ADVANCES IN INTEGRATIVE CAUSAL ANALYSIS

Ioannis Tsamardinos
Associate Professor, CSD,
University of Crete
UAI 2015 Causality Workshop

CREDIT AND ACKNOWLEDGEMENTS

Theory and Algorithms

Sofia Triantafillou

Vincenzo Lagani

Giorgos Borboudakis

Mass-Cytometry Application

Karolinska Insitutet

Angelika Schmidt

Jesper Tegner

David Gomez-Cabrero

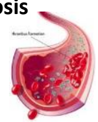
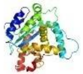
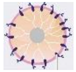
Funding and Hosting

Institute of Computer Science, FORTH

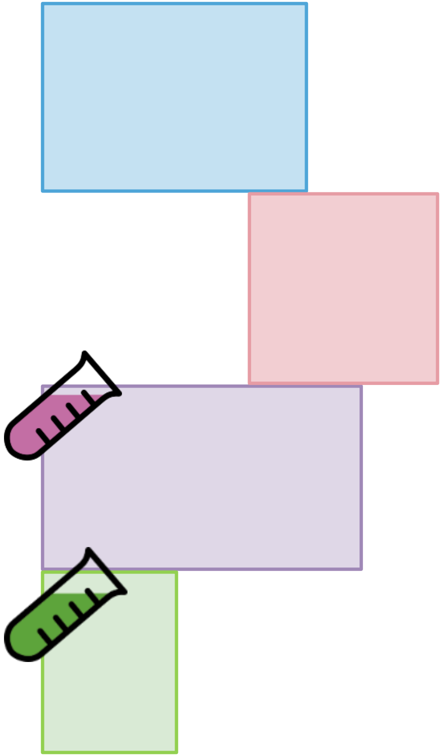
Computer Science Department, University of Crete

EU STATegra, EU ERC Causalpath, GSRT ARISTEIA II Epilogeas

HETEROGENEOUS DATA SETS MEASURING THE SAME SYSTEM UNDER STUDY

Variables / Study	Thrombosis 	Contraceptives 	Protein C 	Breast Cancer 	Protein Y 	Protein Z 
1 observational data	Yes	No	10.5	Yes	-	-
	No	Yes	5.3	No	-	-
					-	-
	No	Yes	0.01	No	-	-
2 observational data	-	-	-	Yes	0.03	9.3
	-	-	-			
	-	-	-	No	3.4	22.2
3 experimental data	No	No	0 (Control)	No	3.4	-
	Yes	No	0 (Control)	Yes	2.2	-
					-	-
	Yes	Yes	5.0 (Treat.)	Yes	7.1	-
	No	Yes	5.0 (Treat.)	No	8.9	-
4 experimental data	No	No (Ctrl)	-	-	-	-
	No	No (Ctrl)	-	-	-	-
			-	-	-	-
	Yes	Yes(Treat)	-	-	-	-

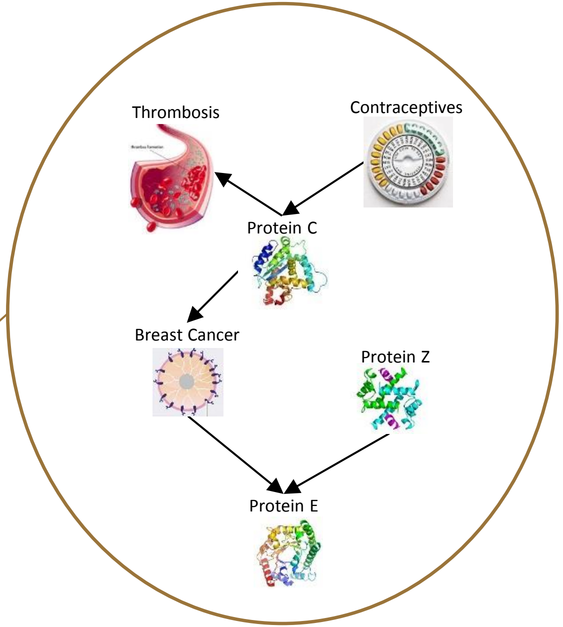
INTEGRATIVE CAUSAL ANALYSIS



Data can not be pulled together:

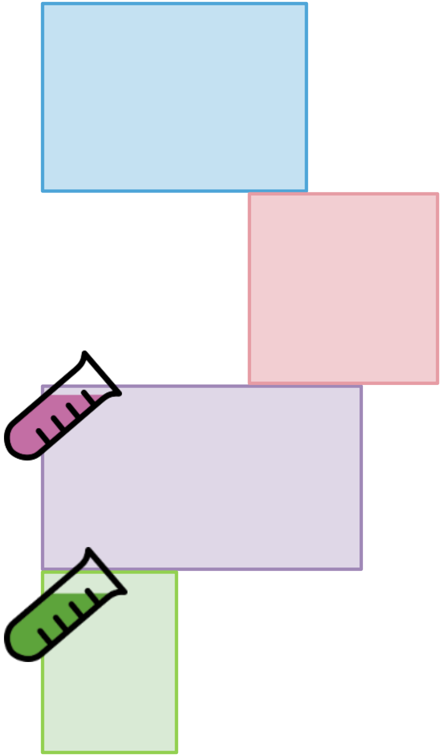
Missing variables cannot be treated as missing values.

They come from **different experimental conditions (different distributions).**

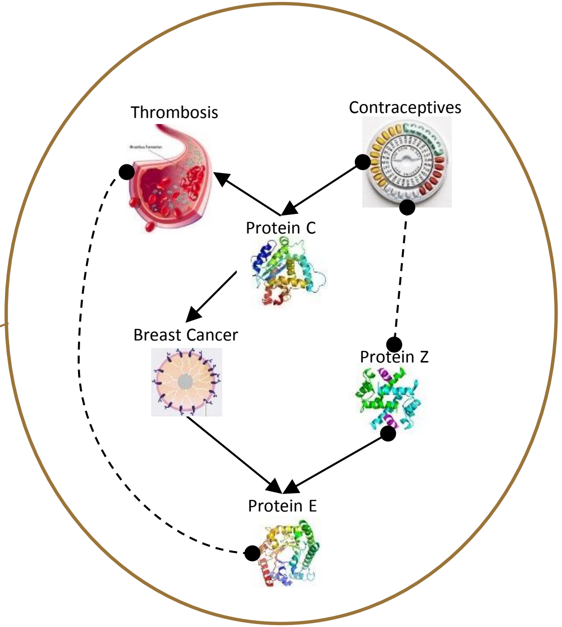


Data come from the **same causal mechanism**.

INTEGRATIVE CAUSAL ANALYSIS

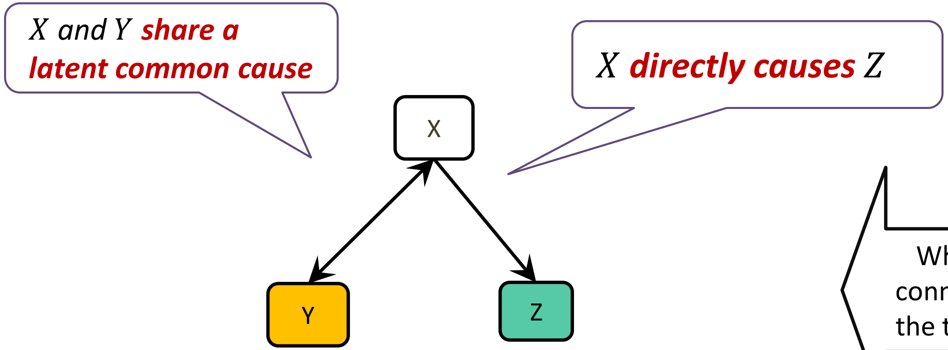


Identify the **causal graphs** that **simultaneously fit** all data.



CAUSAL MODELS

Semi Markov Causal Graph $\mathcal{G}$



What connects the two?

