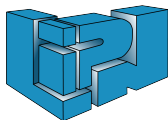


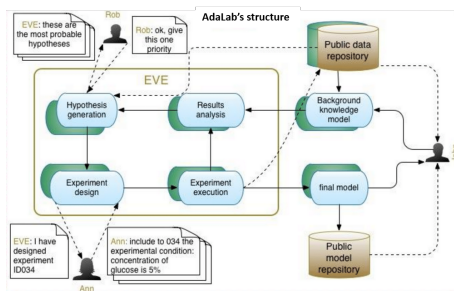
Dynamic genetic models revision for robot assisted scientific discovery

A. COUTANT, C. ROUVEIROL et al.



Eve and the AdaLab (Adaptive Laboratory) project

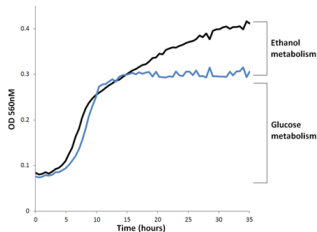
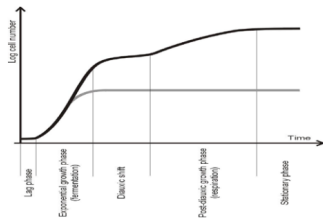
- AdaLab: CHIST-ERA project for cheaper, faster & better scientific discovery.
- Objective: framework for (semi-)automated knowledge discovery in systems biology.



- Eve: Manchester's robot scientist, able to perform experiments and output results [King 2009]
- LIPN role: machine learning for model revision & experiment selection

Diauxic shift (DS) in *S. Cerevisiae*

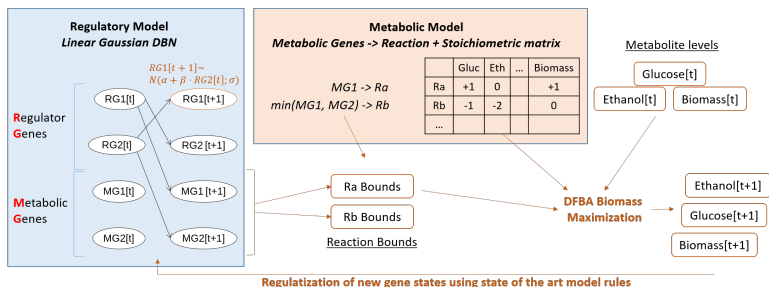
- Diauxic Shift = transition between growth phases: 1) glucose \rightarrow ethanol; 2) ethanol \rightarrow CO_2 .



- Goal: improve state of the art DS model [Geistlinger 2013] using Eve
- DS inference: auto analysis of curve features, mostly the post shift growth rate
- Experiment space: **1-gene KO**, nutrient levels, temperature & O_2

Model description

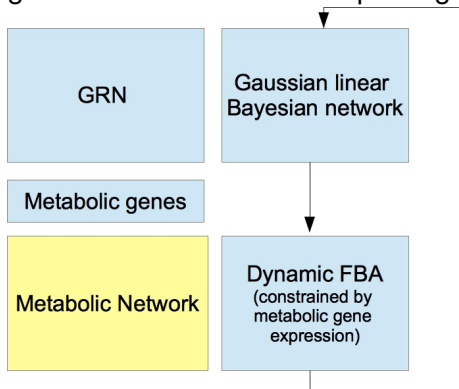
- Yeast model = **regulatory** + **metabolic** model
- **Regulatory model**: dynamic Bayesian network (DBN) with linear Gaussian conditionals, over regulatory and metabolic genes. **TO LEARN / REVISE.**



- **Metabolic model**: Stoichiometric matrix (metabolism chemical reactions \times metabolites) + enzymatic relations between metabolic genes and reactions. **STATIC.**

Model simulation

Simulating for n times = n repeats of: 1) DBN inference; 2) Metabolic inference with dynamic Flux Balance Analysis (DFBA); 3) Regularization of gene state for next time step using step 2 results.



Generative learning w/ spanning arborescences

Few gene regulation data is available for DS. "Fat" data, with 12 time points (samples) for over 3000 involved genes.

D : original (t_i, t_{i+1}) dataset

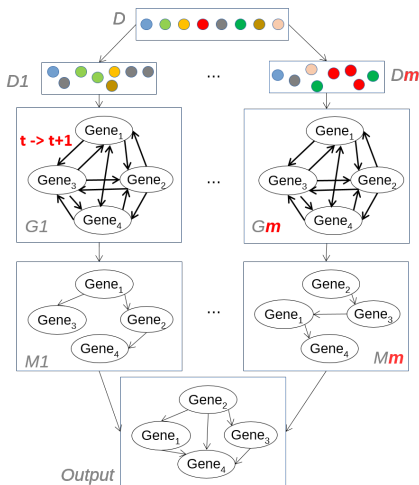
	Gene 1 [t]	Gene 2 [t]	Gene 3 [t]	Gene 4 [t]	Gene 1 [t+1]	Gene 2 [t+1]	Gene 3 [t+1]	Gene 4 [t+1]
● (t1, t2)	+	-	+	+	-	+	+	-
● (t2, t3)	-	+	+	-	+	+	+	-
● (t3, t4)	+	+	+	-	-	-	+	-
● (t4, t5)	-	-	+	-	-	-	-	+
● (t5, t6)	-	-	-	+	-	+	+	+
● (t6, t7)	-	+	+	+	+	-	-	+
● (t7, t8)	+	-	-	+	+	+	+	+
● (t8, t9)	+	+	+	+	-	-	+	-

