$= \frac{\chi^{k}}{m=1} \log p_{\theta}(\mathbf{Y}_{m}^{j} \mid \mathbf{Y}_{1:m-1}^{j}),$$
(11)

where

$$\begin{split} & \mathcal{P}_{\theta}(\mathbf{Y}_{m}^{j} \mid \mathbf{Y}_{1:m-1}^{j}) \\ & = \underbrace{\overset{\boldsymbol{\chi}^{d}}{\boldsymbol{\chi}^{d}}}_{i=1} \mathcal{P}_{\theta}(\mathbf{Y}_{m}^{j} \mid \mathbf{X}_{m} = \mathbf{x}^{i}, \mathbf{Y}_{1:m-1}^{j}) \mathcal{P}_{\theta}(\mathbf{X}_{m} = \mathbf{x}^{i} \mid \mathbf{Y}_{1:m-1}^{j}) \\ & = \underbrace{\overset{\boldsymbol{\chi}^{d}}{\boldsymbol{\chi}^{d}}}_{i=1} \mathcal{P}_{\theta}(\mathbf{Y}_{m}^{j} \mid \mathbf{X}_{m} = \mathbf{x}^{i}) \mathcal{P}_{\theta}(\mathbf{X}_{m} = \mathbf{x}^{i} \mid \mathbf{Y}_{1:m-1}^{j}) \end{split}$$

With $(\beta_m^{\theta,j})_i = p_{\theta}(\mathbf{Y}_m^j \mid \mathbf{X}_m = \mathbf{x}^i) P_{\theta}(\mathbf{X}_m = \mathbf{x}^i \mid \mathbf{Y}_{1:m-1}^j)$ the penalized log-likelihood in (10) be written as

$$L_{k}(\theta) = \frac{1}{kn} \sum_{j=1}^{X^{n}} X^{k} \left\| \beta_{m}^{\theta,j} \right\|_{1} - \eta \sum_{i,j=1}^{d} |a_{ij}|.$$
 (13)

The sequence of alues $\|\beta_m^{\theta j}\|_1$, for $j=1,\ldots,n$ and function Let $\Delta\theta_{\text{ind}}^s(r)=(\Delta\theta_{\text{disc,ind}}^s(r),\Delta\theta_{\text{cont,ind}}^s(r))$ be m = 1, ..., k can be computed by a BKF tuned to partim (reandidate) dividual displacement of fish θ applied to the time serles, (see the Supplementary iteration Vector $\theta_{
m disc,ind}^s$ (r) is drawn from a uniform Material for a description of the BKF). As mentioned is tribetion over the rectangular region d^2+d , while previous section, here we use the auxiliary particle fiftering) is drawn from a uniform distribution over the implementation the BKF for computation efficiency rectangular region $_1(r)$, $\tau_1(r)$] × × $[-\tau_Q(r)$, $\tau_Q(r)$].

2.2.2 Mixed Fish School Search Algorithm

Suppose that the sample data considerendent timen this section, we describe in detail a novel particle-swar optimization algorithm for discrete-continuous parameter rithm,which is an extension of the fish school search (FSS algorithm for continuous parameter spaces proposed in [2 One of the main novelties in the MFSS algorithm is the abi ity to operate on large continuous and discrete parameter (10) spaces simultaneously, which is needed to infer the conti ous noise parameters of the observation process, in addit to the discrete parameters of the GRN itself. As the origin algorithm, MFSS has a few properties that are unique amo most particle swarm optimization techniamesythe ability to switch automatically between exploration and log-likelihood per time series and a negative value time MFSS algorithine objective is to find a model number of edges in the mode in the encourages the model to both fit the data and be sparse, i.e., previous sectio**li**ach candidate mod**ist**,,each candiin agreement with biological knowledge. The calculate parameter vector $(\theta_{\text{disc}}, \theta_{\text{cont}})$, corresponds to a be adjusted by the user to obtain a desired level of sparsity. The length of denoted by . From the previous section $\neq d^{-2} + d + Q$. The fish school is an ensemble ofsuch particles in the parameter space $\Theta = \Theta_{\text{disc}} \times \Theta_{\text{cont}}$. The position of fish iteration ill be denoted $\mathbf{b}_{\mathbf{y}}^{S}(r) = (\theta \underset{\text{disc}}{s}(r), \ \theta_{\text{cont}}^{S}(r))$, for $s = 1, \ldots, S$ and $r = 0, \dots, R$. The number of fishes and the total number of iterations are user-defined parameters (in practice, $S = 3 \times P$ and R = 5000 are found to be good values). In additioneach fish has a weight s(r) at iteration, (11) which reflects the quality of the solution.

