Neural Information Processing Systems Conference

# *Causality Challenge*

# ***Pot-Luck***

## *Bring Your Own Problem*

***Isabelle Guyon, Clopinet***

***Constantin Aliferis and Alexander Statnikov, Vanderbilt Univ.***

***André Elisseeff and Jean-Philippe Pellet, IBM Zürich***

***Gregory F. Cooper, Pittsburg University***

***Peter Spirtes, Carnegie Mellon***

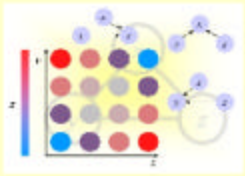
# *Causality Workbench*

- **Goal:** Benchmark causal discovery algorithms.
- **Method:**
  - Challenges.
  - Repository of datasets, tasks, models, software, etc.
  - Interactive workbench.
  - Weekly teleconference seminar.
- **So far…**
  - Causality Challenge #1: Causation and Prediction (WCCI 2008).
  - Causality Challenge #2: Pot-luck (NIPS 2008).
- **Winter, 2008:** Start developing an interactive workbench.
- **June, 2009:** KDD workshop on causality in time series?

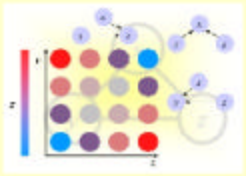
# *Why a new challenge?*

- **Causality challenge #1**
  - **Favor "depth"**
    - Single well defined task
    - Rigorous performance assessment
- **Causality challenge #2**
  - **Favor "breadth"**
    - Many different tasks
    - Encourage creativity

# http://clopinet.com/causality



## Causality Workbench



Home | Challenge | Pot-luck | **Repository** | Resources | Credits | Disclaimer | People

Data | Models | Software | Publications | Annoucements | Deposit | My account | Trash

Logged in as: **Isabelle Guyon - guyon@clopinet.com** [logout]

**Filtering by keyword: time.series** [X]

### PROMO: Simple causal effects in time series

Contact: **Jean-Philippe Pellet** - Submitted: 2008-09-15 18:57 - Views : 63

The PROMO dataset proposes the task to identify which promotions affect sales. Artificial data about 1000 promotion variables and 100 product sales is provided. The goal is to predict a 1000x100 boolean influence matrix, indicating for each (i,j)... [more/question/discuss/rate/edit...]

- **Authors:** Causality workbench team
- **Key facts:** This dataset contains artificial data about product sales and promotions as time series. There are 1000 binary promotions variables and 100 continuous product...
- **Keywords:** time.series, structural.equation.models
- Download BibTeX
- Download the data

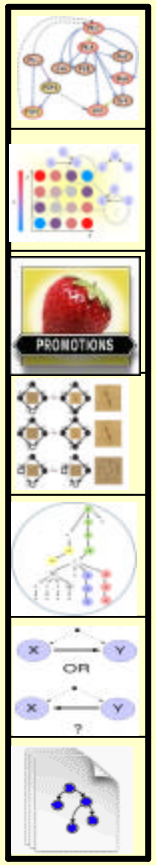
### SIGNET: Abscisic Acid Signaling Network

Contact: **Jerry Jenkins** - Submitted: 2008-09-17 21:53 - Views : 20

The objective is to determine the set of boolean rules that describe the interactions of the nodes within this plant signaling network. The dataset includes 300 separate boolean pseudodynamic simulations of the true rules, using an asynchronous... [more/question/discuss/rate/edit...]

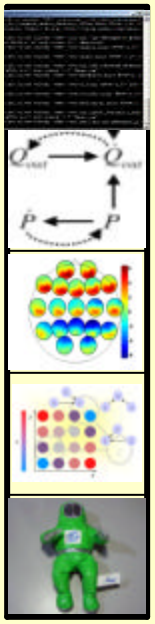
- **Authors:** Jerry W. Jenkins, Abhishek Soni
- **Key facts:** Simulated data with a Boolean network modeling a biological signaling network. Time series of 21
- Download BibTeX
- Download the data