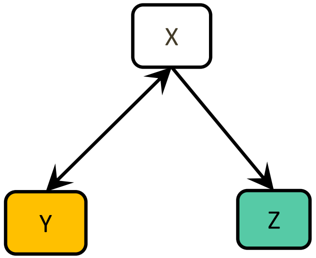
JPD $\mathcal{P}$

		Z	
X	Y	Yes	No
Yes	Yes	0,01	0,04
Yes	No	0,01	0,04
No	Yes	0,000045	0,044955
No	No	0,000855	0,854145

- Edges represent direct causal relations.
- Both edges allowed for a single pair of variables.
- No directed cycles (no causal feedback).

- $P(Y|\boldsymbol{V},Z) = P(Y|\boldsymbol{V})$ or Y is **independent** of Z **given** $\boldsymbol{V}$: $\boldsymbol{Ind}(Y,Z \mid \boldsymbol{V})$
- Otherwise : $\boldsymbol{Dep}(Y,Z \mid \boldsymbol{V})$
- The set of conditional independencies entailed by the JPD is called the **independence model** $\boldsymbol{J}$

CAUSAL ASSUMPTIONS



$$Ind(Y, Z \mid X)$$

$$Dep(Y, Z \mid \emptyset)$$
$$Dep(X, Z \mid \emptyset)$$
$$Dep(X, Z \mid Y)$$
$$Dep(Y, X \mid \emptyset)$$
$$Dep(Y, X \mid Z)$$

Causal Markov Assumption:
Every variable is independent of its **non-effects** (non-descendants in the graph) given its **direct causes** (parents).

Causal Faithfulness Assumption:
Independences stem **only** from the causal structure, **not the parameterization** of the distribution.

All independencies in $\mathcal{J}$ can be identified in $\mathcal{G}$ using the graphical criterion of **m-separation.**

m-SEPARATION

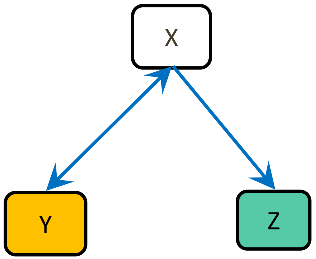
A path $X_1, \dots, X_n$ between X_1 and X_n is **m-connecting given V** if for every triple $\langle X_{i-1}, X_i, X_{i+1} \rangle$ on the path:

- If $X_{i-1} \ast\!\!\rightarrow X_i \leftarrow\!\!\ast X_{i+1}$,
  X_i or one of its descendants $\in \boldsymbol{V}$
- Otherwise, $X_i \notin \boldsymbol{V}$

m-connecting path **=>** information flow **=> dependence**

No **m-connecting** path **=>** no information flow **=> independence** (m-separation)

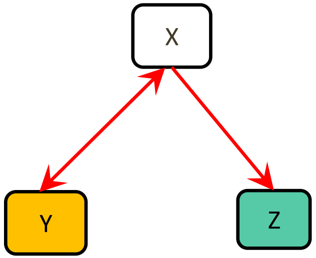
m-SEPARATION



Y $\leftrightarrow$ X $\longrightarrow$ Z

is m-connecting given $\emptyset$

$\Leftrightarrow Dep(Y, Z|\emptyset)$

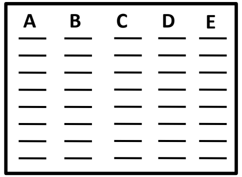


Y $\leftrightarrow$ X $\longrightarrow$ Z

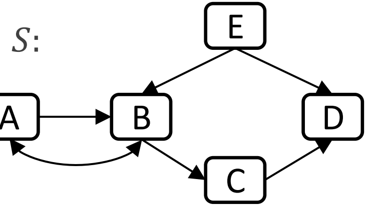
is NOT m-connecting given X

$\Leftrightarrow Ind(Y, Z|X)$

REVERSE ENGINEERING CAUSAL MODELS



$Dep(A, D|\emptyset)_{D_2}$
$Dep(A, D|B)_{D_2}$
$Dep(A, D|E)_{D_2}$
$Dep(A, D|C)_{D_2}$
...
$Ind(A, E|\emptyset)_{D_2}$



$\exists$ *m-connecting path from A to D given* $\emptyset$ *in S*
$\exists$ *m-connecting path from A to D given B in S*
⋮
$\exists$ *m-connecting path from A to B given* $\emptyset$ *in S*
⋮
$\nexists$ *m-connecting path from A to E given* $\emptyset$ *in S*

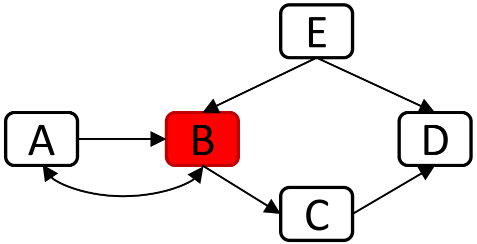
Dataset D_i measuring a set of variables

Independence model J_i

Path constraints on the underlying causal graph

INTERVENTIONS / MANIPULATIONS IN CAUSAL MODELS

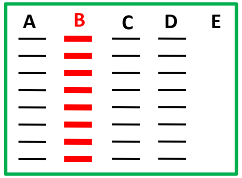
Graph (SMCG) S



Values of B are **set** solely **by the manipulation procedure.**

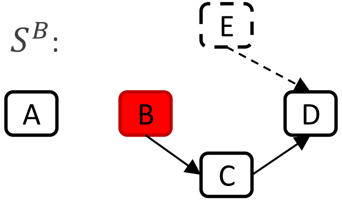
If you know direct causal relations, **remove all edges into the manipulated variable.**

INCA: OVERLAPPING VARIABLES, INTERVENTIONS



$Ind(A,D|\emptyset)_{D_2}$
$Ind(A,D|B)_{D_2}$
$Ind(A,D|C)_{D_2}$
$Ind(A,D|B,C)_{D_2}$
$Dep(B,C|\emptyset)_{D_2}$
...

S^B:



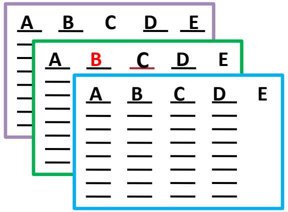
$\nexists$ *m-connecting path from A to D given* $\emptyset$ *in* S^B
$\nexists$ *m-connecting path from A to D given B in* S^B
$\vdots$
$\nexists$ *m-connecting path from A to D given* B, *C in* S^B
$\vdots$
$\exists$ *m-connecting path from B to C given* $\emptyset$ *in* S^B

Dataset D_i measuring a subset of variables, some of which are manipulated

Manipulated independence model J_i

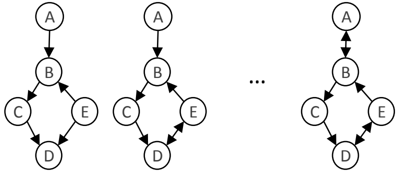
Path constraints on the **manipulated** underlying causal graph

PROBLEM DEFINITION



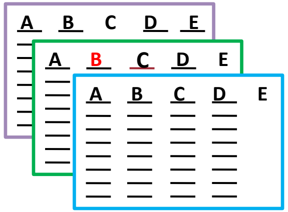
Data sets $D_1, \dots, D_n$ measuring overlapping variables under different manipulations

Causal Markov Assumption
Causal Faithfulness Assumption

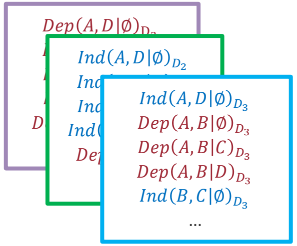


Graphs *S* that simultaneously fit all data

PROPOSED APPROACH



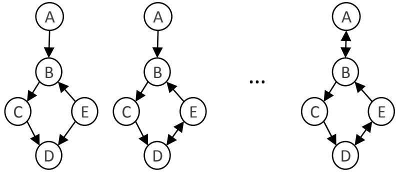
Data sets $D_1, \dots, D_n$



Independence models $J_1, \dots, J_n$

$\exists$ *m-connecting path from A to D given* $\emptyset$ *in* S^{I_1}
$\exists$ *m-connecting path from A to D given B in* S^{I_1}
$\vdots$
$\exists$ *m-connecting path from A to B given* $\emptyset$ *in* S^{I_2}
$\vdots$
$\nexists$ *m-connecting path from B to C given* $\emptyset$ *in* S^{I_n}

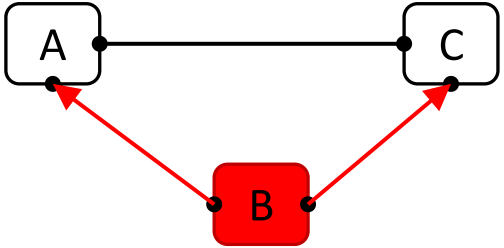
path constraints in $S_1, \dots, S_n$



Graphs S that simultaneously fit all data

CONSTRAINTS AS BOOLEAN FORMULAE

- Suppose you know nothing about the structure S of the three variables.