Initialization. The initial position $\theta^{s}(0)$ $(\theta_{\text{disc}}^{s}(0), \theta_{\text{cont}}^{s}(0))$ of each fish is assigned and omly. The continuous $\mathsf{vect} \mathfrak{S}_{\mathsf{rht}}(0)$ is drawn from a uniform distribution over bont, but for the discret part, it is advantageous to use a non-uniform distribution to initializ the edge parametenssuch a way that (0) is equal to -1 or 1 with probability 4, and 0 with probability 2, for $i, j = 1, \ldots, d_i$ which introduces a bias toweres and-1. This is in agreement with the biological observation that GRNs tend to be sparsely conficiential value -1/2 or 1/2 with equal probabilities, for . . . , d

Individual movement operator. This is an exploratory step

where each fish independently moves adiatratice in a random directions long as this increases the fitness

The step size bounds, for $q = 1, \ldots, Q$ shrink linearly The position of each Fishupdated according to: with r, in order to ensure convergence and emphasize r instance rexploitation over exploration at later itehationsimplementatiothe initial and final value $\mathbf{F}_q(1)$ and $\mathbf{T}_q(R)$ are setrespective by 10% and 0.01% of the range (i.e., the difference between upped lowerbounds)of the

$$\theta_{\text{ind}}^{s}(r) = \theta_{\text{ind}}^{s}(r) + \frac{P \int_{S^{o}=1}^{s} \Delta\theta_{\text{ind}}^{s^{o}}(r)(L \kappa(\theta_{\text{ind}}^{s^{o}}(r)) - L \kappa(\theta^{s^{o}}(r-1)))}{S \int_{S^{o}=1}^{s} (L \kappa(\theta_{\text{ind}}^{s^{o}}(r)) - L \kappa(\theta^{s^{o}}(r-1)))}.$$
(18)

corresponding continuous parameter — these valu**esheatispl**acement in discrete parameter space is quantize modified by the user, desired Now, $\Delta \theta_{\mathrm{disc,ind}}^s$ (r) needs following the same procedure adopted to discretize the to be quantized into the lattice 1 order to be individual movement displacement vector. added to the discrete component of the current fis**ic presting** volitive movementerator his is similar to The quantization scheme we adopt here is a generalization ovemens tep but now the fish move in adaptive thresholds:

of the method for binary parameters in [43]. We define two, depending on whether the fish school is success after the previous steips, its totalweigh increaseer not.If the fish school is succetisful it should contract, changing from exploration to exploitation mode. Otherwis (15) it should expand in order to explore the space insore. is accomplished by first fining the curre lish school barycenter:

thr
$$_{\text{pos}}^{s}(r) = \max_{+} (\Delta \theta_{\text{disc,ind}}^{s}(r)) \times \frac{r}{R}$$
,
thr $_{\text{neg}}^{s}(r) = \min_{-} (\Delta \theta_{\text{disc,ind}}^{s}(r)) \times \frac{r}{R}$, (15)

where the operator $ax_+(v)$ is equal to the maximum of the components vector if at leastone ofthem is positiveand equato zero otherwise imilarlymin - (v)

$$\mathbf{b}(r) = \frac{P \underset{S=1}{S} W^{S}(r)\theta_{\text{inst}}^{S}(r)}{P \underset{S=1}{S} W^{S}(r)}.$$
 (19)

is equal to the minimum of the componerifsableast For each fishafter the collective instinctive movement at one of them is negatived equal to zerotherwise the iteration, let $\xi^{S}(r) = \theta_{inst}^{S}(r) - b(r) = (\xi_{1}^{S}(r), \dots, \xi_{n}^{S}(r))$ factof/R increases the thresholds (in magnitud $m{\epsilon}$) withthe position vector with respect to the school barycen to favor exploitation over exploration at later iteration $\Phi_{\text{disc,vol}}^{\text{S}}(r)$ = $(\Delta \theta \text{ sont,vol} \text{ } (r), \Delta \theta \text{ sont,vol} \text{ } (r))$ be the collective favor exploitation over exploration at later iteration $\Phi_{\text{disc,vol}}^{\text{S}}(r)$ and $\Phi_{\text{cont,vol}}^{\text{S}}(r)$ be the collection over exploration at later iteration $\Phi_{\text{cont,vol}}^{\text{S}}(r)$ and $\Phi_{\text{cont,vol}}^$ ensure convergenExploitation could be understood time volitive displacement tor for fish at iteration. an analogy to tree depth-first search as opposed to **Vergton** $\theta_{\mathrm{disc,vol}}^{s}(r)$ is drawn from a uniform distribution ration, which would equate to tree breadth-first bearoner the rectangulategion $[0, \frac{S}{6}] \times [0, \frac{S}{62}]_{td}$ and exploration mode the algorithm widens the search in the individual mov parameter space, while in exploitation mode, the algorithm $_{\mathrm{cont,vol}}^{\mathrm{S}}(r)$ is drawn from uniform distribution attempts to get a more accurate result in a small acceptable ectangular gion $[0, 2t_1(r)\xi^s_{d^2+d+1}(r) imes 1]$ parameter space. $[0, 2 v_0(r) \xi_{d^2+d+Q}^S(r)]$, where $v_1(r), \ldots, v_r(r)$ are the same The quantized discrete displacement vector is obtained as used in the individual vergent tended by assigning to a positive component if it is larger technols successful if $v_1(r) > v_2(r) > v_3(r) >$

 ${\sf thr}_{\sf pos}^{\sf s}({\it r})$, assigning 1 to a negative component is then its radius should contract, and smaller that $\prod_{n=0}^{s}(r)$, and assigning to all other components (no movement). Then the position isfujishated