# *Pot-Luck challenge*




| Task | Participants (views) | Type |
|---|---|---|
| CYTO | 2 (394) | real, self eval |
| LOCANET | 10 (558) | real, artif |
| PROMO | 3 (570) | artif, self eval |
| SIGNET | 2 (415) | artif |
| TIED | 1 (330) | artif |
| **CauseEffectPairs** | 5 (218) | real |
| **Stemmatology** | 0 (109) | real, self eval |

# *Other donated datasets*

| | Task | Views | Type |
|---|---|---|---|
|  | WebLogs | 90 | real, self eval |
| | MIDS | 65 | artif |
| | NOISE | 43 | real, artif |
| | SECOM | 59 | real |
| | SEFTI | 35 | real |

*http://clopinet.com/causality*

# *Winners*

***PRIZES***

| | |
|---|---|
| ***Best benchmark result:*** | **Kun Zhang and Aapo Hyvärinen** |
| ***Best contributed task:*** | **Guido Nolte** |

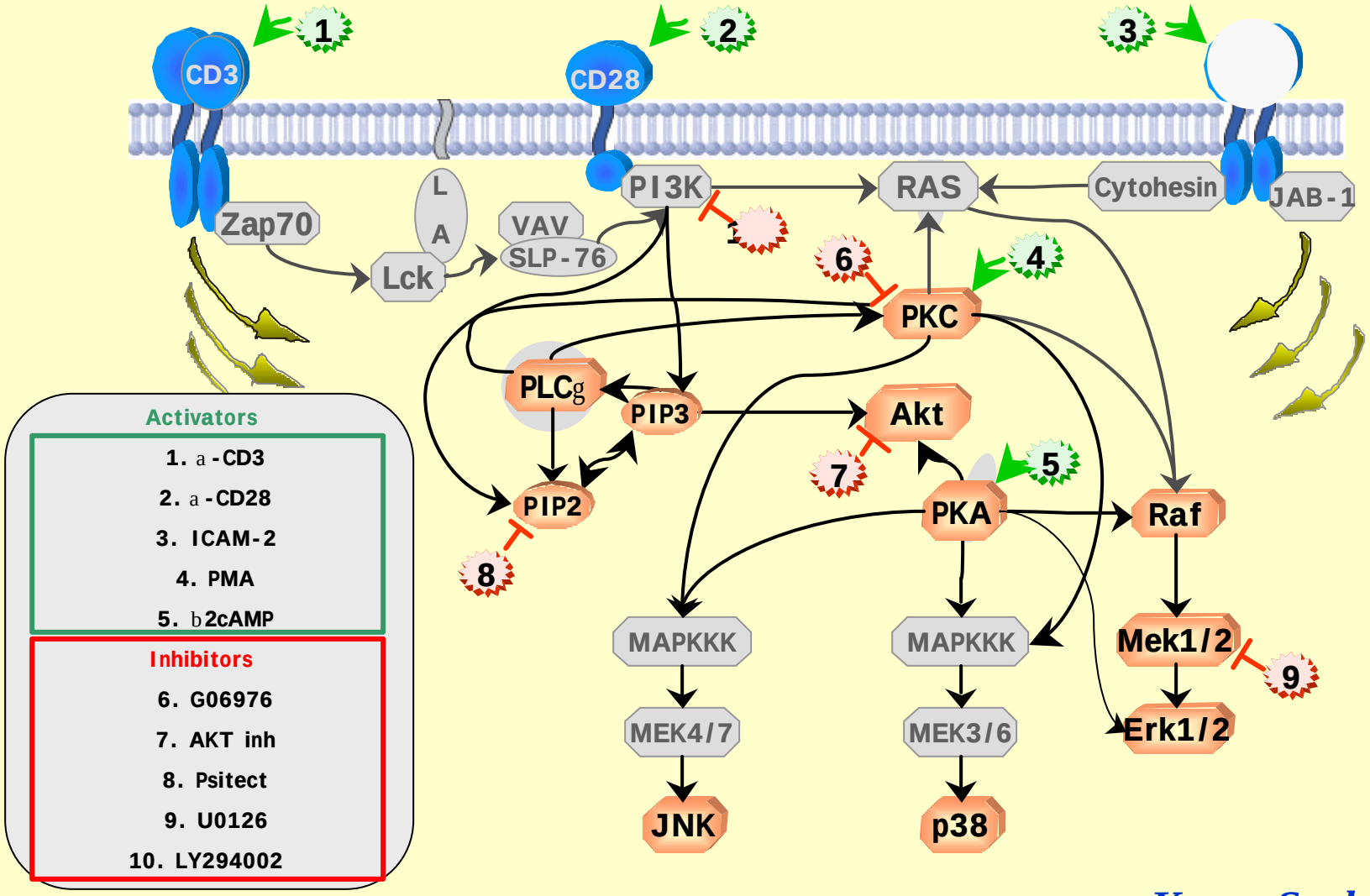
***MENTIONS***

***Significant advance on***

| | |
