Oregon State University

COLLEGE OF ENGINEERING | School of Electrical Engineering and Computer Science

Pooling vs Voting:

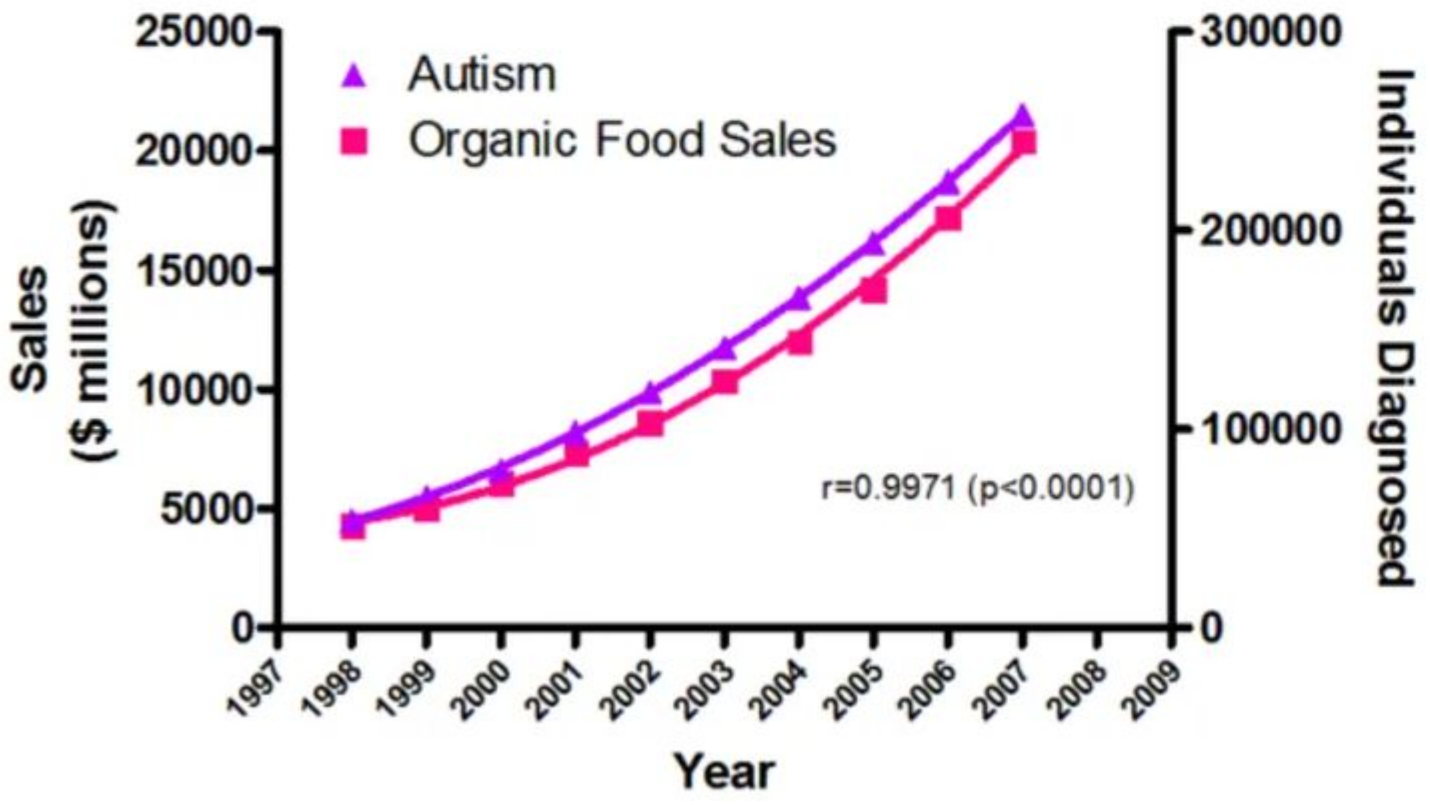
An Empirical Study of Learning Causal Structures

Meghamala Sinha, Prasad Tadepalli, Stephen A. Ramsey

AAAI Spring Symposium Beyond Curve Fitting: Causation, Counterfactuals, and Imagination-based AI, March 27, 2019

Confusion over Causality and Correlation

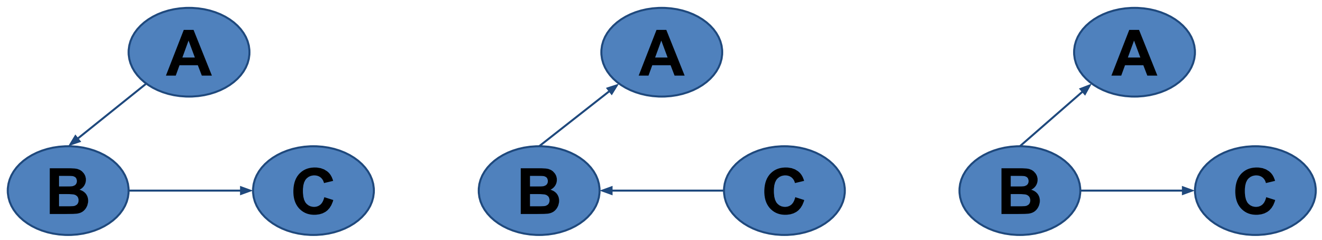
- Causally *unrelated* variables appear to be highly correlated
- *Spurious Correlation*, *Personal Anecdotes* & *Scientific Reporting* are unclear to determine *Causality*



Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Act

Effects of mixing Observational and Interventional data

- Observational data is cheap, accessible and identifies among Markov Equivalent structures
- Interventions are external manipulating of one or more variables in our system
- Enable us to differentiate among different causal structures compatible with an observation



Intervening at **B** ← **Switches on one of the 3 networks !!!**

- A dependence on both observational and interventional experiments is important
- A common approach is to pool all the data and learn a single causal model for Statistical Efficiency.

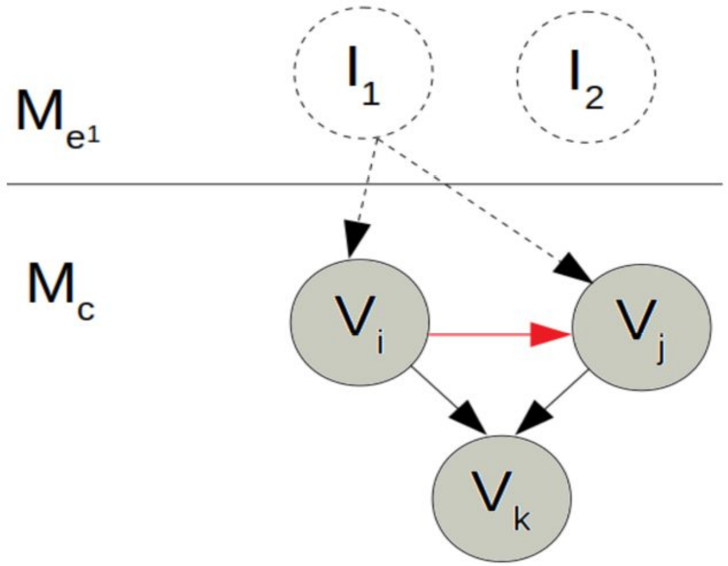
Types of Intervention:

Perfect *Known, fixed target* → Not applicable in Real World settings !!