 $\theta_{\text{vol}}^{\text{S}}(r) = \theta_{\text{inst}}^{\text{S}}(r) - \Delta \theta_{\text{vol}}^{\text{S}}(r)$ (20)

if the exploratory move causes an increase in fitnestherwisthe radius expands, the schockn escape a bad region, and

$$\theta_{\text{ind}}^{s}(r) = \begin{cases} \theta^{s}(r-1) + \Delta \theta & \frac{s}{\text{ind}}(r), \\ \text{if } L_{k}(\theta^{s}(r-1) + \Delta \theta & \frac{s}{\text{ind}}(r)) > L_{k}(\theta^{s}(r-1)), \\ \theta^{s}(r-1), & \text{otherwise.} \end{cases}$$
(16)

$$\theta_{\text{vol}}^{\text{s}}\left(r\right) = \theta_{\text{inst}}^{\text{s}}\left(r\right) + \Delta\theta_{\text{vol}}^{\text{s}}\left(r\right),$$
 (21)

The new position of the fish $\theta \sim 0$ s $\theta \sim 0$ (r).

where $\frac{(16)}{k}$ is the penalized log-likelihood the model, $\frac{(16)}{3}$ RESULTS defined in the previous section.absorbing boundary In this section, e presenthe resulof a comprehensive

where it encounters it.

on the fitness improvement the previous individual noted the default values for all PALLAS fixed parameters movement, if any: and estimation intervals are used, as described in the pre

$$W^{s}(r) = w^{-s}(r-1) + \frac{L_{k}(\theta_{\mathsf{ind}}^{s}(r)) - L_{k}(\theta^{s}(r-1))}{\max_{s} \{L_{k}(\theta_{\mathsf{ind}}^{s}(r)) - L_{k}(\theta^{s}(r-1))\}}.$$

3.1 Network Size

ous sections.

We consider networks with 6 to 10 genes. These are typic Collective instinctive movementatorThis operatornetwork sizes encountered in biomedisearclMost makes the fish thatd successimblividualmovementsknown molecular pathways invofuedastancia, cellinfluence the collective direction of movement of t**signalized,** metabolic processes, and the cell cycle, contain JOURNAL OF TEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015

TABLE 1Number of pathwaysyerage number of general distance is related to the dynamical of the per pathway for different species networks, since it has to do with how the Boolean functio differ.

| H.sapiens | pathways | genes/pathway | |
|--|----------|---------------|--|
| KEGG | 237 | 72.4 | |
| WikiPathways | 135 | 46.3 | |
| HumanCyc | 290 | 7.2 | |
| M.musculus | pathways | genes/pathway | |
| KEGG | 218 | 74.6 | |
| WikiPathways | 140 | 57.8 | |
| MouseCyc | 323 | 8.0 | |
| S.cerevisiae | pathways | genes/pathway | |
| KEGG | 98 | 35.2 | |
| WikiPathways | 125 | 11.8 | |
| YeastCyc | 184 | 6.5 | |
| M.tuberculosis H37Rvpathways genes/pathway | | | |
| KEGG | 110 | 32.5 | |
| WikiPathways | 8 | 22.3 | |
| MTBRvCyc | 234 | 5.7 | |

3.2.2 Edge-Calling Accuracy Rates An edge in the groundtruth network represents a relationship between two genHere we consider directionality (an edge from geneto gene is distinctrom an edge from gene to gene), but disregard activation/inhibition relationships (this is done because some of the methods t which PALLAS is compared in this section do not capture activation/inhibitiobatTP andFN be the total number of directional edges that are correctly detected (irrespect of inhibition/activation) and incorrectly missed by the infe ence algorithm, respectively. Similarly det be the total number of directional edges that are incorrectly four and correctly missespectiveWe define the following

small number of components. This could be a bias towards ling accuracy rates: experiments, but it is more likely to be a result of the fundamental constraints of biological processes: huge state spaces could be prone to non-robust, chaotic dynamical behaviour

$$TPR = \frac{TP}{TP + FN}.$$
 (23)

extracted from [44], substantiates this claim using a diverige specificity/True Negative Rate (SPC): sample of metabolic pathways for four medganisms from majorpublic databases (KEGG [45]kiPathways [46],[47],BioCyc [48],49],[50]).The average number of genes per pathway in KEGG is around 61, in WikiPathwayfii) Precision/Positive Predictive Value (PPV):

biologicaprocesseshatimpinge on a single substrate,

and unnecessary energy expenditure by thatdelf,

$$SPC = \frac{TN}{FP + TN}.$$
 (24)

it is 39, and in BioCycjt is 8. In [51], the BioCyc and KEGG databases are compared, and it is shown that KEGG pathways are substrate-ceinerthey combine multiple

$$PPV = \frac{TP}{TP + FP}.$$
 (25)

meaning that individual KEGG pathways can contaîn3alfeeperiments with Synthetic Data native routes of biosynthesis (or catabolism) of a substrate mmalian Cell-Cycle Network with Synthetic RNAeither from one organism or from multiple organism or from the fro conclusion is that the BioCyc pathway conceptualization is

closer to a single conserved biological process than that of KEGG. That is to say 10-gene networks are not small fo single biological process, and in fact it is above the av-BioCyc network size in all species considered.

CycD

3.2 Performance Criteria

The problem ocomparing networks is a nontriviral; there is not a single metric that captures both the topo and dynamical roperties of the networks [52] ere we consider two classes of metrics, one based on the diffe between the network functions; hakes into account the full regulatory relationships among genes, and the based on edge-calling accuracy rates, which considers only the network topology.