- In a data set where B is manipulated, $\mathrm{Ind}(\mathrm{A}, \mathrm{C}|\emptyset)$
- In path terms:

  $\nexists$ m-connecting path between A and C given $\emptyset$ in S^B

A-C does not exist

$\neg edge(A, C)$

AND

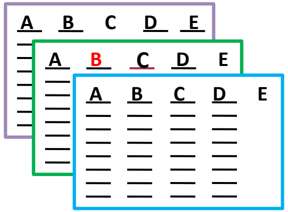
A ←B→C does not exist

$[\neg edge(A, B)$
$\vee \neg tail(A, B)$
$\vee \neg edge(B, \mathrm{C})$
$\vee \neg tail(C, B)]$

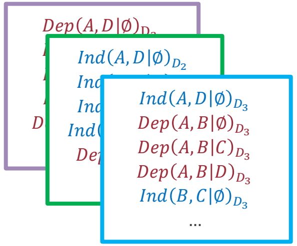
In SATisfiability terms:

$$\neg edge(A, C) \wedge$$
$$[\neg edge(A, B) \vee \neg tail(A, B) \vee \neg edge(B, \mathrm{C}) \vee \neg tail(C, B)]$$

PROPOSED APPROACH



Data sets $D_1, \dots, D_n$



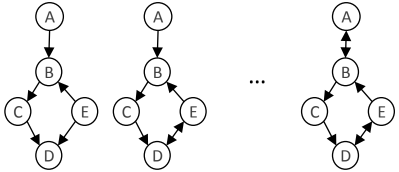
Independence models $J_1, \dots, J_n$

$\exists$ *m-connecting path from A to D given* $\emptyset$ *in* S^{I_1}
$\exists$ *m-connecting path from A to D given B in* S^{I_1}
$\vdots$
$\exists$ *m-connecting path from A to B given* $\emptyset$ *in* S^{I_2}
$\vdots$
$\nexists$ *m-connecting path from B to C given* $\emptyset$ *in* S^{I_n}

path constraints in $S_1, \dots, S_n$

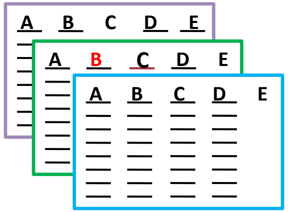
$[edge(A\text{-}D) \vee (\, edge(A\text{-}C)$
$\wedge\, edge(C\text{-}D) \wedge tail(C, A) \dots] \wedge$
$\vdots$
$[edge(A\text{-}D) \vee (edge(A\text{-}C) \dots] \wedge$
$\vdots$
$[\neg\, edge(A\text{-}D) \wedge (\neg\, edge(A\text{-}C) \dots)]$

Formula Φ encoding path constraints in S



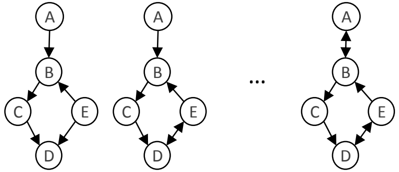
Graphs *S* that simultaneously fit all data

SCALING UP



Data sets $D_1, \ldots, D_n$

How many constraints?



Graphs S that simultaneously fit all data

$Dep(A,D|\emptyset)_{D_1}$

$Ind(A,D|\emptyset)_{D_2}$

$Ind(A,D|\emptyset)_{D_3}$
$Dep(A,B|\emptyset)_{D_3}$
$Dep(A,B|C)_{D_3}$
$Dep(A,B|D)_{D_3}$
$Ind(B,C|\emptyset)_{D_3}$
...

Independence models $J_1, \ldots, J_n$

$\exists$ *m-connecting path from A to D given* $\emptyset$ *in* S^{I_1}
$\exists$ *m-connecting path from A to D given B in* S^{I_1}
⋮
$\exists$ *m-connecting path from A to B given* $\emptyset$ *in* S^{I_2}
⋮
$\nexists$ *m-connecting path from B to C given* $\emptyset$ *in* S^{I_n}

path constraints in $S_1, \ldots, S_n$

[*edge(A-D)* ∨ (*edge(A-C)*
∧ *edge(C-D)*∧ $tail(C, A)$ …] ∧
⋮
[*edge(A-D)* ∨ (*edge(A-C)* …] ∧
⋮
[¬ *edge(A-D)* ∧ (¬ *edge(A-C)* …)]

Formula Φ encoding path constraints in S

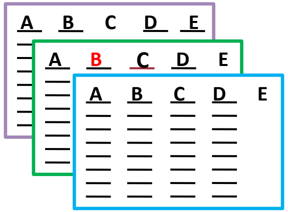
SCALING UP

How many constraints?

- Independence model: 2^n conditional (in)dependencies.
  - You only need a subset that entail all others ($\binom{n}{2}$).
  - FCI (Spirtes, Glymour, Scheines 2000, Zhang 2008) finds this subset of conditional (in)dependencies
    - also outputs a graph that summarizes the characteristics of every possible causal structure that entails them.

  - Use FCI and only convert
    - (non) adjacencies to $(\nexists)\exists$ inducing path.
    - Colliders with order to $(\nexists)\exists$ inducing and directed paths.

SCALING UP



Data sets $D_1, \dots, D_n$

$Dep(A,D|\emptyset)_{D_1}$

$Ind(A,D|\emptyset)_{D_2}$

$Ind(A,D|\emptyset)_{D_3}$
$Dep(A,B|\emptyset)_{D_3}$
$Dep(A,B|C)_{D_3}$
$Dep(A,B|D)_{D_3}$
$Ind(B,C|\emptyset)_{D_3}$
…

Independence models $J_1, \dots, J_n$

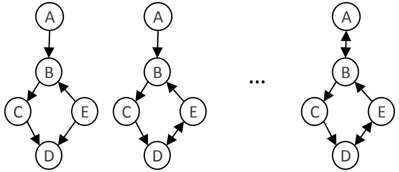
How many paths?

$\exists$ *m-connecting path from A to D given* $\emptyset$ *in* S^{I_1}
$\exists$ *m-connecting path from A to D given B in* S^{I_1}
⋮
$\exists$ *m-connecting path from A to B given* $\emptyset$ *in* S^{I_2}
⋮
$\nexists$ *m-connecting path from B to C given* $\emptyset$ *in* S^{I_n}

path constraints in $S_1, \dots, S_n$

[*edge(A-D)* ∨ (*edge(A-C)* ∧ *edge(C-D)*∧ $tail(C, A)$ …] ∧
⋮
[*edge(A-D)* ∨ (*edge(A-C)* …] ∧
⋮
[¬ *edge(A-D)* ∧ (¬ *edge(A-C)* …)]

Formula Φ encoding path constraints in S



Graphs *S* that simultaneously fit all data

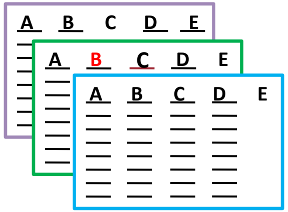
SCALING UP

How many paths?

Reduce the number of paths.