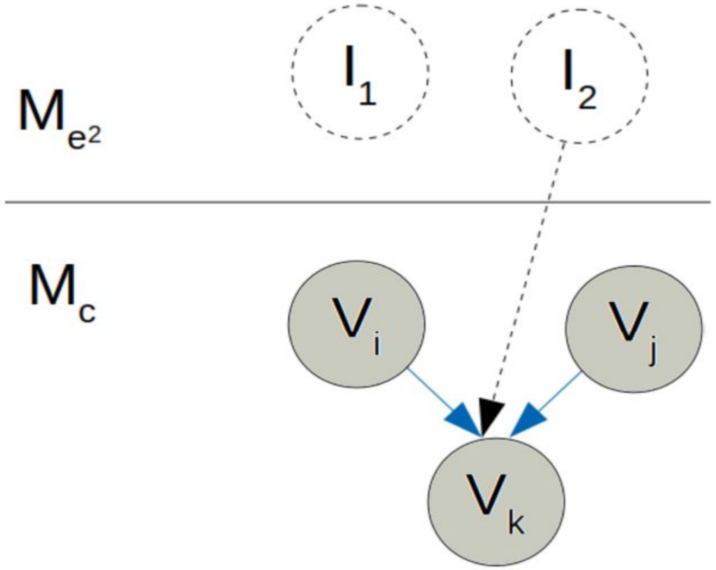
Imperfect *Unreliable, Soft targets* → "off-target" effects of drugs, gene knockouts etc

Uncertain *Unknown targets, Fat hands* → "off-target" effects of drugs, gene knockouts etc

Spurious Causal links



False Causal Dependence



False Causal Independence

Eberhardt, Frederick. "Sufficient condition for pooling data from different distributions." *First Symposium on Philosophy, History, and Methodology of Error*. 2006.

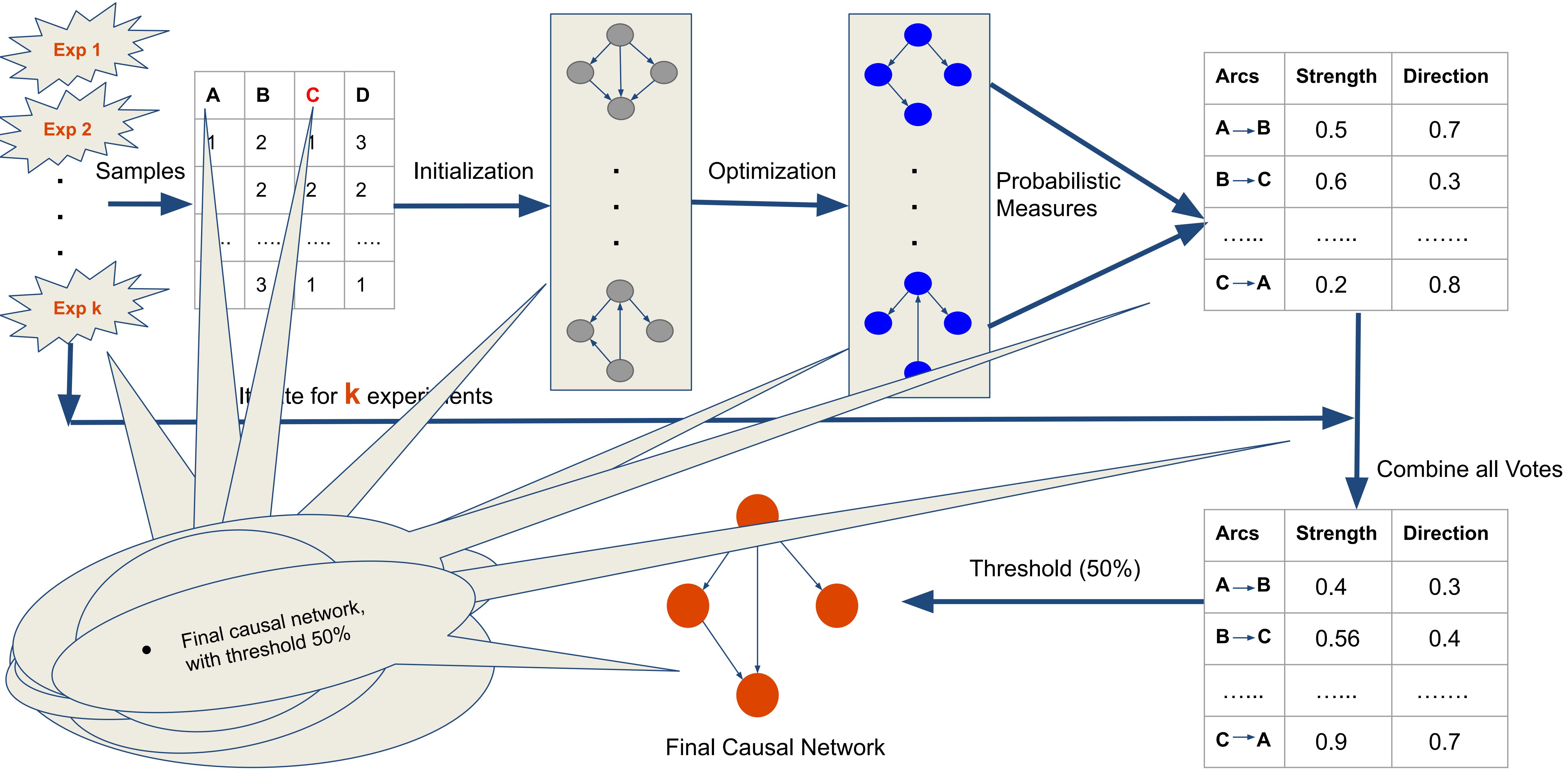
Contribution

- Detecting such spurious causal links might lose the very purpose of learning these networks

- We propose a way of handling uncertain interventions by learning causal information from different experiments separately and combining the results using a simple approach called ***"Learn and Vote"***.

- We found that our approach achieves a significant reduction of false causal discovery.

Our Approach: *Learn and Vote*



Our Approach: Learn and Vote

- Total number of “k” experiments
- Can be observational or interventional

Algorithm 1 Learn and Vote

Input: set of k experiments with dataset $D_1, D_2 \ldots D_k$

Output: DAG G^f = (E, V), final causal network

```
procedure OUR APPROACH
    for j=1 to k do
        N=nodes In D_j
        intv=Intervened nodes in D_j
        randomNet=createRandNet(N, 100)
        for l=1 to 100 do
            Net[l]=Tabu(randomNet[l], intv)
        arcProb[j]=arcStrength(Net)
    avgArcs = avgNetwork(arcProb)
    G^f = learnDAG(avgArcs,Threshold)
```

- Observed variables in our system

- Known Targets of intervention

- Initialization
- 100 DAGs

- Score based searching

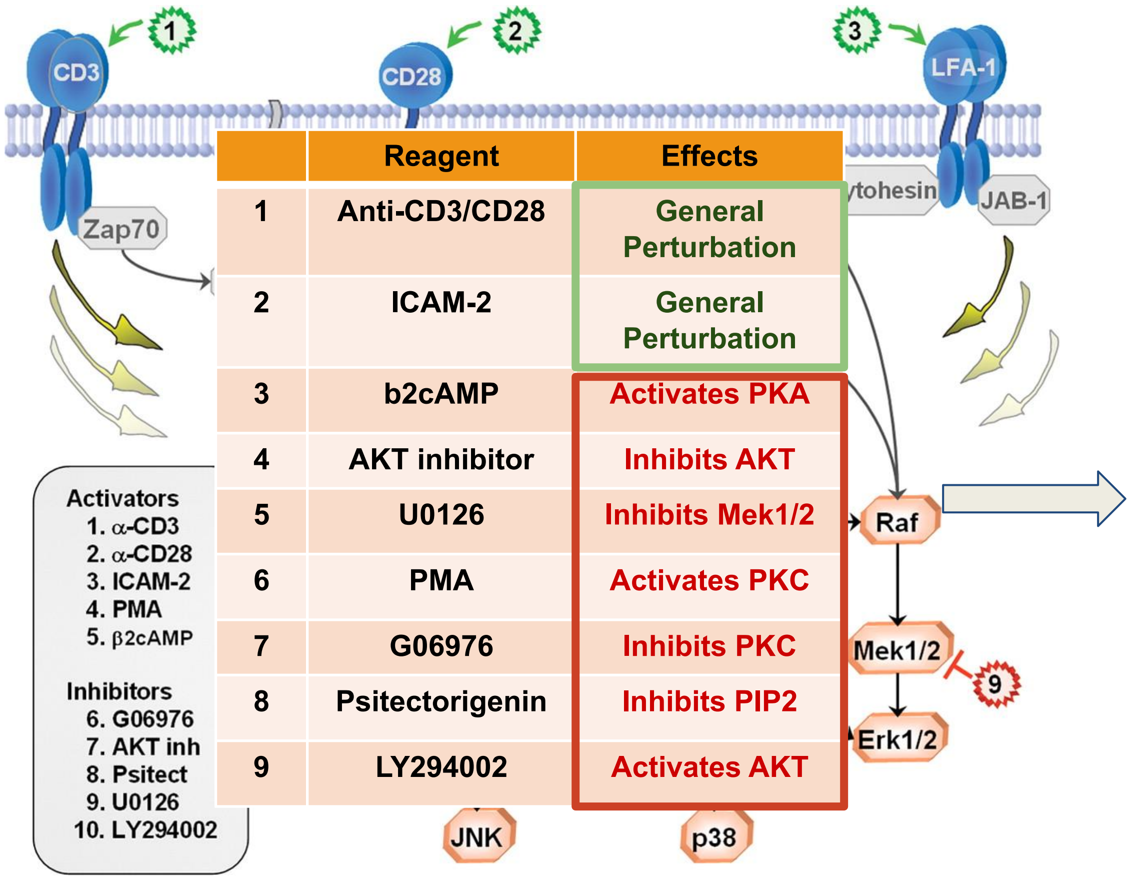
- List containing arc strength and direction in probability