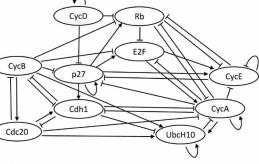


Fig. 2: Mammalian cell cycle network.

3.2.1 Network Function Distance

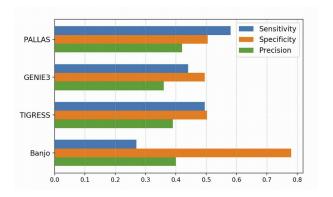
Here, we present results based on the well-known Man Let $f = (f_1, \dots, f_d)$ and $\hat{f} = (f_1, \dots, f_d)$ be the networkmalian Cell-Cycle network [58]ch is displayed in Fig. functions of the groundtruth and inferred networksurcher (Results for a different GRN are presented in the Su the compone interctions, and f_i are Boolean functions, lementary Materially estate vectors (CycD,Rb, on d variables $pr^{j}=1,\ldots,d$ see (1)The performance 27, E2F, CycE, CycA, Cdc20, Cdh1, UbcH10, CycB). This is criterion is the average number disagreeing Boolean large network, with a huge parameter space, for which estimation problem is hard. The gene interaction paramet functions between the two networks

$$\phi(\mathbf{f}, \hat{\mathbf{f}}) = \frac{1}{d \times 2^d} \sum_{i=1}^{X^d} \sum_{j=1}^{X^d} [f_i(\mathbf{x}^j) \oplus f_i(\mathbf{x}^j)].$$
 (22)

 $\phi(\mathbf{f},\hat{\mathbf{f}}) = \frac{1}{d \times 2^d} \sum_{i=1 \ i=1}^{d'} [f_i(\mathbf{x}^i) \oplus \hat{f}_i(\mathbf{x}^j)].$ (22) $\text{MDM2 network in the Supplementary Material. Once again the regulation biases are set to 2, for <math>i=1,\ldots,10$ } a_{ij} can be read from Figure 2 in the same way as in the p5

The transition noise paramistselected randomly in fluections or which the network function distance is apinterva[0.01, 0.1] The RNA-Seq data model parameteproperate on the other hand, the output of GABNI consist $\mu_i \equiv \mu = 0.1$, $\delta_i \equiv \delta = 3$, $\varphi_i \equiv \varphi = 5$, for $i = 1, \ldots, 10$. of positive (activating) or negative (inhibitory) interactions The sequencing depth is set 22.52 (500K-550K reads) r which we use the edge-calling accuracy rates defined and the time series length is fixed Hatr 50 we compare previously.

PALLAS with the GENIE3 [28] GRESS [30] and Banjo Average network function distances and edge-calling a [31] algorithms. Like PALLAS, these algorithms carcopecutes obtained over 20 repetitions of the experime directly on the noisy time series, without a need fo(2atbheach of the 10 networks) (with are displayed binarizationHoweverthey do not estimate observationAligures 4 and 5 (corresponding results for are parameters or provide activation/inhibition information/soin the Supplementary Material). Figure 4 shows that only the edge-calling accuracy rates in Section 3.2.12 experformance diest-Fitand PALLAS increases with appropriate here. Average rates obtained over 20 répetitions eries lengalique the performance of REVEAL of the experiment are displayed in Figure 3. One cansle EBN Metare mostly stable ALLAS perform better with similar specificity, PALLAS displays higher settbitivthe Best-Fätgorithmespecially whenis smaller. and precision than GENIE3 and TIGRESS. Although This as flects the fabatad-hoc binarization tife data not possible to adjust the specificity of Banjo to the sames less accurate with a smaller difference between levels we can see that its sensitivity is quitentant, activation/inactivation levels in the observed blizta, Banjo returned a very small number of edges overals idelleismined by. Figure 5 shows that ALLAS beats experimenPALLAS also displayed the highrectision GABNI in sensitivity throughout, wellas in specificity among all the algorithms. under sufficient data. Indeed, it is shown in Supplementar



3.3.2 Artificial Networks with Synthetic RNA-Seq Data

In this section we report results obtained on an ensection we report results of the report report results of the report results 10 randomly generated networks withgenes, where is displayed in Figure 6 [557]. We attempt to infer this each gene is regulated by 3 other genes on a subgegenetwork from gene expression time series datasets gener connectivity cluding activation and inhibition, ell as ated by [56] http://www.weizmann.ac.il/mcb/UriAlon/ regulation biases, are randomly chosen. The transitiownbiad/downloadable-daEa)ch time series contains parameteris selected randomly in the interval [0.1] 50 measurements fevery 6 minutes including the ini-RNA-Seq synthetidata are generated with parametetial zero concentrations pick the third datasetthe $\varphi_i \equiv \varphi = 1$ or 5, for $I = 1, \ldots, 8$ In the database or this experiment and compare the results first case, there is more observation noise, and theapailleththose found in [36], [37], [27]. is hardeithe parametersare allowed to vary uniformly The sparsity parametein (10) is chosen to produce over the interv $\frac{1}{2}$ or [1, 5] for i = 1, ..., 8 In the first about half of the possible edges in the six-gene network. case, the problem is harder, since the differences inrebs disptays in red the edges of the original network tha expression are smaller. Sequencing depth is £8€0 at were successfully recovered by a consensus of the top th (500K-550K reads). networks found by PALLAScording to the penalized