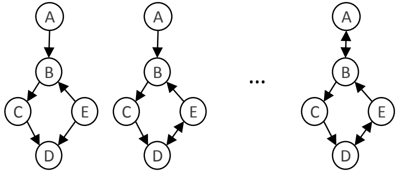
- Remove edges based on preprocessing.
- Restrict the path length.

SCALING UP



Data sets $D_1, \dots, D_n$

How many solutions?



Graphs S that simultaneously fit all data

$Dep(A,D|\emptyset)_{D_1}$

$Ind(A,D|\emptyset)_{D_2}$

$Ind(A,D|\emptyset)_{D_3}$
$Dep(A,B|\emptyset)_{D_3}$
$Dep(A,B|C)_{D_3}$
$Dep(A,B|D)_{D_3}$
$Ind(B,C|\emptyset)_{D_3}$
…

Independence models $J_1, \dots, J_n$

$\exists$ *m-connecting path from A to D given* $\emptyset$ *in* S^{I_1}
$\exists$ *m-connecting path from A to D given B in* S^{I_1}
$\vdots$
$\exists$ *m-connecting path from A to B given* $\emptyset$ *in* S^{I_2}
$\vdots$
$\nexists$ *m-connecting path from B to C given* $\emptyset$ *in* S^{I_n}

path constraints in $S_1, \dots, S_n$

[*edge(A-D)* ∨ (*edge(A-C)*
∧ *edge(C-D)*∧ $tail(C, A)$ …] ∧
$\vdots$
[*edge(A-D)* ∨ (*edge(A-C)* …] ∧
$\vdots$
[¬ *edge(A-D)* ∧ (¬ *edge(A-C)* …)]

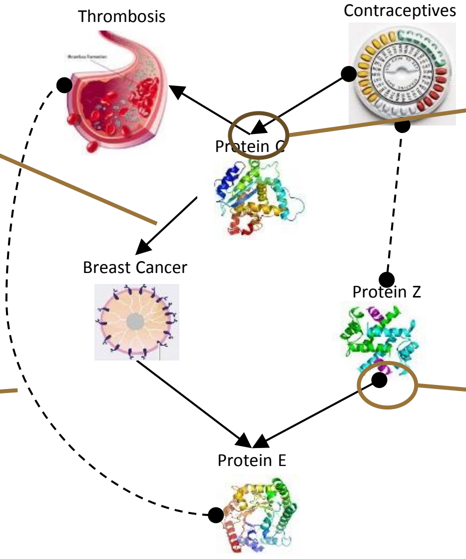
Formula Φ encoding path constraints in S

SCALING UP – QUERY-BASED LEARNING

How many solutions?

solid edges: present in **all** solutions

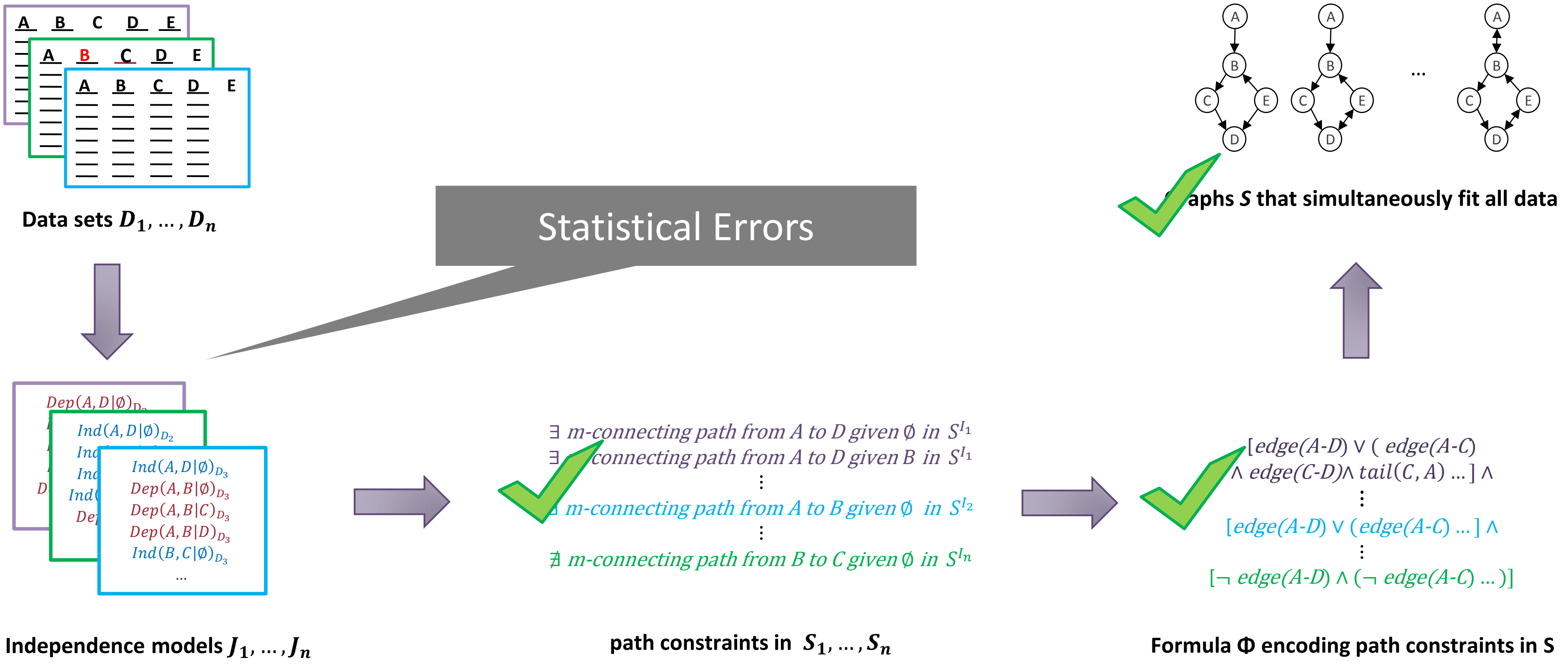
solid endpoints: same orientation in **all** solutions



dashed edges: present in **some** solutions

dashed endpoints: different orientation in **different** solutions

STATISTICAL ERRORS



STATISTICAL ERRORS

dependencies	independencies
▼	▼
set of constraints	set of constraints

Select non conflicting constraints!

- How?
- Assign confidence to every constraint.

p-VALUES TO PROBABILITIES

low p-values suggest **dependence**

high p-values suggest **independence**

p-values p coming from dependencies	p-values p coming from independencies

$H_1 : p \sim Beta(\xi, 1), \quad \xi \in (0, 1)$

$H_0 : \; p \sim Beta(1,1)$

$$f(p|\pi_o, \xi) = \pi_0 + (1 - \pi_0)\xi p^{\xi - 1}$$

The proportion of p-values coming from H_0

If you know $\widehat{\pi_0}$, $\hat{\xi}$ you can find posterior probabilities

$$P(H_0|p) = \frac{\frac{\widehat{\pi_0}}{(1-\widehat{\pi_0})\hat{\xi}p^{(1-\hat{\xi})}}}{1 + \frac{\widehat{\pi_0}}{(1-\widehat{\pi_0})\hat{\xi}p^{(1-\hat{\xi})}}}$$

$$P(H_1|p) = 1 - P(H_0|p)$$

ALGORITHM PROPeR

PROPeR

identify $\widehat{\pi_0}$:
(Storey and Tibshirani, 2003)

Identify $\hat{\xi}$:
Minimize negative log likelihood of
$f(p|\widehat{\pi_0}, \xi) = \widehat{\pi_0} + (1 - \widehat{\pi_0})\xi p^{\xi - 1}$