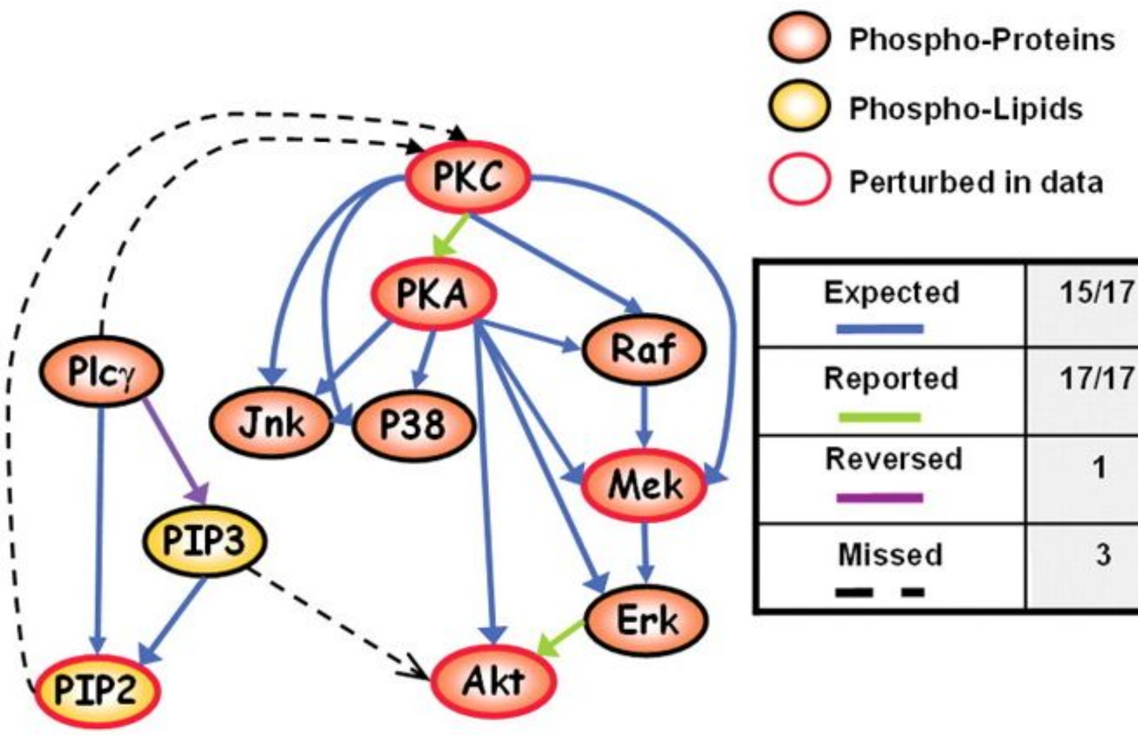
- Averaging over all experiments

- Final causal network, with threshold 50%

An Application: Cell Signalling Networks

	Reagent	Effects
1	Anti-CD3/CD28	General Perturbation
2	ICAM-2	General Perturbation
3	b2cAMP	Activates PKA
4	AKT inhibitor	Inhibits AKT
5	U0126	Inhibits Mek1/2
6	PMA	Activates PKC
7	G06976	Inhibits PKC
8	Psitectorigenin	Inhibits PIP2
9	LY294002	Activates AKT



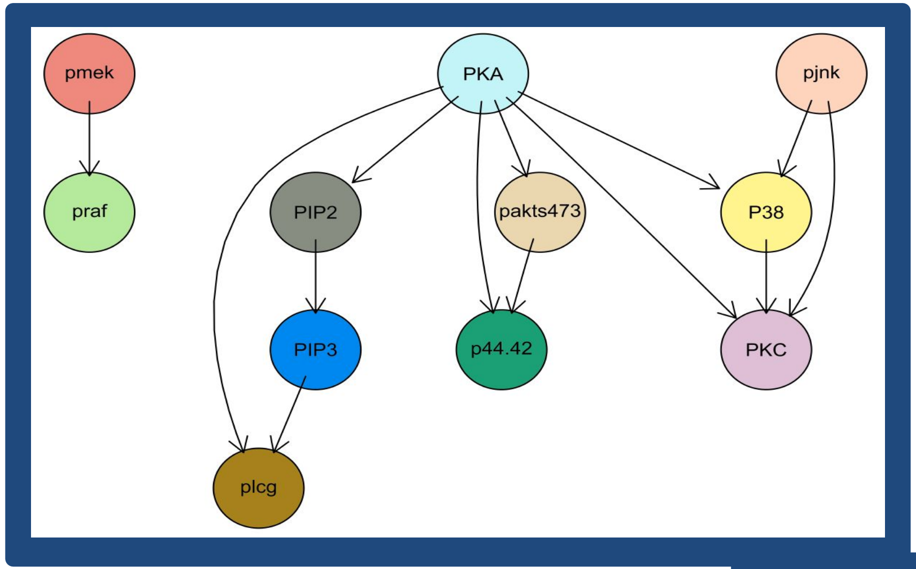


Expected	15/17
Reported	17/17
Reversed	1
Missed	3

Sachs, Karen, et al. "Causal protein-signaling networks derived from multiparameter single-cell data." *Science* 308.5721 (2005): 523-529.

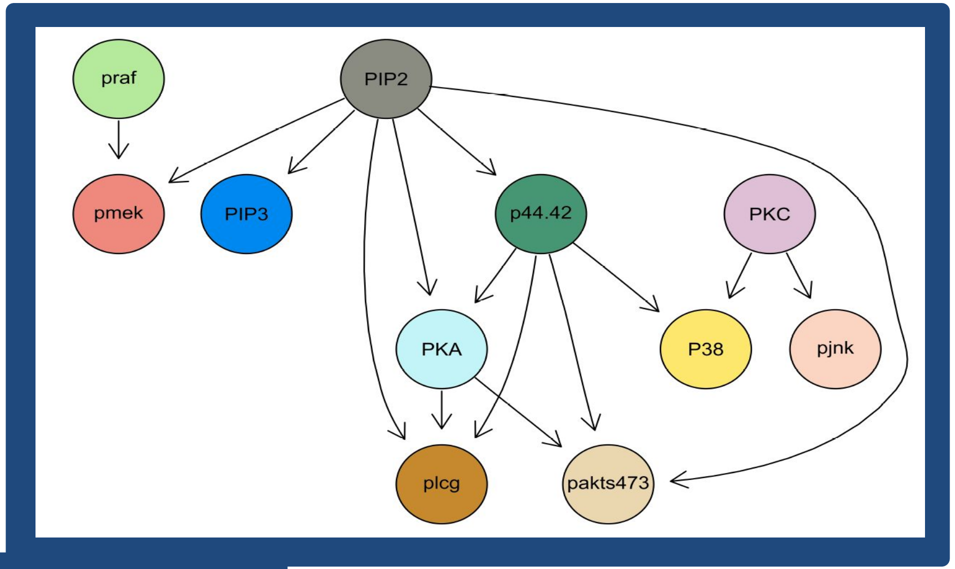
An Application: Biological Signalling Networks