Here,we compare PALLAS with the Best-FittB2],likelihood score (the full network is displayed in the Suppl VEAL [33], FBNNet[35], and GABNI [34] algorithms.riTelestery Material). We can see that all inhibitory edges fro methodsapply to Boolean time seriess, they need to lexA were successfully detectioning PALLAS infers employ ad-hoc binarizationtbe gene expression datthe wrong direction between recA and lexA, the connection For the firsttwo, [54] recommends the use the KM3 detected. With a similar total number of inferred edges, [3 binarization method, while for GABNI, [9] recommefidathee opposite regulations all the inhibitory edges use of K-means binarization, as well as FBNNet; heareinferred as activating edges. While [36] finds most of t use those binarization methods here. The output ofiable best-edges, it misses the important edge from lexA to Fit, REVEAL and FBNNet algorithms are Boolean transitionally, [27] recovers only two of the edges.

3.4 Experiments with Real Data

inflates its specificity.

In this sections demonstrate the application of PALLAS to real microarray data from well-known biological system The complete resultscluding both false positives and false negatives, can be found in the Supplementary Mater

MateriathatGABNI detects very few edges under small sample size ohigh observation noisehich artificially

3.4.1 E. Coli SOS DNA Repair System

First,we consider the SOS DNA repair systemCpIE. In the normal tate the protein LexA is known to be a repressor to the SOS geiles on DNA is damaged he Fig. 3: Mammalian cell cycle experiment results rotein RecA becomes activated and mediates LexA autocleavage which causes activation of the SOS Acters. the activated SOS genes repair the damaged DNA, stops mediating LexA autocleavage and LexA represses the

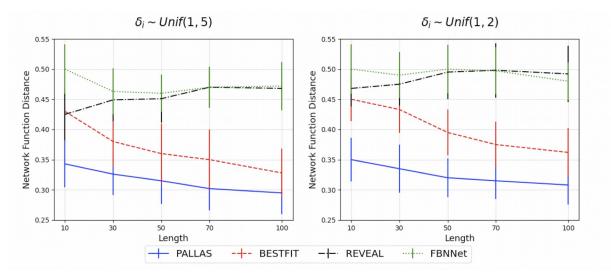


Fig.4:Comparison of network function distance among the BALLIAGEVEAL and FBNNet algorithms der different ranges.

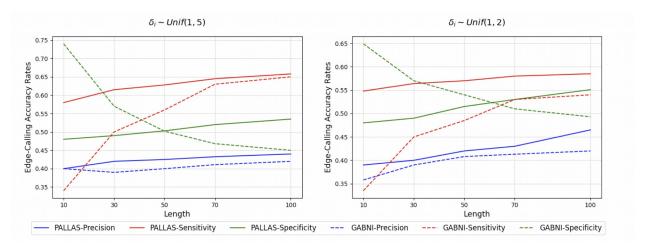
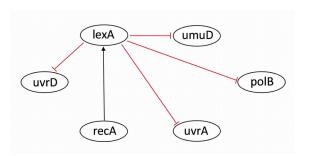


Fig. 5: Comparison of edge-calling accuracy rates between the PALLAS and GABNI algorithms angular differen



the ones successfully recovered by PALLAS).

3.4.2 E. Coli Biofilm Formation Pathway

[57],[58],[59]. Figure 7 displays a consensus gene network derived from these sources. The gene expression da used is from the Eoli Strain B/REL606 and is available at the Dryad DigitaRepository (https://datadryad.org/ resource/doi:10.5061/dryad.hj@@@j)This datasedonsists of 3 bacterial samples and 9 time points evenly space for each sample. The genes in this pathway display simila values atow expression levelst vary considerably at high expression levelscordinglywe assume a single Fig.6:SOS DNA repair system in E.coli (the red edgesaseline parameter μ for all genes, but the parameters δ_i and φ_i are allowed to differrom gene to generor i = 1, ..., 8. The sequencing depth is set at 02 (1k-50k reads) reflecting the low read counts in the data set.

database (https://www.genome.jp/kegg/)well as in

As in the previous experiment, sparsity parameter In this section demonstrate the performan \mathbf{PA} bf λ in (10) is chosen to produce abbat of the possible LAS on RNA-Seq time series expression data from ædgels-in the eight-gene networks 7 displays in red way involved in biofilm formation bootil, namely the the edges of the original twork that were successfully Rpos(sigmaS)/MIrA/CsgD cascadejch involves eightrecovered by a consensus of the top three networks found genes Rpos, MirA, CsgD, YciR, YoaD, BcsA, YaiC, YdaM. by PALLAS, according to penalized likelihood score (the Information on this pathway can be found in the KEGB network is displayed in the Supplementary Material).

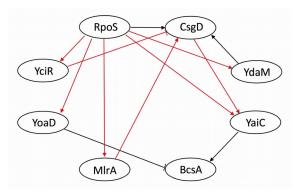


Fig. 7: Biofilm architecture of Escherichia coli (the reducestand H. G. Othmer, The topology office regulatory are the ones successfully recovered by PALLAS).

PALLAS successfully infers five out of the six important and the twork is robustly designed," Proceedings of the National vating interactions from RpoS. Most of the other connections of Sciences, vol. 101, no. 14, pp. 4781–4786, 2004. in the original network were correctly detected.