|---|---|
| ***REGED:*** | **Ernest Mwebaze and John Quinn** |
| ***SIDO:*** | **You Zhou, Changzhang Wang, Jianxin Yin, Zhi Geng** |
| ***SIGNET:*** | **Mehreen Saeed** |
| ***SIGNET:*** | **Cheng Zheng and Zhi Geng** |
| ***TIED:*** | **Advanced Analytics, Intel, LTD** |

*Causal discovery from real manipulations*

# *The CYTO problem*






*Karen Sachs et al*

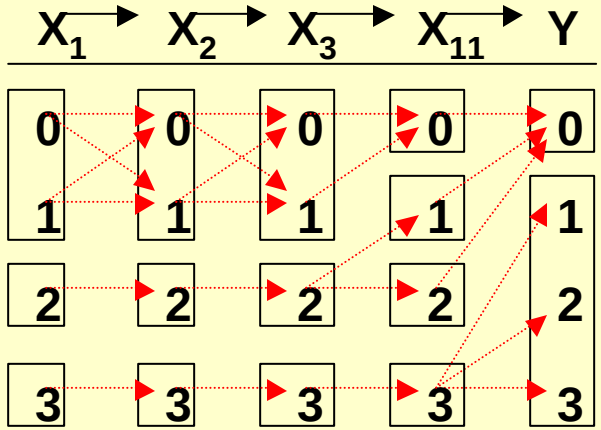
# *What if we cannot experiment?*
# *The LOCANET problem*

- Experiments may be infeasible, costly or unethical.
- Using only observations we may want to predict the effect of new policies.
- Policies may consist in manipulating several variables.
- **Task:** Find the local causal structure around a given target variable (depth 3 network) in four datasets (REGED, CINA, SIDO, and MARTI).