Observational Study



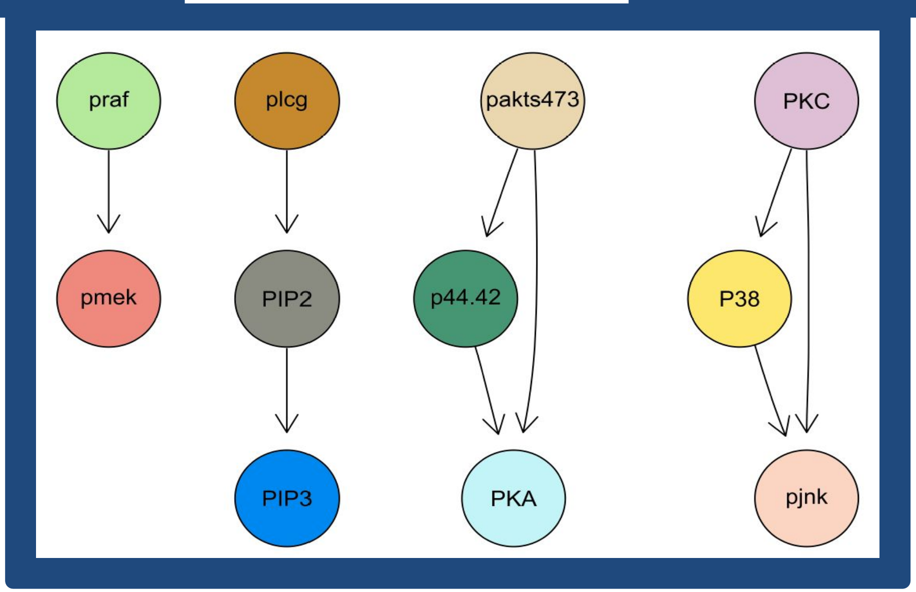
TP: 3, FP: 10, FN: 17

Observational + Interventional Study



TP: 5, FP: 10, FN: 15

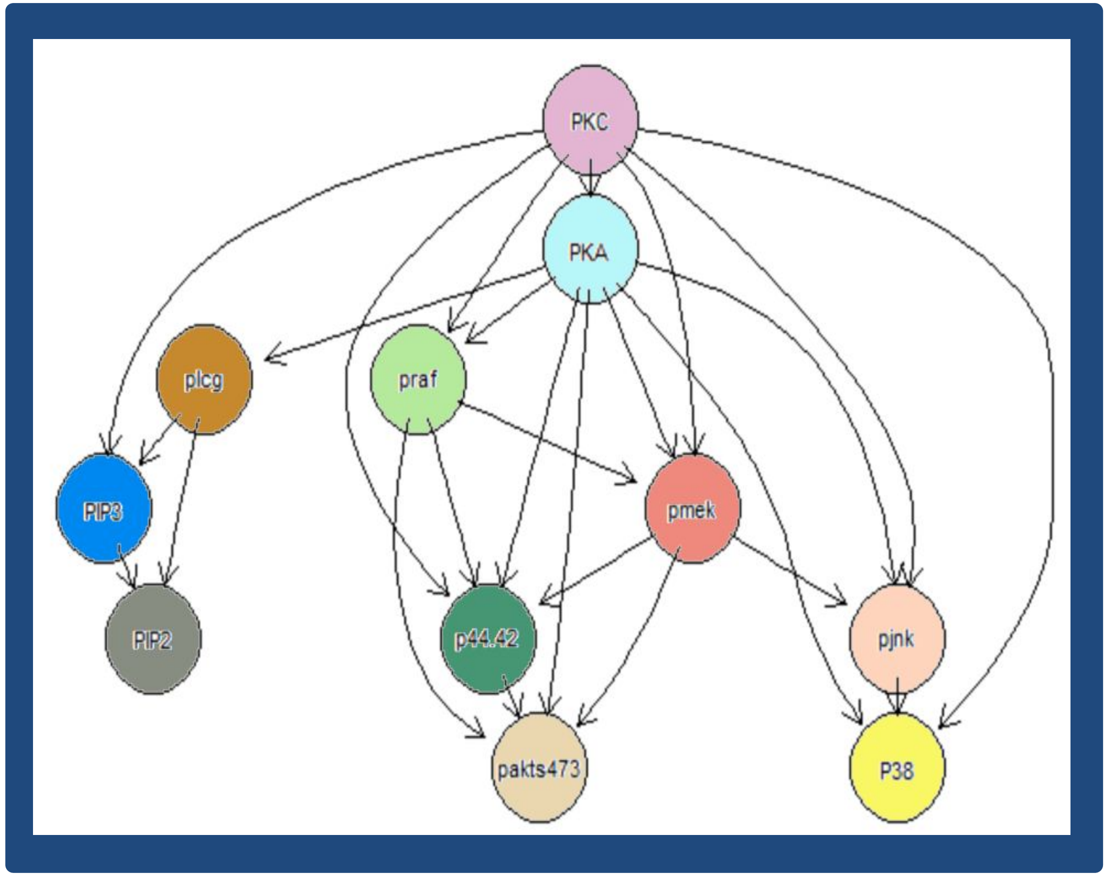
Learn and Vote



TP: 4, FP: 5, FN: 16

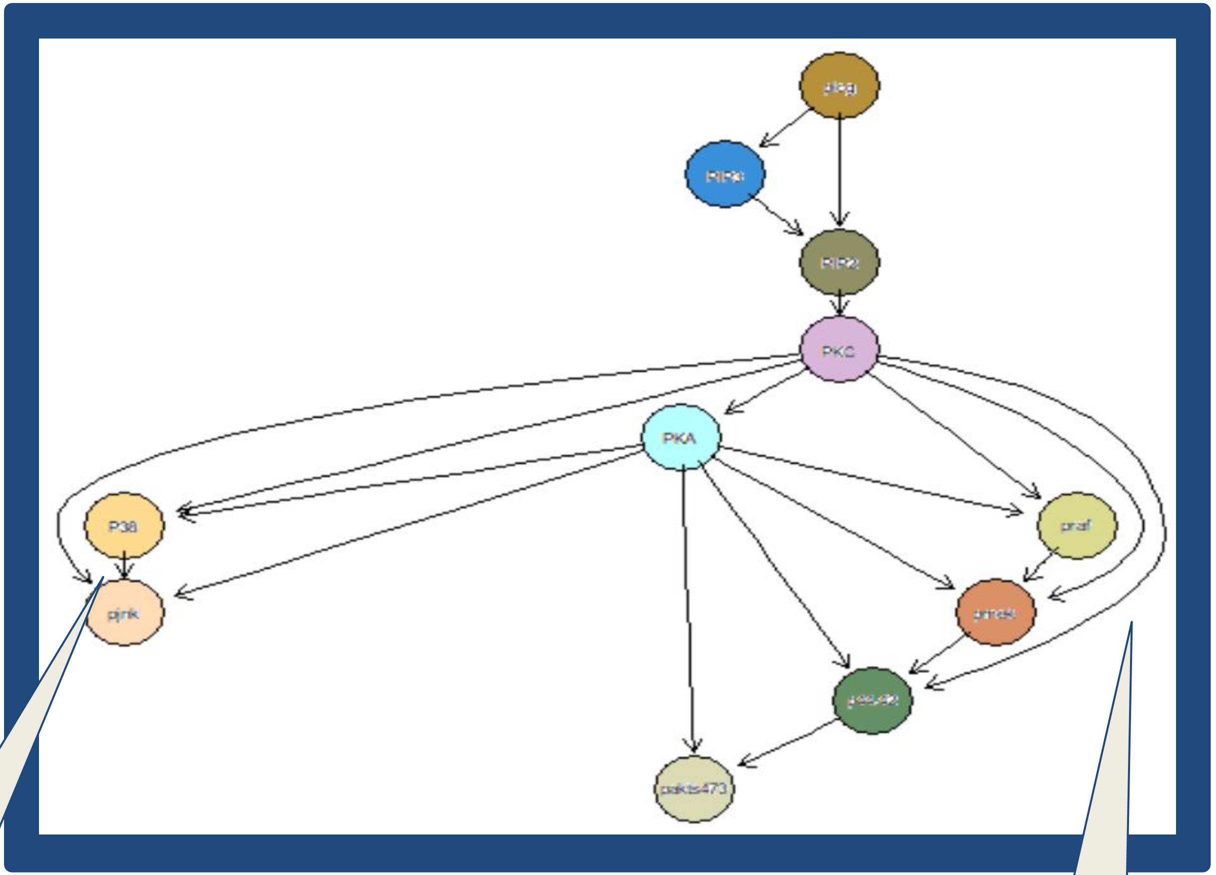
Networks Inferred

Sachs method reimplemented



TP: 17, FP: 8, FN: 3

Learn and Vote



TP: 18, FP: 2, FN: 2

P38 -> pjnk

PKC -> P44.42

1 PCViz: http://www.pathwaycommons.org/pcviz/
2 PubMed : https://www.ncbi.nlm.nih.gov/pubmed/

Empirical Study

Datasets

- **Lizards** — Real world dataset of 3 nodes
  Mutated Network: Intervened at 2 points
- **Asia** — Synthetic dataset of 8 nodes
  Mutilated 1: 1 observation and 1 interventional study
  Mutilated 2: 1 observation and 2 interventional study
- **Alarm** — Synthetic dataset of 37 nodes
  Mutilated 1: 3 observation and 6 interventional study
  Mutilated 2: 5 observation and 10 interventional study