4 CONCLUSION

We presented in this paper PALAA6w framework for inference of Boolean gene regulatory networks from general 1999.
[15] I. Shmulevich, I. Gluhovsky, R. F. Hashimoto, E. R. Dougherty, an expression time series dense algorithm avoids ad-hoc W. Zhang Steady-state analysis of genetic regulatory and the series of genetic re ence of large networks by employing penalized maxim<u>fumctional genomics, vol. 4, no. 6, pp. 601–608, 2003.</u> likelihood as regularization methospplying particle [16] I. Shmulevic E. R. Doughertys. Kim, and W. Zhang, "Probafiltering for the computation of the likelihooding a novel version of the fish school search particle swarm 2002rithm to search the parameter space. Numerical experimental experiment using synthetic time series data show that PALLAS outo probabilistic boolean networks as models of genetic regulato forms other well-known GRN inference methods: formance of PALLAS was also demonstrated genue all 8] D. Cheng and H.Qi, "A linear representation dynamics of expression time series data from the SOS DNA repair and lean networks," IEEE Transaction tutomatic Control, Biofilm formation pathways is di. As a sophisticated 19] H. Qi and D.Cheng, Analysis and control of boolean networks: state-space method for Boolean GRN inference directly tramtensoproductapproach," in 2009 7th Asian Control noisy gene expression datathoutthe need ofad-hoc binarization, PALLAS is computationally expensive Resultagea-Net#Optimal state estimation for boolean dynamical tion time increases linearly with the number of fish usesein 2011, pp. 1050-1054. the MFSS algorith The user can adjust the running timeR. Tibshiran Regression shrinkage and selection via the lasso," Future work will include the implementation of PAL 22 M. Mani and UM. Braga-NetőMaximum-likelihood adaptive high-performance paradie hitectures bich willenable the inference of larger networks.

REFERENCES

- rent network inference methods," Nature Reviews Microbiol**seg**ychalgorithm,"in Computationalntelligencend 11th vol. 8, no. 10, p. 717, 2010.
- [2] E. P. van Somererl, F. Wesselsand M. J. Reinders; Linear modeling of genetic networks from experimental data." [agj ்டிகுan, F. B. L. Neto, and U. B. Neto, "Inference of gene regulato 2000, pp. 355-366.
- [3] P. D'haeseleex, Wen, S. Fuhrmanand R. Somogyi, Linear modeling of mrna expression levels during cns developmentMachine Learning for Signal Processing (MESP)2018, pp. injury," in Biocomputing'90orld Scientific, 1999, pp. 41–52.
- dynamic bayesian networks," Technical report, Computer Science egulatory networks with hybrid differential evolution and Division, University of California, Berkeley, CA, Tech. Rep., 1925 icle swarm optimization," Neural Networks, vol. 20, no. 8, pp
- [5] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using bayestan927, 2007. networks to analyze expression data," dóœmaþutational[27] M. Julfikar Islamī, M.S.R.,and A. M.A.H., "Gene regulatory biology, vol. 7, no. 3-4, pp. 601-620, 2000.

- [6] D. C. WeaverC. T. Workmanand G. D. Stormo, Modeling regulatory networks with weight matrices," in Biocomputing'99. World Scientific, 1999, pp. 112-123.
- [7] T. Chen, H. L. He, and G. M. Church, "Modeling gene expression with differentiaquations," in Biocomputing Marld Scientific, 1999, pp. 29-40.
- [8] M. B. EisenP.T. SpellmarP.O. Brownand D.BotsteiffCluster analysis and display@fnome-wide expression patterns," Proceedings of the National Academy of Sciences, vol. 95, no. 2 pp. 14 863-14 868, 1998.
- [9] S. Barman and Y.-Kwon,"A novelmutuainformation-based boolean network inference method from time-series gene expre sion data," PloS one, vol. 12, no. 2, p. e0171097, 2017.
- [10]S.A. Kauffman Metabolic stability and epigenesis in randomly constructed genetic nets," Jouthebreticalologyyol.22, no. 3, pp. 437-467, 1969.
 - interactions predicts the expression pattern of the segment pol genes in drosophila melanogaster," Journal of theoretical biolog vol. 223, no. 1, pp. 1–18, 2003.
- [13] A. Faure, A. Naldi, C. Chaouiya, and D. Thieffry, "Dynamical analysis of a generic boolean model for the control of the mammalia cell cycle," "Bionformatics", vol. 22, no. 14, pp. e124-e131, 2006
- [14] S. Huang, "Gene expression profiting etic networks d cellular statesm integrating concepttumorigenesis and drug discovery," Journal of molecular medicine, vol. 77, no. 6, pp. 469
- expression time series dana.algorithm avoids ad-noc "w. Zhang, Steady-state analysis of genetic regulatory networks binarization of the gene expression data and allows in the state of the gene expression data and allows in the state of the
 - bilistic boolean networks: a rule-based uncertainty model for ge regulatory networks," Bioinformatids,,no.2,pp.261-274,
 - 2002.
 - vol. 55, no. 10, pp. 2251–2258, 2010.
 - ConferenceIEEE, 2009, pp. 1352-1356.
- provided in the Supplementary Material indicate that execution on Signal stems and Computers (ASILOMAR).
- - filter for partially observed boolean dynasyistelms," IEEE Transactions on Signal Processing, vol. 65, no. 2, pp. 359-371, 2
 - [23] C. J. Bastos Filho, F. B. de Lima Neto, A. J. Lins, A. I. Nasciment and M.P.Lima,"A novel search algorithm based on fish school behavior," in SystemMan and Cybernetic2908SMC 2008. IEEE International ConferenceEEE, 2008, pp. 2646-2651.
- [1] R.De Smet and MarchalfAdvantages and limitations of 🕍 C. Bastos-Filho and BlascimentőAn enhanced fish school Brazilian Congress on Computational Intelligence (BRICS-CCI & CBIC), 2013 BRICS Congress Ene., 2013, pp. 152-157.
 - networks by maximum-likelihood adaptive filtering and discrete fish school search," in 2018 IEEE 28th International Workshop o
- 1–6. [4] K. Murphy, S. Mian et al., "Modelling gene expression dalactifungu, G. K. Venayagamoorthy, and D. C. Wunsch II, "Modeling of
 - network inference using prominent swarm intelligence methods