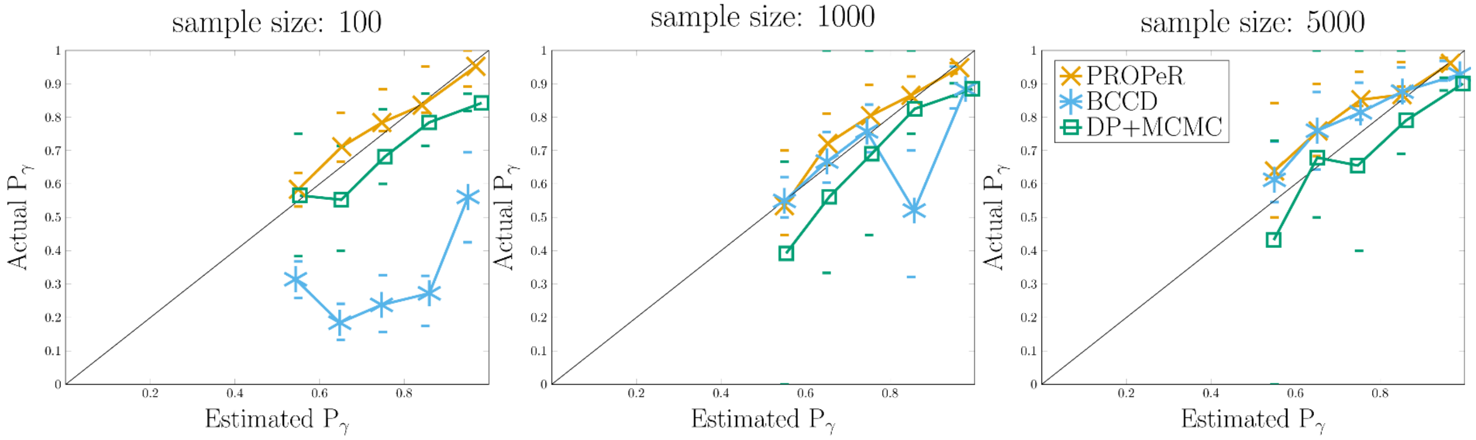
Estimate
$P(H_0|p), P(H_1|p)$

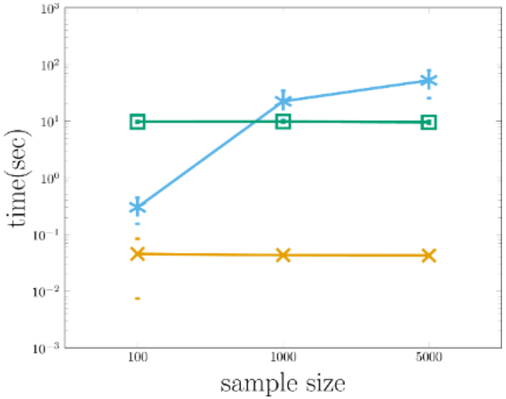
- Constraints correspond to adjacencies (or absence thereof), not (in) dependencies
  - adjacent(X, Y): The maximum p-value for X, Y was $< \alpha$
  - $\neg$ adjacent(X, Y): The maximum p-value for X, Y was $> \alpha$
- Maximum p-values may not follow a uniform distribution
  Samples (p-values) are not i.i.d.
- Did we cut too many corners?

Calibration of probability estimates compared to

- BCCD: Posterior probability of a feature is obtained by the weighted sum of the likelihoods of all networks with < 6 variables [Claassen and Heskes, 2012]
- DP+MCMC: Exact method, scales up to ~20 variables, MCMC priors [Eaton and Murphy, 2007]

Discrete variables





- PROPeR has no significant computational overhead
- PROPeR performs on par with more expensive Bayesian methods

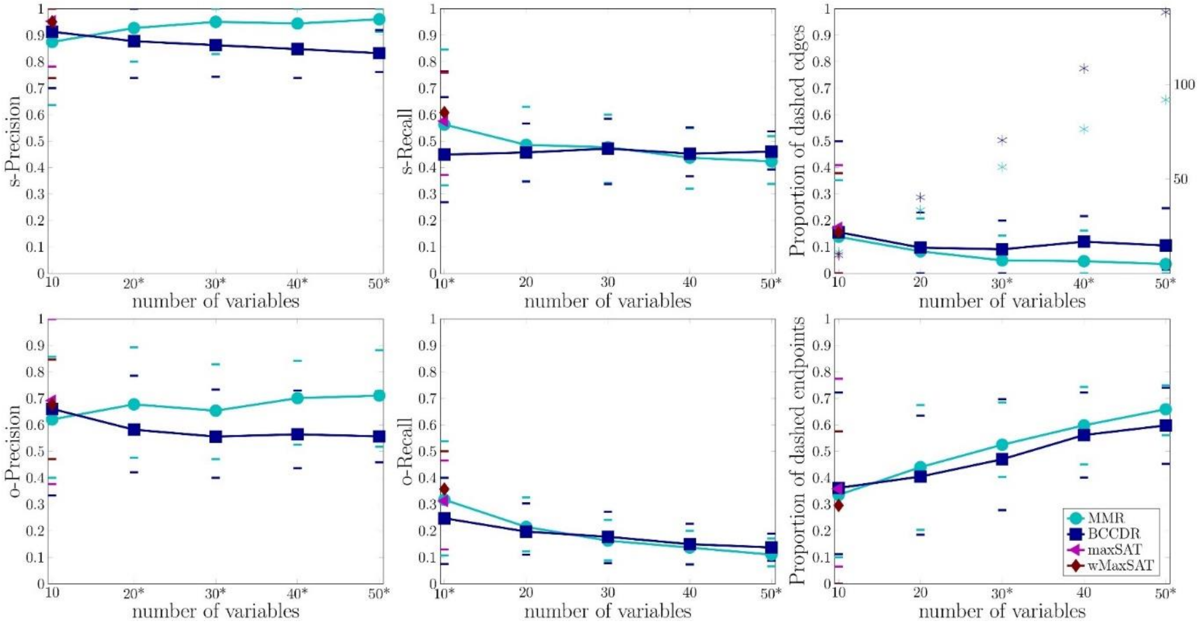
CONFLICT RESOLUTION STRATEGIES

P(constraint)	constraint
0.999	∃ m-connecting path from A to D given ∅ in S^{I_n}
0.998	∄ m-connecting path from A to D given ∅ in S^{I_1}
⋮	⋮
0.510	∃ m-connecting path from A to B given ∅ in S^{I_1}

- MMR: Estimate confidence using PROPeR, satisfy constraints in decreasing order of confidence.
- BCCDR: Estimate confidence using BCCD, satisfy constraints in decreasing order of confidence.
- maxSAT: satisfy maximum number of constraints.
- wMaxSAT: satisfy constraints with maximum sum of weights, where weights are PROPeR estimates.

CONFLICT RESOLUTION STRATEGIES

20 variables, 5 overlapping data sets, 100 samples each



- Greedy strategies perform similarly
- Max/ weighted max strategies do not scale up.
- All strategies are equally conservative.

PERFORMANCE EVALUATION

- ~2000 runs in simulated networks
- vs maximum path length, number of variables, sparseness of the ground truth networks, sample size, number of input data sets, proportion of non-overlapping variables.

- Performance improves with more data sets.
- Performance improves with more samples.
- Performance is better for sparser networks.
- COmbINE scales up to 100 variables.
- Maximum path length and number of non-overlapping variables do not influence COmbINE's performance.