- **Insurance** — Synthetic dataset of 27 nodes
  Mutilated 1: 1 observation and 5 interventional study
  Mutilated 2: 3 observation and 8 interventional study
- **GmInt** — Synthetic dataset of 8 nodes provided in the pcalg-R package

Popular Algorithms

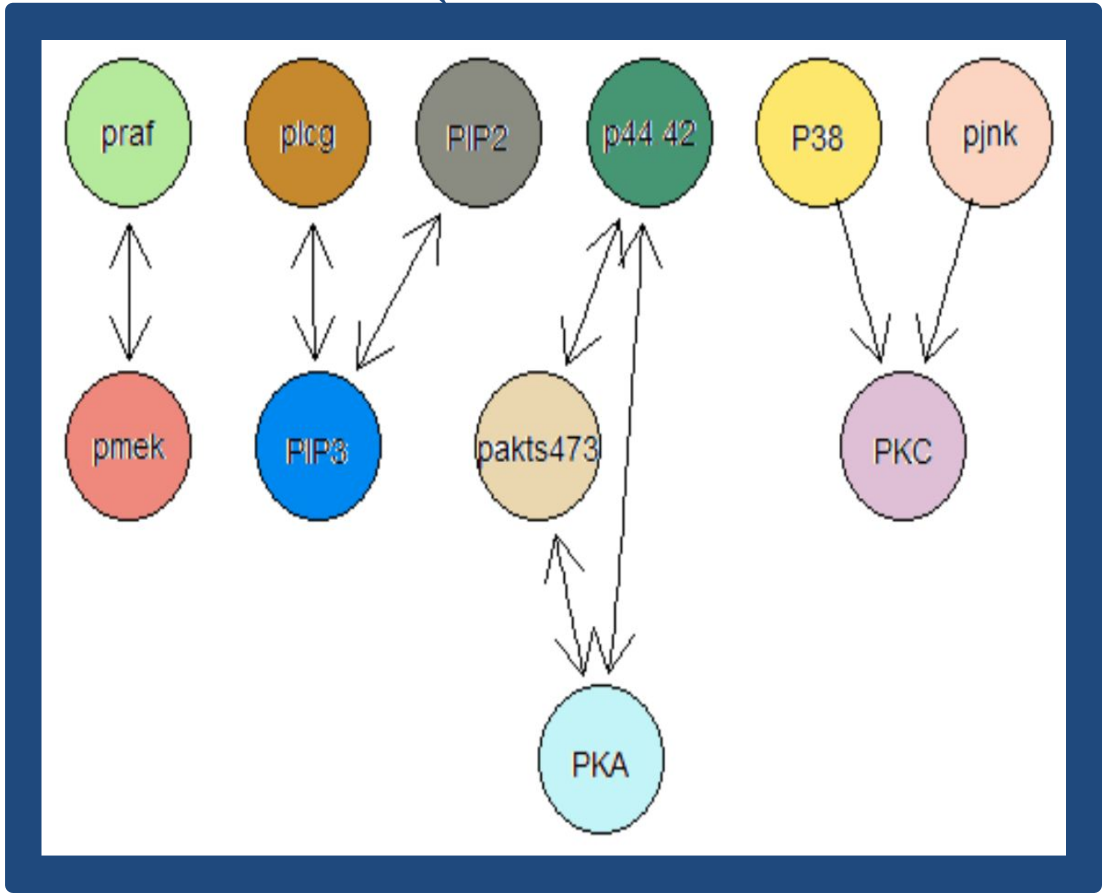
- **PC** — An algorithm for fast recovery of sparse causal graphs (P Spirtes, C Glymour, 1991)
- **GDS** — Greedy DAG Search proposed by (A. Hauser and P. Bühlmann, 2012)
- **GIES** — (Hauser and Bühlmann, 2012) extended the greedy equivalence search to include interventions.
- **ICP** — "Causal inference using invariant prediction: identification and confidence intervals" Peters, B¨uhlmann and Meinshausen, 2015
- **simy** — Dynamic programming approach of Silander and Myllymäki (2006)

Networks Inferred

Good with Observational data

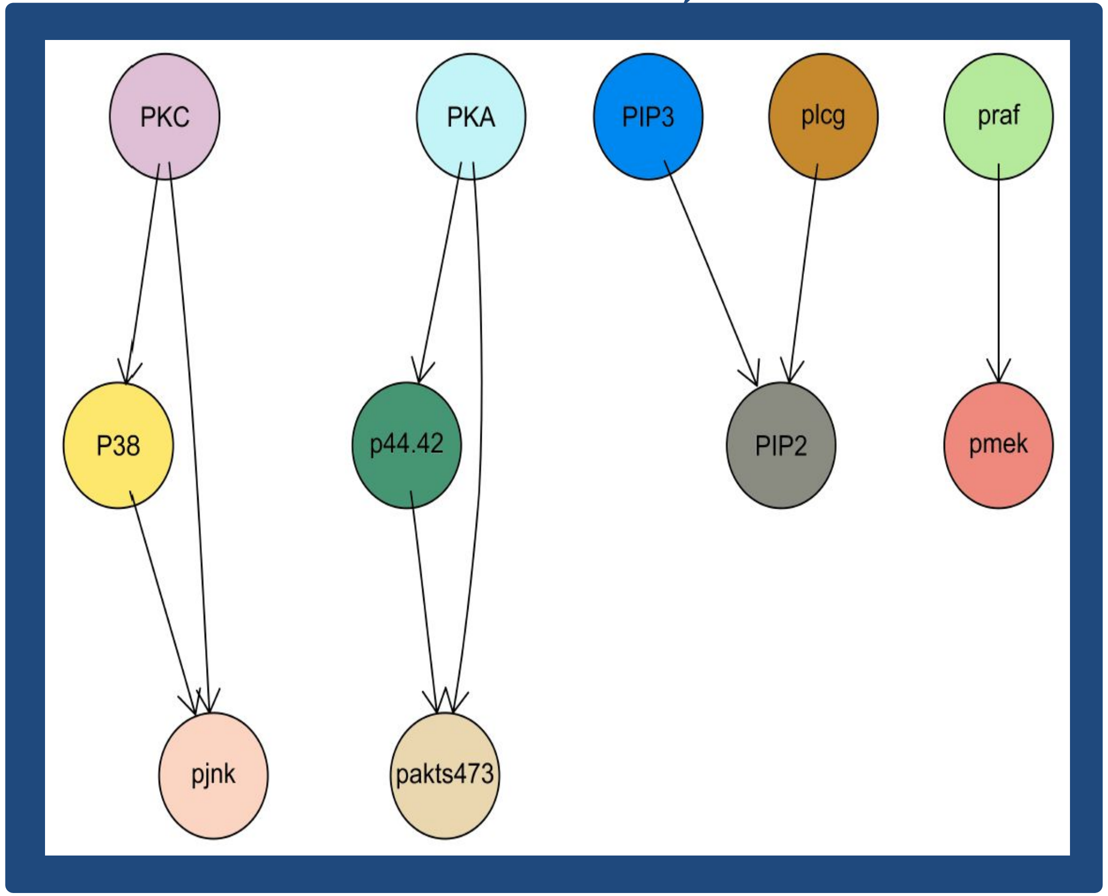
Too Strict, but Reliable !!!