- [28] A. Irrthum, L. Wehenke, Geurts et l., "Inferring regulatory networks from expression data using tree-based method521 Pld6 Dougherty; Validation oinference procedures gene one, vol. 5, no. 9, p. e12776, 2010.
- [29]G. Sanguinettit al., "Gene regulatory network infereace: 2019, pp. 1-23.
- [30] A.-C. Haury, F. Mordelet P. Vera-Liconand J.-FVert, Tigress: BMC systems biology, vol. 6, no. 1, p. 145, 2012.
- [31] V. A. Smith, J. Yu, T. V. Smulders, A. J. Hartemink, and E. D. War 6isp. e66031, 2013. "Computational inference of neural information flow net[55]Ns,D. Sutton, B. T. Smith, V. G. Godoy, and G. C. Walker, "The so PLoS computational biology, vol. 2, no. 11, p. e161, 2006.
- [32] H. Lähdesäki, I. Shmulevickand O. Yli-Harja, On learning Machine learning, vol. 52, no. 1-2, pp. 147-167, 2003.
- engineering algorithm for inference of genetic network <u>archacademy of scie</u>nces, vol. 99, no. 16, pp. 10 555–10 560, 2002. tures," 1998.
- Bioinformatics, vol. 34, no. 17, pp. i927-i933, 2018.
- [35] L. Chen.D. Kulasiriand SSamarasing M. novel data-driven boolean modefor genetic egulatory networks," Frontiers [59] physiology, vol. 9, p. 1328, 2018.
- [36] S. Kimura, K. Sonoda, S. Yamane, H. Maeda, K. Matsumura, and M. HatakeyamæFunction approximation approach to the ²⁰¹⁴.

 inference ofeduced ngnemodels ofgenetic networks," BMG⁶⁰JJ. R. Houser, C. Barnhart, D. R. Boutz, S. M. Carroll, A. Dasgupta, bioinformatics, vol. 9, no. 1, p. 23, 2008.
- [37] S.KimuraŞ. Nakayamand M.HatakeyamaGenetic network inference as seriesof discrimination tasks," Bioinformatics, ysis ofcellular components incelireveals broad regulatory vol. 25, no. 7, pp. 918-925, 2009.
- [38] N. Ghaffari,M. R. Yousefi,C. D. Johnson, Ivanovand E.R. Dougherty Modeling the nexteneration sequencing sample processing pipeliner the purposes of classification <u>bioinformat</u>ics, vol. 14, no. 1, p. 307, 2013.
- [39] T. J. Hardcastle and M. Kelly, "bayseqempiricabayesian methods for identifying differential expression in sequer data," BMC bioinformatics, vol. 11, no. 1, p. 422, 2010.
- [40] S. Anders and W.Huber, "Differential expression analysis fo sequence count data," Genome bioob.gyl,,no.10,p. R106, 2010.
- [41] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, rial on particle filters for online nonlinear/non-gaussian k tracking," IEEE Transactions on signal proceds50;999.2, pp. 174-188, 2002.
- [42]M. Imaniand U.M. Braga-Net&Particle filters for partially observed boolean dynamigatems," Automatical, 87,pp. 238-250, 2018.
- [43]J.A. Sargo,S. M. Vieira,J.M. Sousaand C. J.Bastos Filho, signal processing. "Binary fish school search applied to feature selection: Application to icu readmissions," in Fuzzy Systems (FUZZ2021115); EE International Conferencel EEE, 2014, pp. 1366–1373.
- [44] H. Zhou, J. Jin, H. Zhang, B. Yi, M. Wozniak and L. Wong, "Intpath–an integrated pathway gene relationship datab modelorganisms and important thogens," in BMC systems biology, vol. 6, no. SEpringer, 2012, p. S2.
- [45] M. Kanehisa and S. Goto, "Kegg: kyoto encyclopedia of g genomes," Nucleic acids research, vol. 28, no. 1, pp. 27
- [46] T. Kelder, A. R. Pico, M. Iersel, B. Conklin, and C. Evelo, "Wikipathways: Pathway editing for the people," in 11Th Internation@bnference on Research in Computationgly, SanFrancisco (April 2007), 2007.
- [47] T. Kelder, A. R. Pico, K. Hanspers, M. P. Van Iersel, C. Ev B.R. Conklin, Mining biological pathways using wikipathways web services," PloS one, vol. 4, no. 7, p. e6447, 2009.
- [48] P. D. Karp, C. A. Ouzounis C. Moore-Kochlads, Goldovsky, P. Kaipa, D. Ahrén, S. Tsoka, N. Darzentas Y. Kunin, and N. López-Bigas Expansion ofhe biocyccollection of athvol. 33, no. 19, pp. 6083–6089, 2005.
- [49]P.D. Karp, "Pathway databases: ase study in computation in modeling/simulation, and intelligent and semiotic decision support