# *Multiple alternative solutions*
# *The TIED problem*

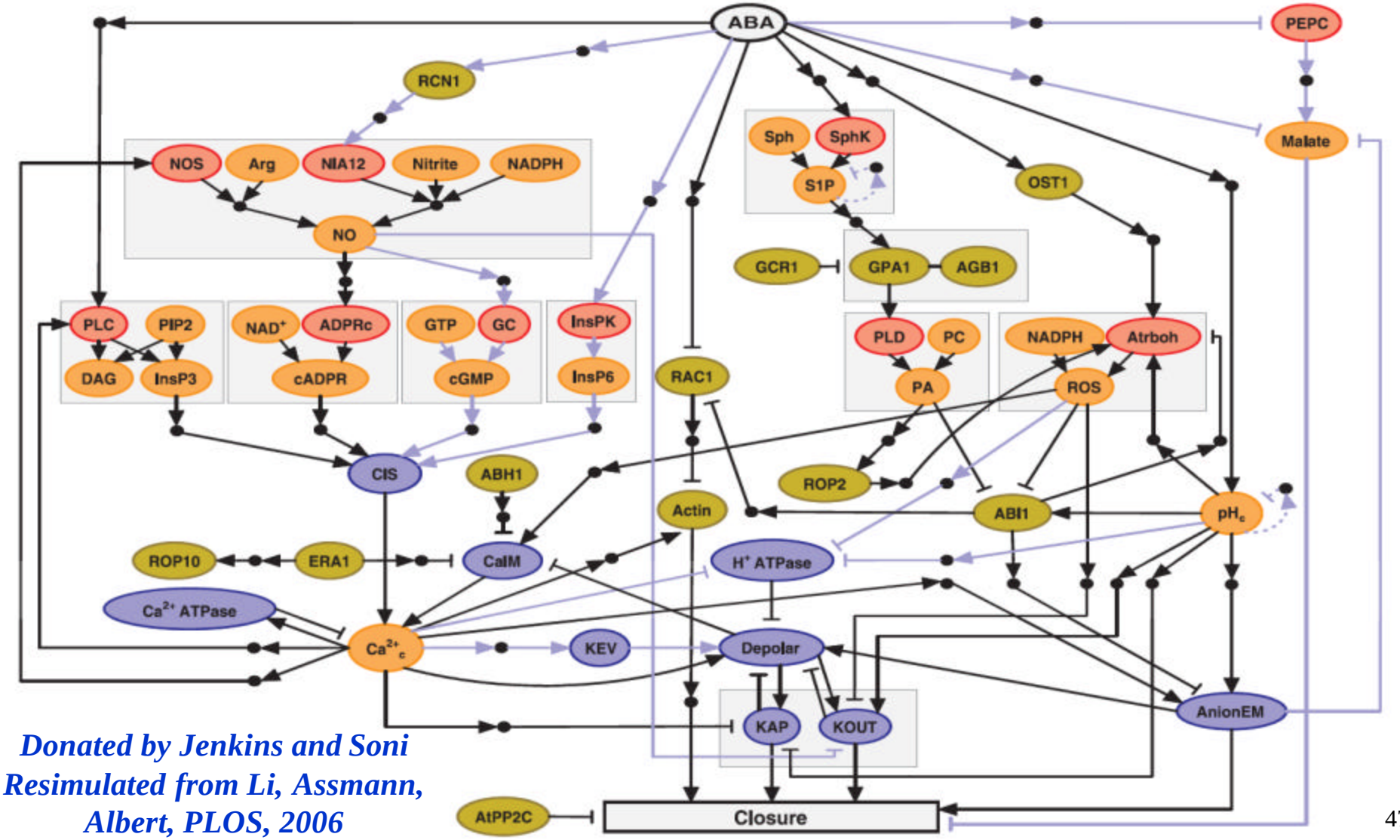
Two disjoint subsets of variables $\mathbf{V_1}$ and $\mathbf{V_2}$ are **Target Information Equivalent** w.r.t. target Y $TIE_Y(\mathbf{V_1}, \mathbf{V_2})$, iff:

- $\mathbf{V_1} \not\perp Y$
- $\mathbf{V_2} \not\perp Y$
- $\mathbf{V_1} \perp Y \mid \mathbf{V_2}$
- $\mathbf{V_2} \perp Y \mid \mathbf{V_1}$