PC



TP: 8, FP: 6, FN: 12

ICP

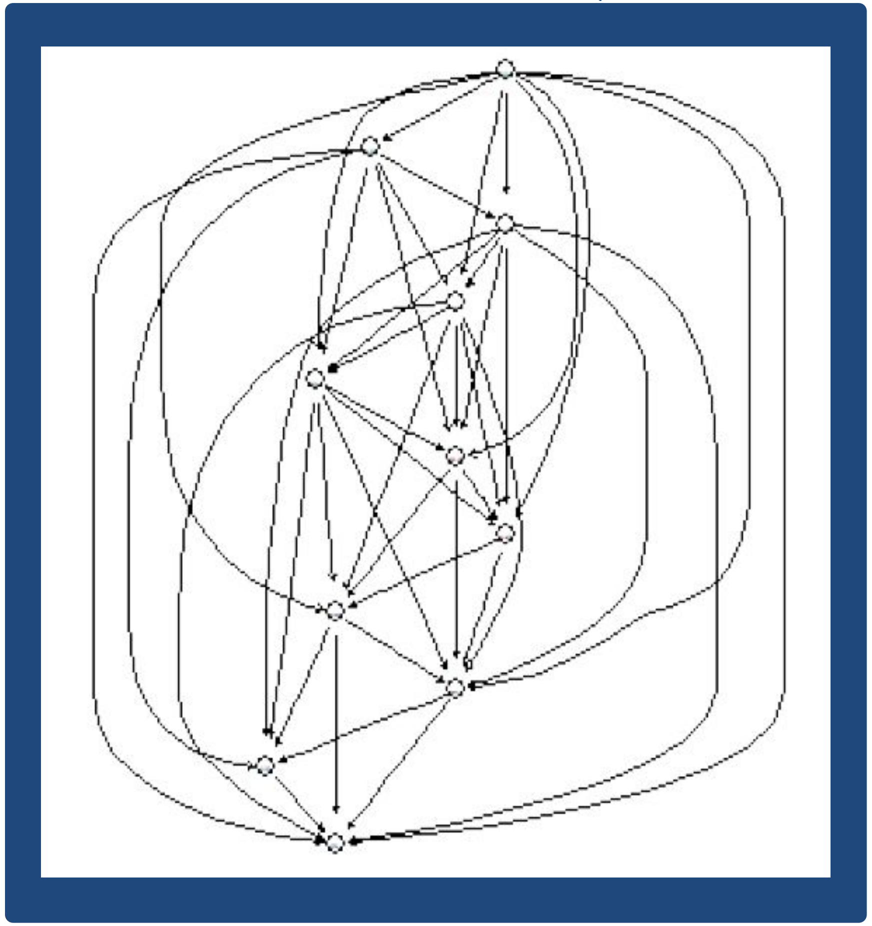


TP: 9, FP: 0, FN: 11

Networks Inferred

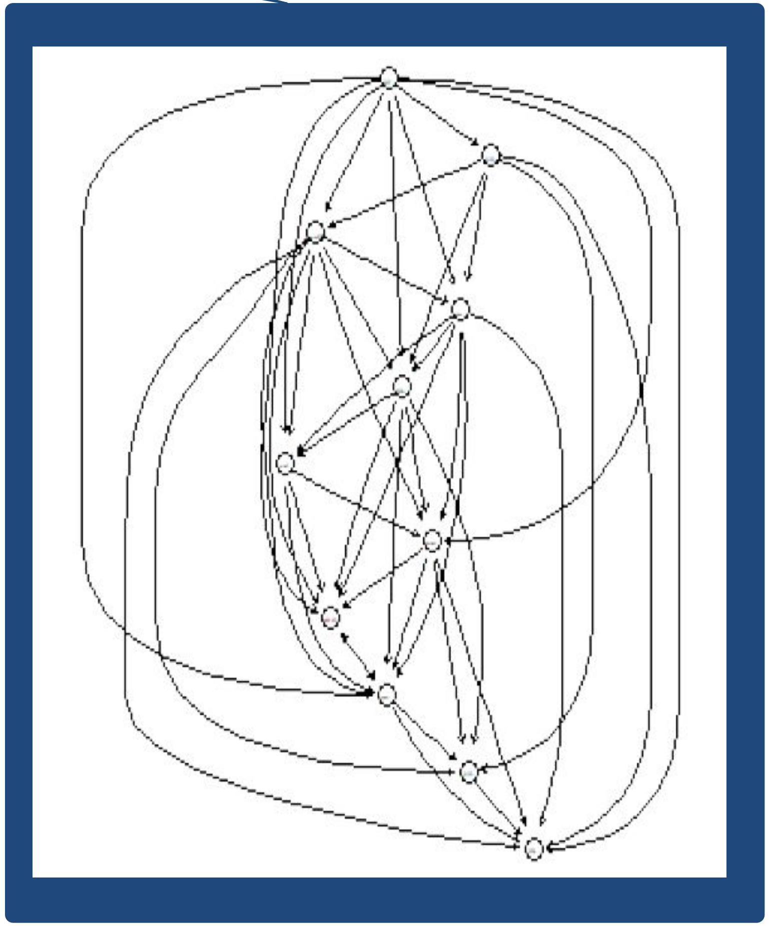
Greedy methods not good with uncertain interventions :-(