APPLICATION ON MASS CYTOMETRY DATA

observed in some data set

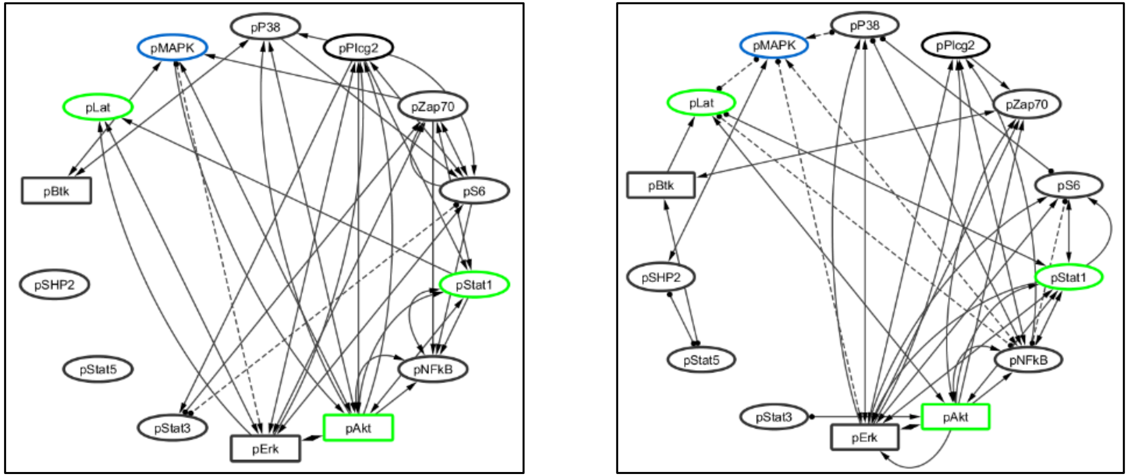
manipulated in some data set

not measured in D_1, D_2, D_3

not measured in D_4

cd4+ T-cells

cd8+ T-cells



Data set	Source	latent ($\mathbf{L_i}$):	manipulated($\mathbf{I_i}$)
$\mathbf{D_1}$	Bodenmiller et al. (2012)	pMAPK	pAkt
$\mathbf{D_2}$	Bodenmiller et al. (2012)	pMAPK	pBtk
$\mathbf{D_3}$	Bodenmiller et al. (2012)	pMAPK	pErk
$\mathbf{D_4}$	Bendall et al. (2011)	pAkt, pLat, pStat1	pErk

- 4 data sets.
- 3 different manipulations.
- 14 variables (phosphorylated proteins).
- 2 different cell populations.

INCA RULE

Dataset D_1

X	Y	Z	W

Independence model J_1

$Dep(X,Y|\emptyset)_{D_1}$
$Dep(X,W|Y)_{D_1}$
$Dep(X,Z|\emptyset)_{D_1}$
$\boldsymbol{Ind(X,W|Y)_{D_1}}$
$Dep(Y,W|\emptyset)_{D_1}$
$Dep(Y,W|X)_{D_1}$

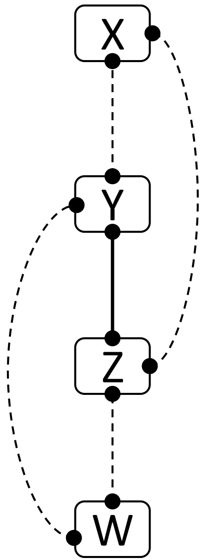
Dataset D_2

X	Y	Z	W

Independence model J_2

$Dep(X,W|\emptyset)_{D_2}$
$Dep(X,W|Z)_{D_2}$
$Dep(X,Z|\emptyset)_{D_2}$
$\boldsymbol{Ind(X,W|Z)_{D_2}}$
$Dep(Z,W|\emptyset)_{D_2}$
$Dep(Z,W|X)_{D_2}$

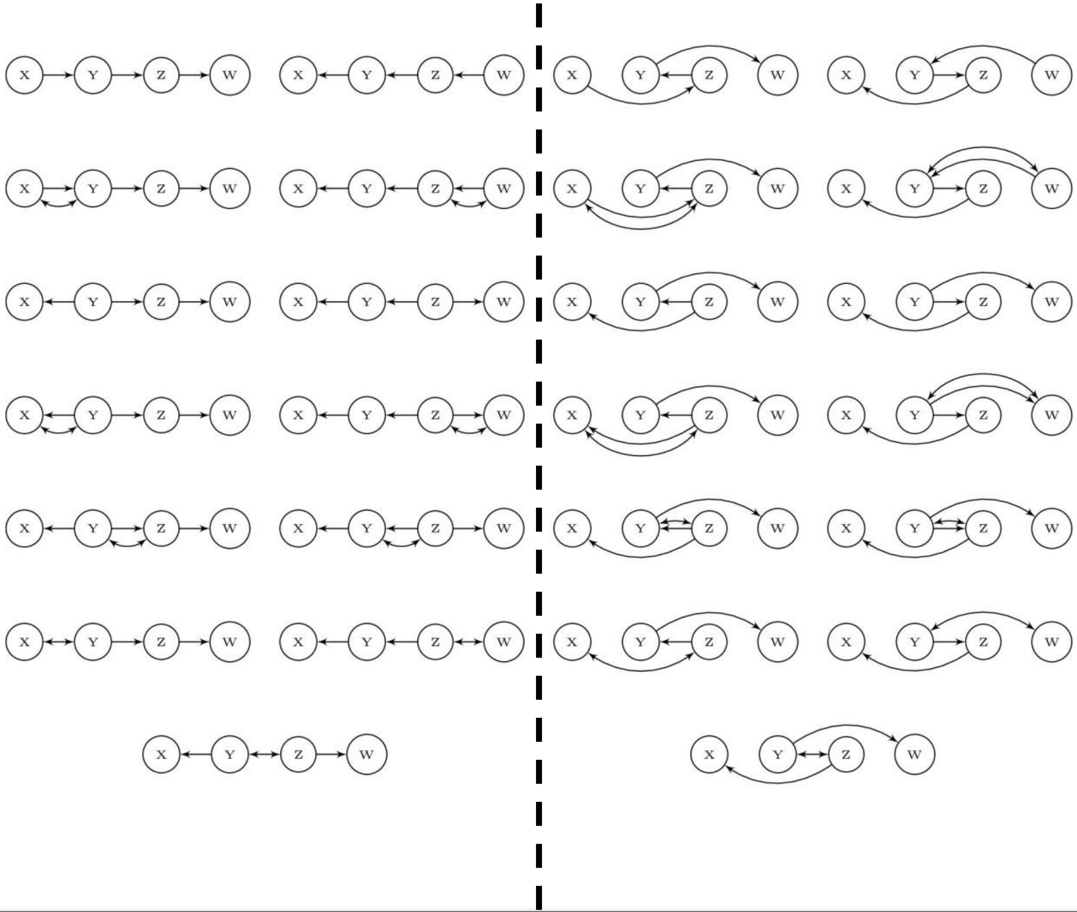
Predict X and Y are associated – even though they are never measured together!



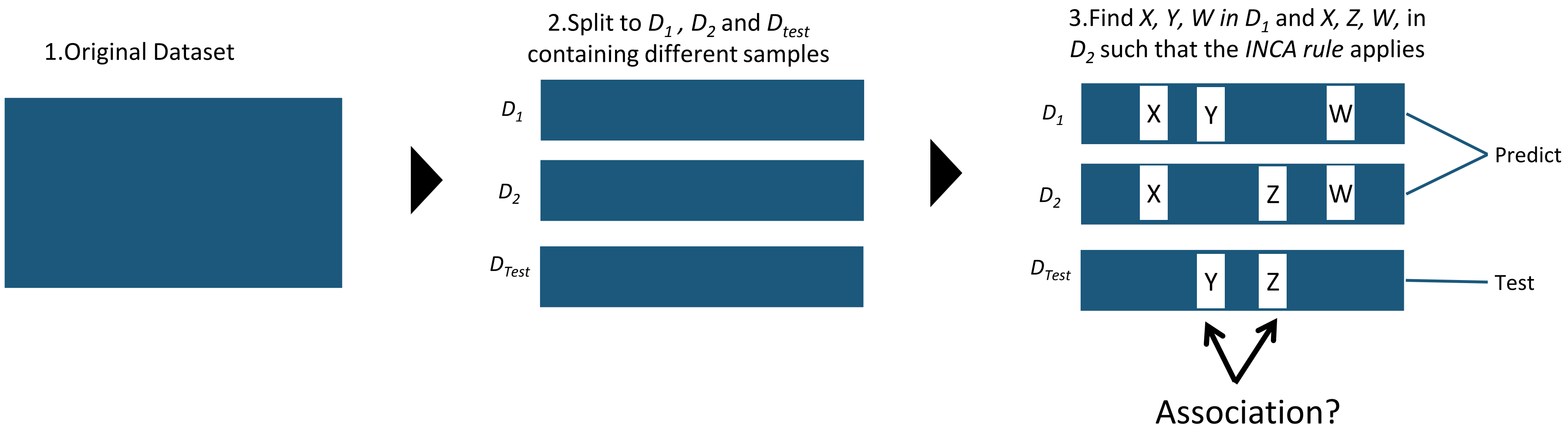
Massively test the assumptions

- Find data sets D_1, D_2 where the rule holds.
- Predict Y, Z are dependent. (given $\emptyset$)
- Check in an independent data set D_{Test} where Y, Z are measured.