$TIE_Y(X_1, X_2)$
$TIE_Y(X_1, X_3)$
$TIE_Y(X_1, X_{11})$
$TIE_Y(X_2, X_3)$
$TIE_Y(X_2, X_{11})$
$TIE_Y(X_3, X_{11})$

# *A dynamical system: SIGNET Abscisic Acid Signaling Network*




***Donated by Jenkins and Soni***
***Resimulated from Li, Assmann, Albert, PLOS, 2006***

# *Causality in time series: PROMO*
# *A simulated marketing task*

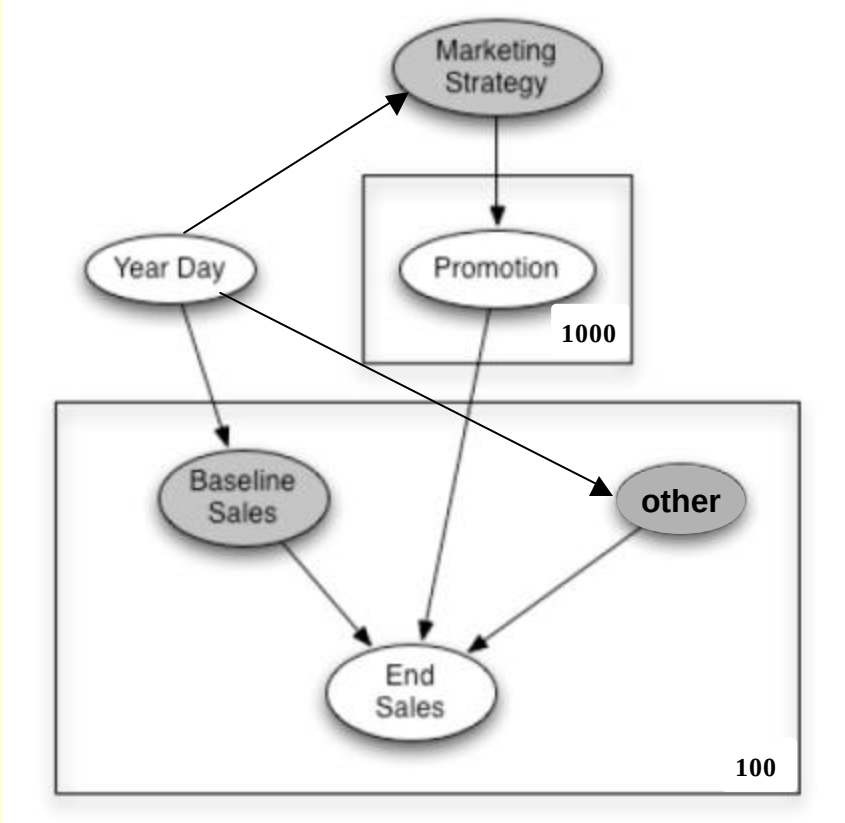
- 100 products
- 1000 promotions
- 3 years of daily data
- **Goal:** quantify the effect of promotions on sales.

The difficulties include:

- non iid samples

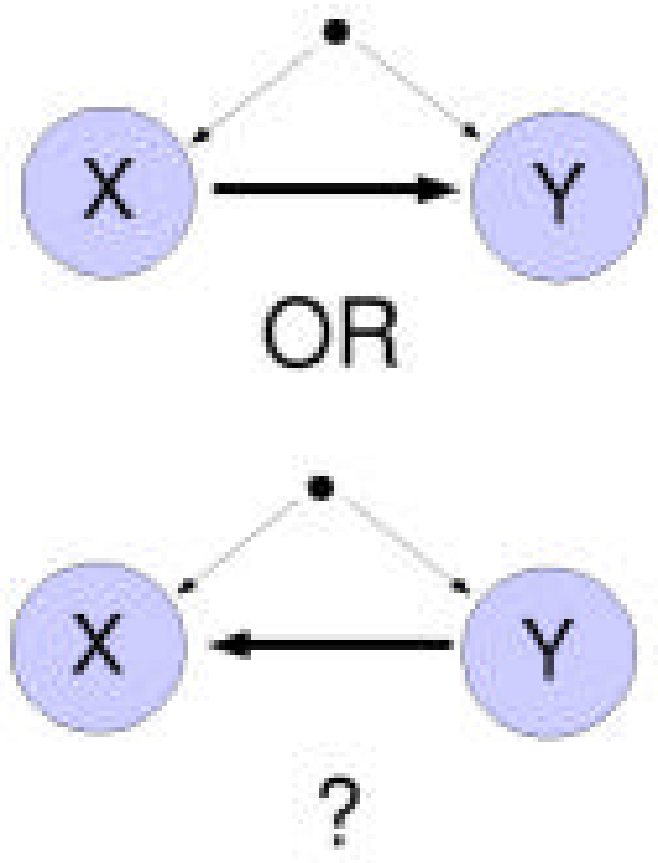
- seasonal effects

- promotions are binary, sales are continuous




***Jean-Philippe Pellet & André Elisseeff***

# *Causal direction among only two variables?*
# *The CauseEffectPairs problem*




***Dominik Janzing***

- Many causal discovery methods rely on tests of conditional independence between 3 or more variables.
- **Task:** Find the causal direction among pairs of variables (real data, e.g. temperature and altitude).

# *What's next?*

- **Proceedings of NIPS workshop (JMLR, early 2009):**
  - Submit revised 10-page paper by December 19, 2008.
- **Depth *vs.* breadth ® focus months:**
  - Teleconference presentations on one particular challenge.
  - Deadline for submission; result analysis and debate.
- **Causality challenge #3:**
  - Focus on time series.
  - Target KDD, June 2009.
- **Interactive workbench:**
  - Under development; target next NIPS.

*http://clopinet.com/causality*