- <u>Computational Biology and Bioinfo</u>rmatics, vol. 4, no. 5, **ቯ1]3//.-64**een and P. Karp, "The outcomes of pathway database com putations depend on pathway ontology," Nucleic Acids Researc vol. 34, no. 13, pp. 3687–3697, 2006.
 - regulatory networks," Current genomics,no.6, pp. 351-359, 2007.
- introductory survey," in Gene Regulatory Netv&missger, [53]A. Faue, A. Naldi, C. Chaouiya, and D. Thieffry, "Dynamical analysis of a generic boolean model for the control of the mammalia cell cycle," Bioinformatics, vol. 22, no. 14, pp. e124-e131, 2006. trustfuInference ofene regulation using stability selecti65/4] N. Berestovsky and L. Nakhleh, "An evaluation of methods for in ferring boolean networks from time-series data," PloS one, vol. 8
- response: recent insights into umudc-dependent mutagenesis a dna damage tolerance," Annual review of genetics, vol. 34, 2000 gene regulatory networks under the boolean network mback. Ronen, R. Rosenberg, B. I. Shraiman, and U. Alon, "Assigning numbers to the arrows: parameterizing a gene regulation netwo [33]S. Liang, S. Fuhrman, and R. Somogyi, "Reveal, a general reverse ing accurate expression kinetics," Proceedings of the national states of the states of [57] R. Hengge, "Principles of c-di-gmp signalling in bacteria," Nature
- Reviews Microbiology, vol. 7, no. 4, p. 263, 2009. [34]S. Barman and Y.-Kwon, "A boolean network inference from Reviews Microbiology, vol. 7, no. 4, p. 263, 2009. time-series eneexpression data using a genealgorithm," [58] F. Mika and R. Hengge, "Small regulatory rnas in the control of the contro of motility and biofilm formation in eoli and salmonella, International journal of molecular sciences, vol. 14, no. 3, pp. 45
 - 4579, 2013. , "Small rnas in the control of rpos, csgd, and biofilm archite ture of escherichia coli," RNA biology, vol. 11, no. 5, pp. 494-507 2014.
 - J.K. Micheneß. D. NeedhamQ. Papoulasy. SridharaD. K. Sydykova et al., "Controlled measurement and comparative ana changes in response to glucose starvation," PLoS computationa biology, vol. 11, no. 8, p. e1004400, 2015.

Yukun Tan received the B.Sc. degree in Information Science and Engineering from Northeastern University, Shenyang, Liaoning, China, and the M.Sc.degree in Electrizad Computer Engineering from University of Illinois at Chicago, ChicagoUSA, in 2015 and 2016; spectively. He is currently pursuing the Ph.D. degree in the Department Electricalnd Computer Engineering, Texas A&M University, College Station, TX,USA.His current research interests include estimation of stochastic Boolean dynamical sys-

tems by particle swa**mith** applications in genomic and proteomic

Fernando B. Lima Neto received the graduation degree in computer science from the Catholic University of Pernambuc Recife Brazil, in 1990the masterss degree in computer science from the FederaUniversity oPernambuco, Recifein 1998the Diploma degree in artificial neuralnetworksfrom ImperiaCollege London,London, J.K., in 2002 and the Ph.Ddegree in artificial intelligence from the University of London London in 2002 In 2012 he was with INRIA (Paris-Rocquencourt), Paris, France,

as Professeur invit and in 2045yas on sabbatileadve with the University of MnstemsterGermanylle is an Associate Professor with the University Fernambucele is a HumboldFellow and Accredited Researcher by the Brazilian Control arc Braslia, way/genome databases to 160 genomes," Nucleic acids Presciar His current research interests include computational intellig (evolutionary, social, and hybrid metaheuristics), complex and stocl

symbolic theori<u>es," S</u>cience, vol. 293, no. 5537, pp. 204@**xxten,\$20d**e Lima Neto is a member of several scientific organiza-[50] A. V. Evsikov M. E. Dolan, M. P. Genrich E. Patekand C. J. tions among them the Brazilian Computer Society, the Brazilian Soc Bult, "Mousecyc: a curated biochemical pathways databas<mark>e գուր</mark>աքեն onal Intelligence (past pre**tsideնա**)որսtational Inlaboratory mouse," Genome Biology, vol. 10, no. 8, pp. 114 liganose Society, and the Systems, Man & Cybernetics Society.

JOURNAL OFIEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015



Ulisses Braga-Neto received the PhD degree in electricand computer engineering from The Johns Hopkins University!;imor@D, and has held post-doctgrastitions with the University of Texas Manderson Cancer Center, Houston, TX, and with the Oswaldo Cruz Foundation, Recife, Brazil.

He is a professor with the Department of Electrical and Computer Engineering and a member of the Center for Bioinformatics and Genomic Systems Engineering at Texas A&M University.

He is author of two books and nearly 150 peer-revieweathplapers, received the NSF Career Award in 2008.

11