FURTHER INFERENCE: PREDICT CORRELATION STRENGTH



MAKING IT WORK ON REAL DATA



- Restrict inferences only to cases where the probability of errors is small, i.e. *p*-values are extreme

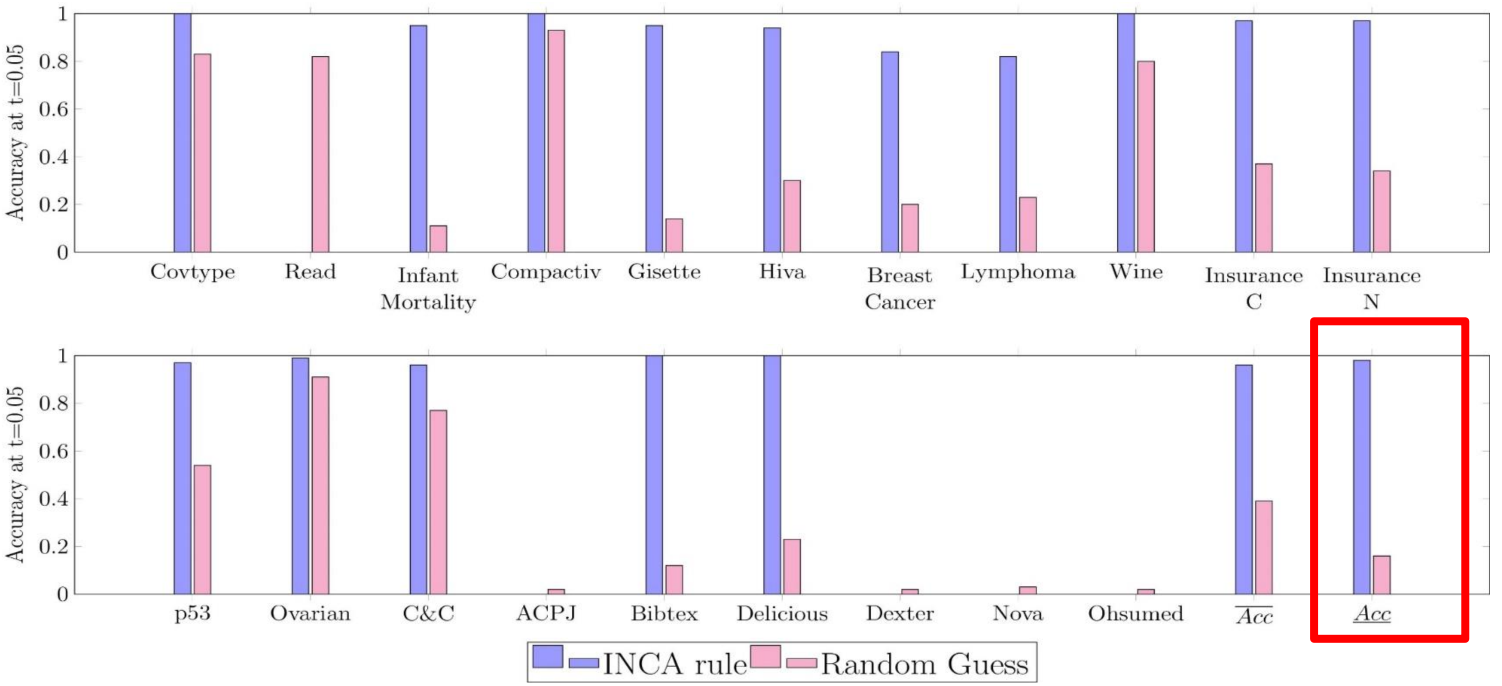
  $p_{XY.Z}$ < 0.05 accept $Dep(X, Y|\mathbf{Z})$
  $p_{XY.Z}$ > 0.3 accept $Ind(X, Y|\mathbf{Z})$
  Else, undecided (forgo making any inferences)

DATASETS

Name	# instances	# variables	Group Size	Variables type	Scientific domain
Covtype	581012	55	55	Nominal/Ordinal	Agricultural
Read	681	26	26	Nominal/Continuous/Ordinal	Business
Infant-mortality	5337	83	83	Nominal	Clinical study
Compactiv	8192	22	22	Continuous	Computer science
Gisette	7000	5000	50	Continuous	Digit recognition
Hiva	4229	1617	50	Nominal	Drug discovering
Breast-Cancer	286	17816	50	Continuous	Gene expression
Lymphoma	237	7399	50	Continuous	Gene expression
Wine	4898	12	12	Continuous	Industrial
Insurance-C	9000	84	84	Nominal/Ordinal	Insurance
Insurance-N	9000	86	86	Nominal/Ordinal	Insurance
p53	16772	5408	50	Continuous	Protein activity
Ovarian	216	2190	50	Continuous	Proteomics
C&C	1994	128	128	Continuous	Social science
ACPJ	15779	28228	50	Continuous	Text mining
Bibtex	7395	1995	50	Nominal	Text mining
Delicious	16105	1483	50	Nominal	Text mining
Dexter	600	11035	50	Nominal	Text mining
Nova	1929	12709	50	Nominal	Text mining
Ohsumed	5000	14373	50	Nominal	Text mining

predictions
222
0
22
135
423
554
1833
7712
4
1839
226
46647
539165
99241
0
1
856
0
0
0

HOW DID WE DO?



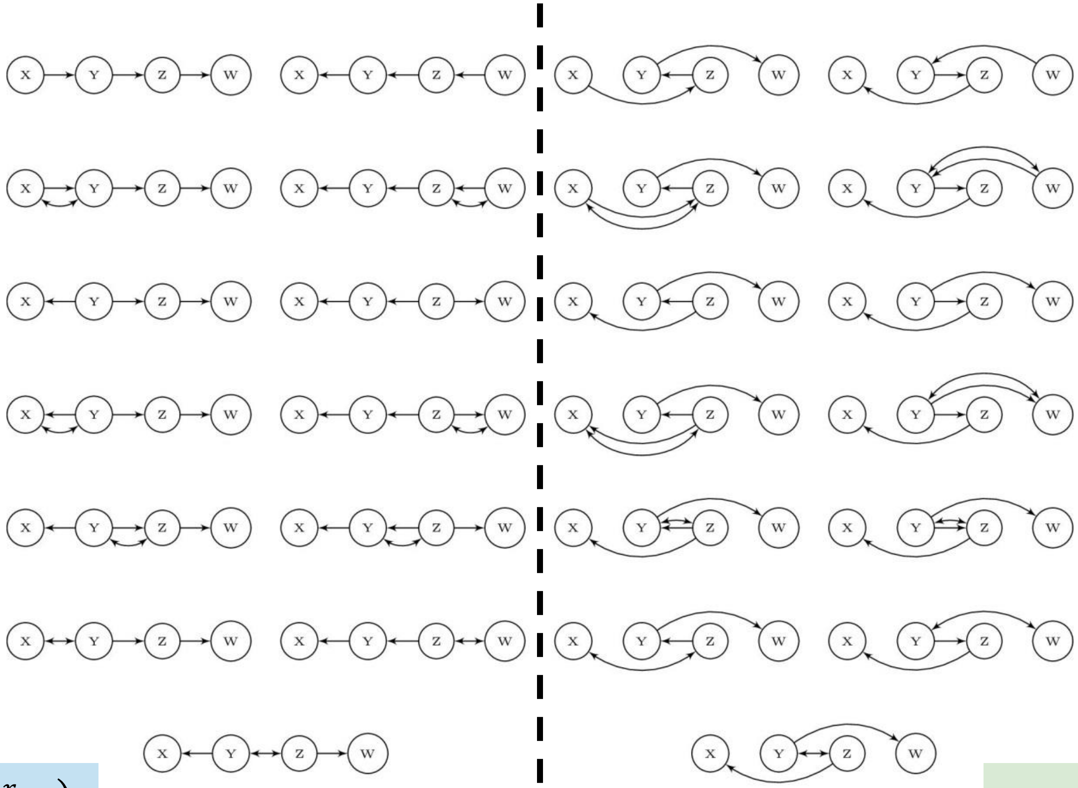
- About 700000 predictions in 20 datasets.
- Accuracy: The percentage of p-values < 0.05.
  - May include false positives and exclude false negatives.