Nodes at **t** time Nodes at **t+1** time

Algorithm

- Ensemble approach for DBN learning
- Combination of Edmonds arborescences
- Output = top edges by frequency ranking

Generative learning w/ spanning arborescences



Sample n individuals
(with replacement)

Build weighted DiGraph
edge probability = ep

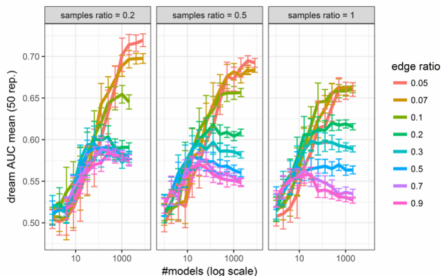
$$\text{Weight}(D_i, A \rightarrow B) = \text{BDeu}(D_i, A^T \rightarrow B^{T+1}) - \text{BDeu}(D_i, \emptyset \rightarrow B^{T+1})$$

Use Edmonds'
Spanning Arborescence
Algorithm

Choose the top k edges
according to
edges frequency

Results

- Lower edge ratio \rightarrow better results, but needs more models for convergence



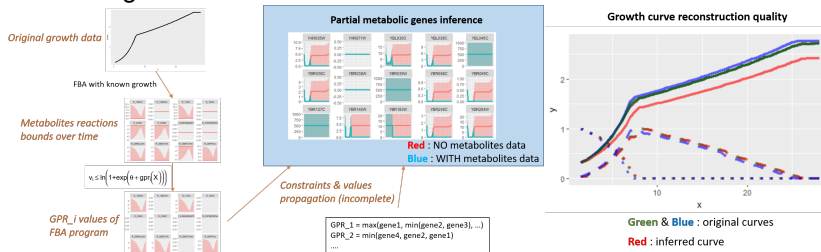
- Good ranking on DREAM 8 Challenges:
 - ▶ in-silico (80 samples, 20 features): 3rd
 - ▶ in-vivo (32 datasets, < 20 samples, 45 features): 2nd among no prior solutions
- **AdaLab: improved the fit to real curves as compared to state of the art DS model, using it as prior**

Active experiment selection w/ error blaming

From Eve growth curves, how to detect uncertainties in a model, to actively choose next experiments?

1 "Reverse" simulation of Eve curves

- ▶ DFBA with fixed growth + reaction constraint propagation to retrieve metabolic genes



- ▶ Results: good reconstruction for many strains despite partial reverse inference

Active experiment selection w/ error blaming

2 Gene ranking and experiment selection

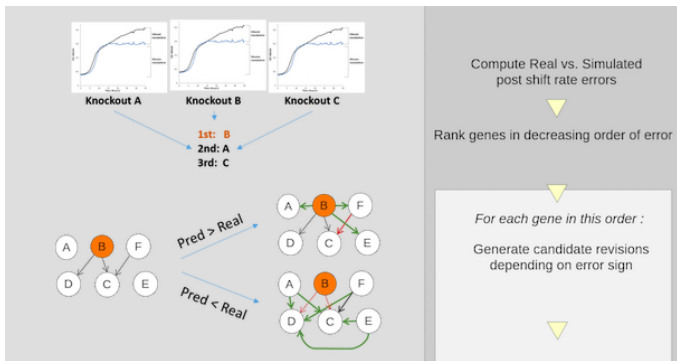
- ▶ Compare regular vs. reverse simulation deviations for available Eve curves
- ▶ Rank genes according to their mean deviations across curves
- ▶ Select genes with most errors as KO experiments to perform next