GDS



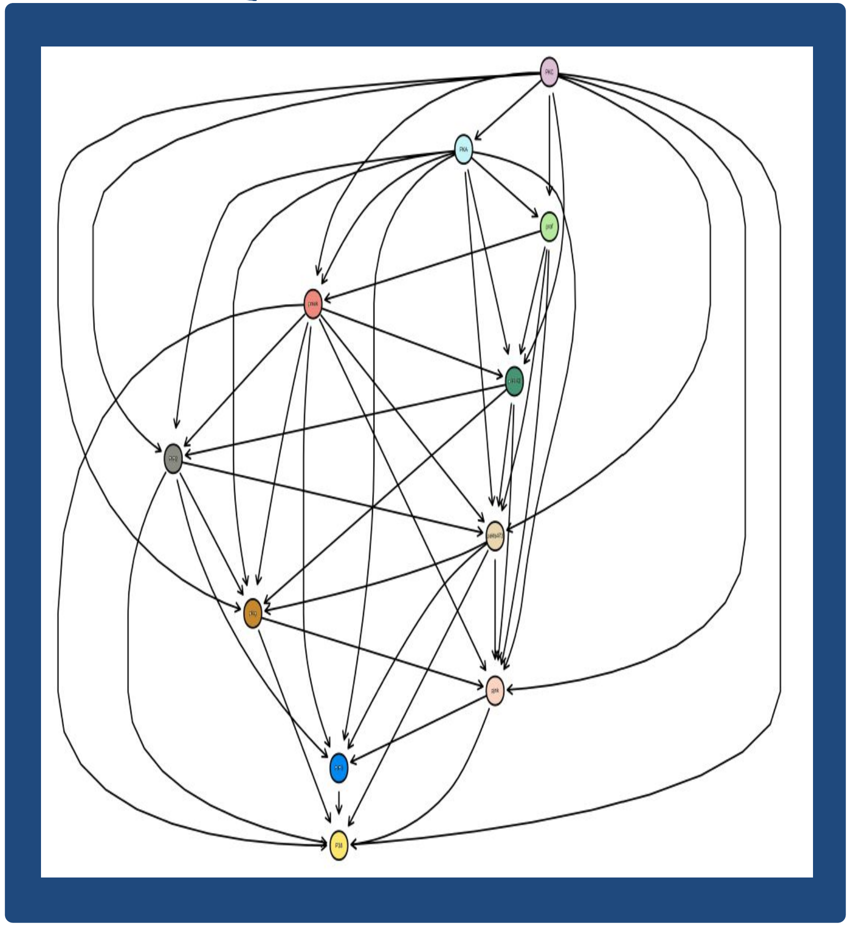
TP: 18, FP: 25, FN: 2

GIES



TP: 17, FP: 28, FN: 3

simy

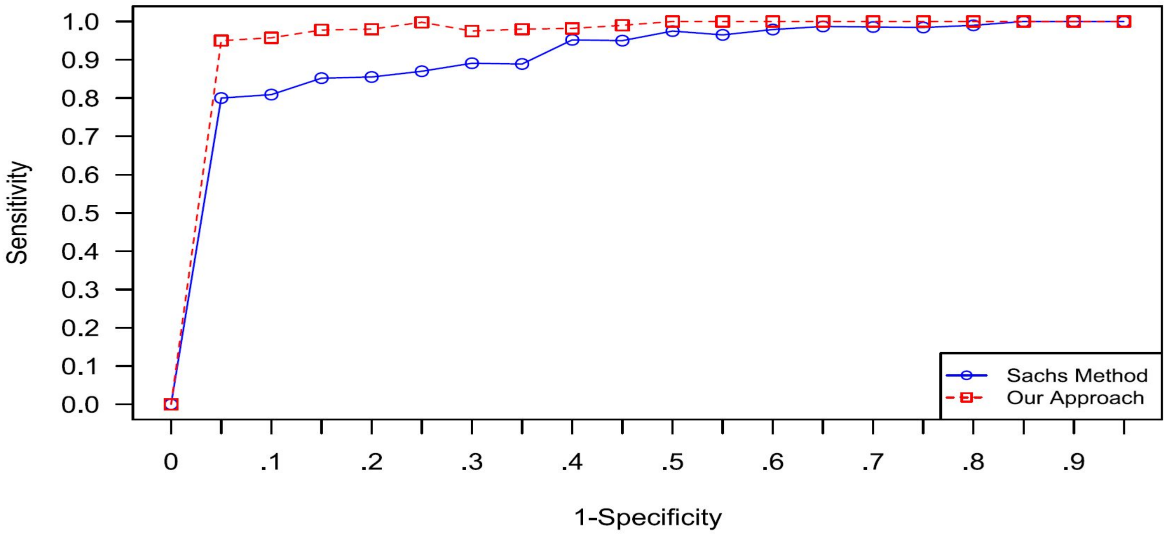


TP: 19, FP: 26, FN: 1

Comparative Benchmark

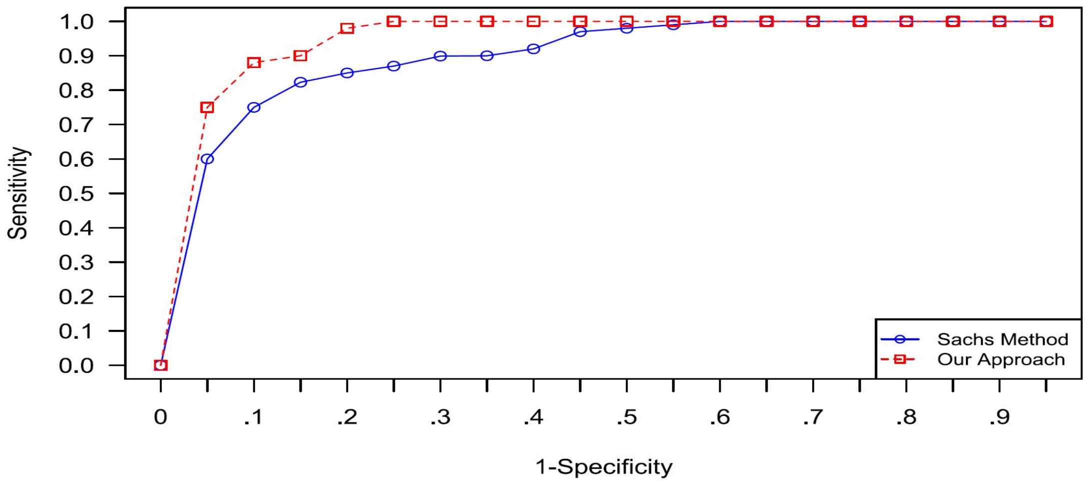
Dataset	*Metric*	PC	GDS	GIES	ICP	simy	Sachs et al	Learn and Vote
		Causal Discovery Algorithms						
Flow Cytometry	Precision	0.5714	0.4186	0.377	**1**	0.4222	0.68	0.89
	Recall	0.4	**0.9**	0.85	0.45	0.95	0.85	0.89
	F1 score	0.47	0.572	0.522	0.62	0.584	0.7558	**0.89**
Lizards	Precision	**1**	**1**	**1**	0	**1**	**1**	**1**
	Recall	**1**	**1**	**1**	0	**1**	0.5	0.5
	F1 score	**1**	**1**	**1**	0	**1**	0.667	0.667
Asia_mut1	Precision	**1**	0.625	0.625	**1**	0.31578	0.77	**1**
	Recall	0.75	0.625	0.625	0.5	0.75	**0.875**	0.75
	F1 score	**0.857**	0.625	0.625	0.666	0.4444	0.8237	**0.857**
Asia_mut2	Precision	**1**	0.85714	0.85714	**1**	0.3043	0.666	**1**
	Recall	0.75	0.75	0.75	0.5	**0.875**	0.75	0.75
	F1 score	**0.857**	0.8	0.8	0.666	0.4928	0.7058	**0.857**
gmInt	Precision	0.75	0.889	0.889	**1**	0.889	0.8571	**1**
	Recall	0.75	**1**	**1**	0.375	**1**	0.75	0.75
	F1 score	0.75	**0.94**	**0.94**	0.5454	**0.94**	0.8	0.857
Alarm_mut1	Precision	0.666	0.25	0.26	**0.7**	n/a	0.625	0.564
	Recall	0.434	0.217	0.26	0.26	n/a	**0.4464**	0.4
	F1 score	**0.526**	0.2325	0.26	0.38	n/a	0.52	0.468
Alarm_mut2	Precision	0.666	0.411	0.5128	0.6	n/a	0.725	**0.769**
	Recall	0.434	0.456	0.434	0.21	n/a	0.63	**0.642**
	F1 score	0.526	0.432	0.47	0.3115	n/a	0.675	**0.7**
Insurance_mut1	Precision	0.7143	0.36	0.3617	0.7	n/a	**0.857**	0.8
	Recall	0.288	0.3461	0.327	0.25	n/a	**0.577**	0.538
	F1 score	0.4107	0.352	0.3435	0.368	n/a	**0.689**	0.643
Insurance_mut2	Precision	**0.7143**	0.355	0.366	0.64	n/a	0.676	0.6857
	Recall	0.288	0.423	0.423	0.21	n/a	0.4423	**0.4615**
	F1 score	0.4107	0.386	0.392	0.316	n/a	0.535	**0.5517**

Performance Analysis with respect to ROC curve

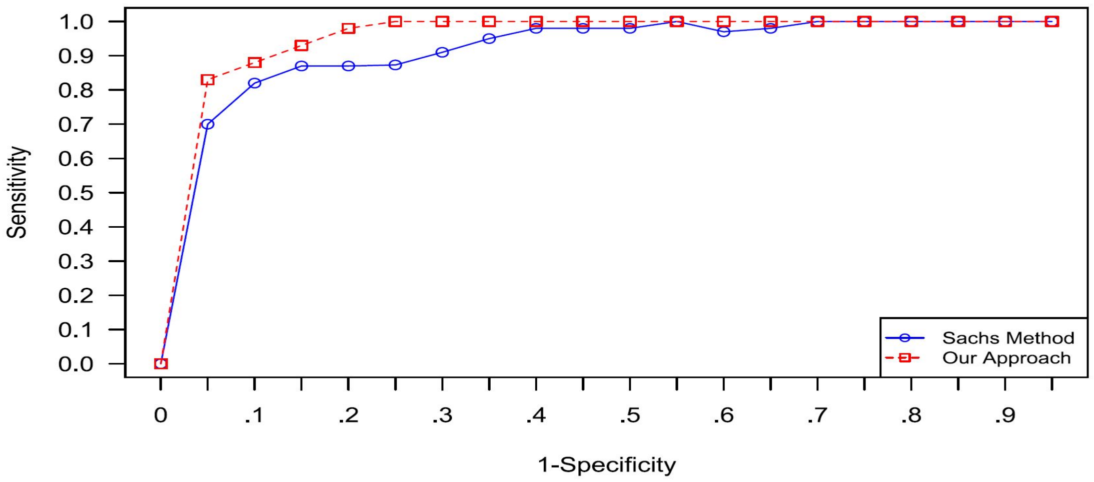


Flow Cytometry data

- We plot the ROC curve for the Flow cytometry data
- Compared the baseline method to Learn and vote
- Area under curve is greater across all thresholds
- Performance improves for “Learn and Vote”
- Intervening at informative targets improves performance

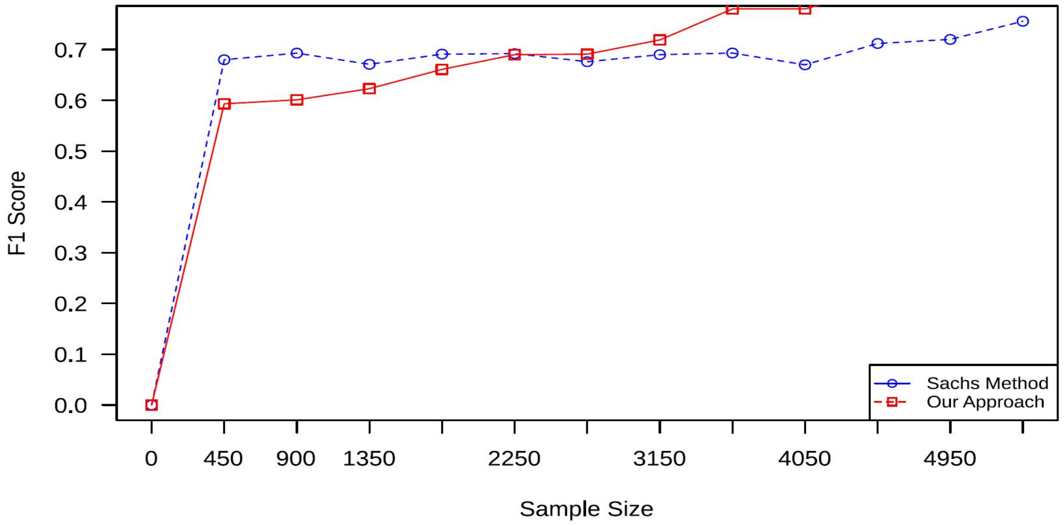


Asia_mut 1



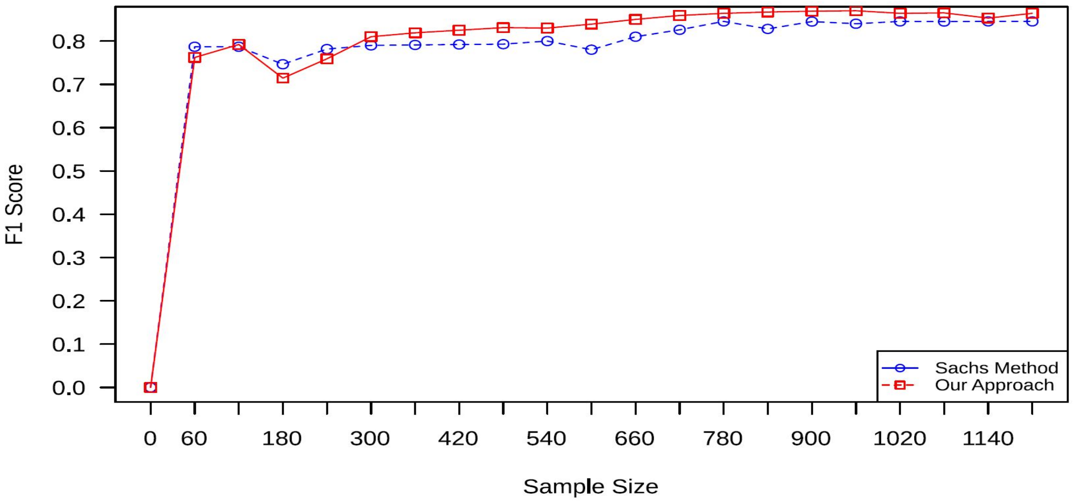
Asia_mut 2

Performance Analysis with respect to Sample size

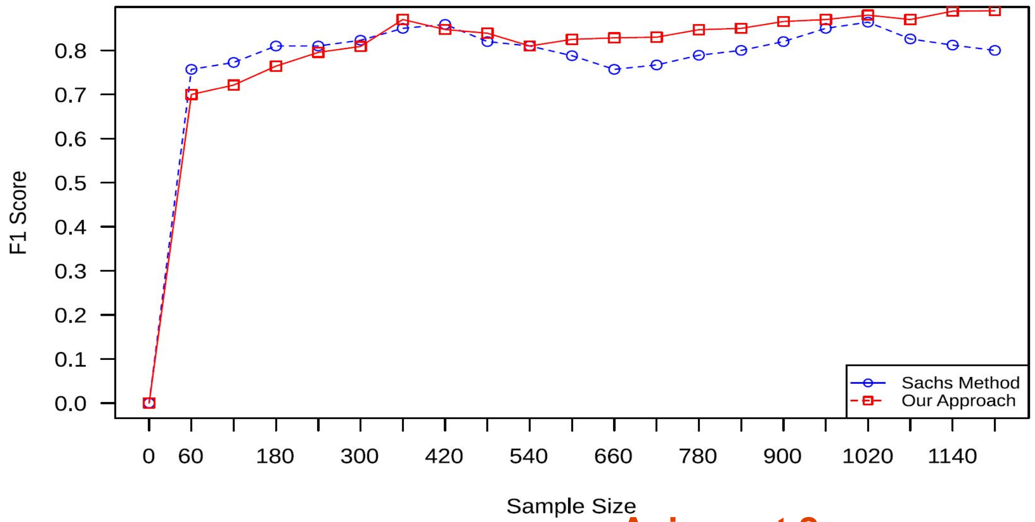


Flow Cytometry data

- We plot the F1 score against various sample sizes
- Equal number of sample per experiment
- Compared the baseline method to Learn and vote
- Performance improves for “Learn and Vote” with larger sample size
- Pooling is a better than “Learn and Vote” with small sample size per experiments



Asia_mut 1



Asia_mut 2

Limitations

- Approach is preliminary, needs more theoretical backing
- Does not work if dataset is too small
- We need equal samples of data per experiments

Future Directions

- Learning better given only Observations
- Categorizing which interventions are more informative
- Detecting presence of Latent variables

Thank You!

Any Questions?