Model revision from Eve experiments

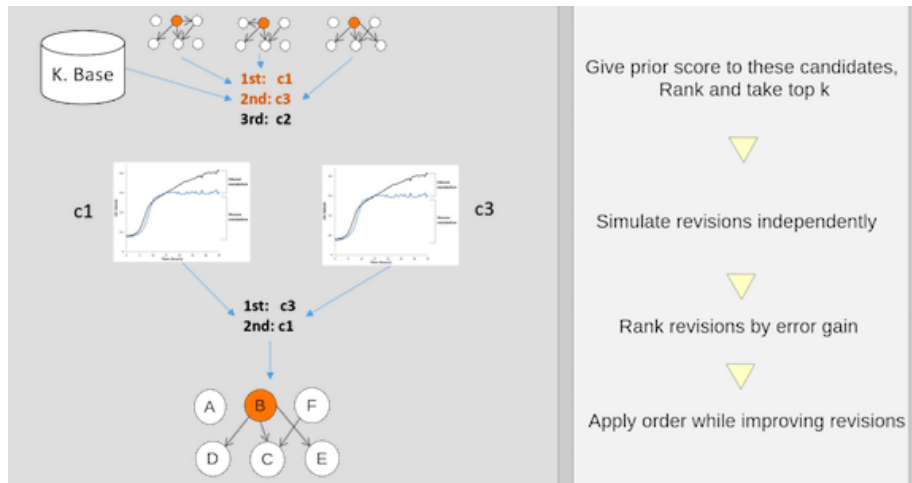
How to perform regulatory model updates w/ Eve's phenotypic growth curves & no transcriptomic data?

Algorithm

- KO prediction error driven
- Genes w/ high abs error are revised first
- Prior score on revisions before simulation \Rightarrow Restricted set of costly simulations

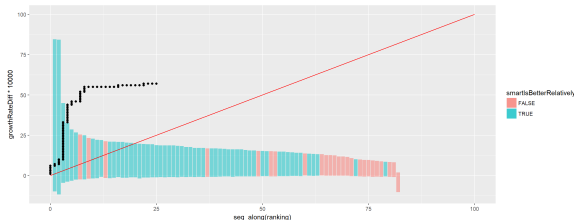


Model revision (ctd)



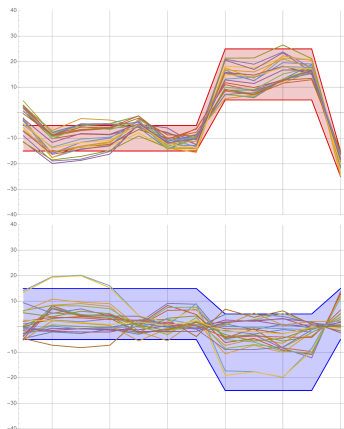
Results

- High reduction of strains prediction error (from 30% to 75%, depending on variant)
- Initializing w/ extended state of the art DS model, using spanning arborescences generative learning, gave better results than state of the art DS model alone
- Active selection of genes for revision gives better model improvement than random selection, validating active learning possibilities in AdaLab.

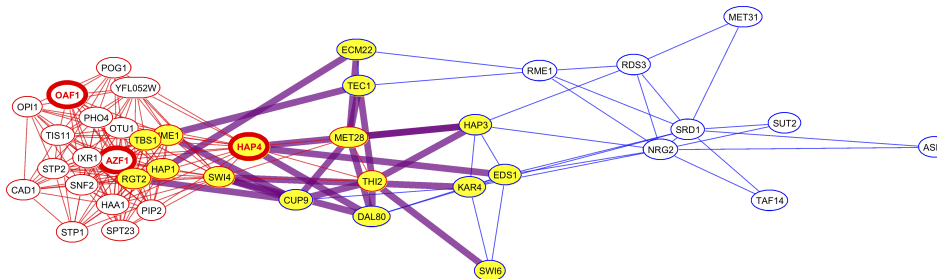


An alternative view for gene selection : Identification of DS genes w/ graph abstraction

- From co-regulation net labelled with activity profile (AP) in KB.
- Enum of (closed abstract pattern, graph extension) = (c, ext) (RG groups w/ \approx AP).
- Selection of $[(c_{\nearrow}, ext_{\nearrow}), (c_{\searrow}, ext_{\searrow})]$ linked in co-regulation net.



Identification of DS genes w/ graph abstraction



On-going/Future work

- First proof of concept, "pragmatic" approach
- Enforce empirical evaluation/formalisation

References & Acknowledgements

- A. Coutant and C. Rouveirol. Network inference of dynamic models by the combination of spanning arborescences. JOBIM 2017. Lille, France, 2017.
- L. Geistlinger et al. 2013. A comprehensive gene regulatory network for the diauxic shift in *Saccharomyces cerevisiae*. Nucleic acids research, 41(18), 8452-8463.
- R. D. King, et al. "The automation of science." Science 324.5923 (2009): 85-89.
- H. Soldano, G. Santini, and D. Bouthinon. "Formal Concept Analysis of Attributed Networks". Formal Concept Analysis in Social Network Analysis, Springer, 2017.
- D. Trejo Banos, P. Trebulle, and M. Elati. Integrating transcriptional activity in genome-scale models of metabolism. BMC Systems Biology (2017).

References & Acknowledgements

This work was funded by CHIST-ERA's 2013 call, and supported by other AdaLab participants: University of Manchester - Ross King, Katherine Roper & Martin Carpenter; Brunel University - Larissa Soldatova & Jacek Grzebyt; KU Leuven - Jan Ramon & Alex Sarafianos; Université d'Évry Val-d'Essonne - Mohammed Elati & Daniel Trejo; Technical advisors - Patrick Courtney (SiLA) & Steve Oliver (University of Cambridge)