98% accuracy vs.
16% of random guessing

FURTHER INFERENCE: PREDICT CORRELATION STRENGTH

$\rho_{XZ} = \rho_{XY} \times \rho_{YZ}$

$\rho_{YW} = \rho_{YZ} \times \rho_{ZW}$



$\rho_{XY} = \rho_{XZ} \times \rho_{YZ}$

$\rho_{ZW} = \rho_{YZ} \times \rho_{YW}$

Predict $\widehat{r_{YZ}}^{1} \approx \frac{1}{2}\left(\frac{r_{XZ}}{r_{XY}} + \frac{r_{YW}}{r_{ZW}}\right)$

Only one of $\left|\widehat{r_{YZ}}^{1}\right|, \left|\widehat{r_{YZ}}^{2}\right|$ **is < 1**

Predict $\widehat{r_{YZ}}^{2} \approx \frac{1}{2}\left(\frac{r_{XY}}{r_{XZ}} + \frac{r_{ZW}}{r_{YW}}\right)$

VS STATISTICAL MATCHING

Conditional Independence Assumption (CIA)

Non common variables are **independent** given the **common variables**

+ Multivariate Normality

vs.

1. **(Causal) Markov**
2. **(Causal) Faithfulness**
3. **Acyclicity**
4. **Multivariate normality**

Data Sets	SMR	INCA rule
ACPJ	0.00 [0.00;0.01]	-
Breast-Cancer	0.25 [0.24;0.25]	0.88 [0.87;0.90]
C&C	0.68 [0.65;0.71]	0.91 [0.91;0.91]
Compactiv	0.49 [0.44;0.54]	0.88 [0.83;0.92]
Insurance-C	0.47 [0.42;0.51]	0.90 [0.89;0.91]
Lymphoma	0.32 [0.31;0.32]	0.50 [0.47;0.52]
Ohsumed	0.01 [0.00;0.01]	-
Ovarian	0.50 [0.50;0.51]	0.14 [0.14;0.14]
Wine	0.58 [0.52;0.64]	0.99 [0.47;1.00]
p53	0.45 [0.44;0.45]	0.87 [0.87;0.87]
Mean over data sets	0.38 [0.35;0.40]	0.76 [0.68;0.77]
On all predictions	0.58 [0.57;0.58]	0.89 [0.89;0.89]

- When predictions are based on only 2 common variables, statistical matching is unreliable.
- INCA rule's predictions are highly correlated with sample estimates (0.89 correlation.

CONTRIBUTIONS

- Comparison of causal models under causal insufficiency.
- Introduction of SAT-based causal analysis: exploit 40 years of SAT-solving technology.
- Query-based approach to avoid explosion of possible solutions.
- Method for estimating posterior probabilities from p-values.
- Scalable INCA algorithm.
- A proof of concept that causal assumptions can make testable qualitative and quantitative predictions.
- Being local and conservative improves applicability of causal methods.

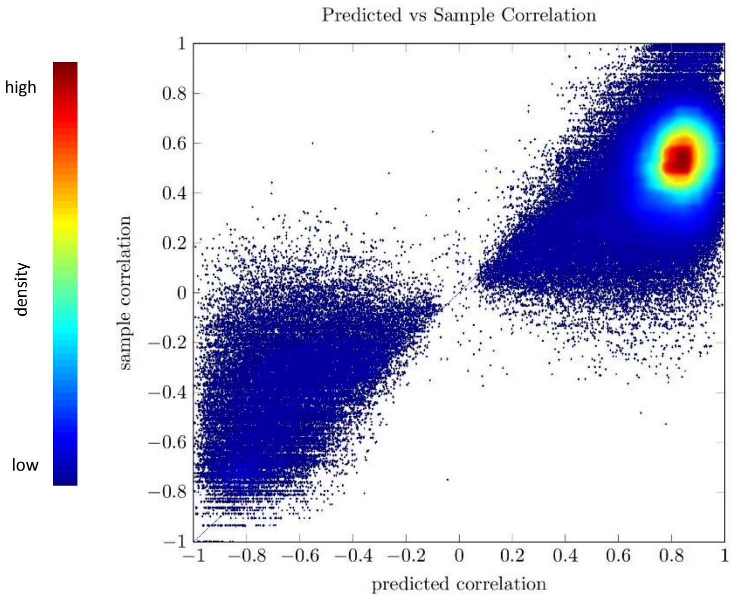
CONCLUSIONS AND FUTURE WORK

- Beyond one dataset at a time
  - Vision of automatically analyzing a large portion of available datasets in a domain
- Inclusion of Prior Causal Knowledge [ICML 2012, UAI 2013]
- Handling Case-Control Data [UAI 2015]
- Handling batch effects [upcoming]
- Handling temporal data and temporal information
- Improve reliability
- Ability to work with semantically similar data
- Quantitative Algorithms?

PUBLICATIONS

- S. Triantafillou and I. Tsamardinos, Causal Discovery from Multiple Interventions, JMLR, to appear.
- Giorgos Borboudakis, Ioannis Tsamardinos (2015).**Bayesian Network Learning with Discrete Case-Control Data.** Uncertainty in Artificial Intelligence (UAI), 2015.
- S. Triantafillou, I. Tsamardinos and A. Roumpelaki ,Finding Neighborhoods of High Confidence in Constraint-based Causal Discovery, *PGM 2014*.
- Borboudakis G., & Tsamardinos, I. (2013). Scoring and Searching over Bayesian Networks with Informative, Causal and Associative Priors. Uncertainty in Artificial Intelligence (UAI) 2013
- Borboudakis, G., & Tsamardinos, I. (2012). Incorporating Causal Prior Knowledge as Path-Constraints in Bayesian Networks and Maximal Ancestral Graphs. Proceedings of the 29th International Conference on Machine Learning, ICML 2012
- I. Tsamardinos, S. Triantafillou and V. Lagani ,Towards Integrative Causal Analysis of Heterogeneous Datasets and Studies. *JMLR.*, 13:1097-1157, 2012.
- S. Triantafillou, I. Tsamardinos, and I. G. Tollis. Learning causal structure from overlapping variable sets. In *Artificial Intelligence and Statistics*, pages 860-867, 2010.
- S. Triantafillou and I. Tsamardinos, Predicting associations from multiple “omics” data sets, HSCBB12, Best Poster Award Recipient.
- V. Lagani, S. Triantafillou, G. Ball, I. Tsamardinos and J. Tegner, Probabilistic Computational Causal Discovery for Systems Biology, In Uncertainty in Biology: a computational modeling approach, Amit Gefen (Editor), to appear.
- G. Borboudakis, S. Triantafillou and I. Tsamardinos Tools and Algorithms for Causally Interpreting Directed Edges in Maximal Ancestral Graphs, PGM 2012.
- V. Lagani, I. Tsamardinos and S. Triantafillou Learning from mixture of experimental data: a constraint-based approach In Hellenic Conference on Artificial Intelligence (SETN 2012).
- G. Borboudakis, S. Triantafillou, V. Lagani, I. Tsamardinos, A constraint-based approach to incorporate prior knowledge in causal model, ESANN 2011.

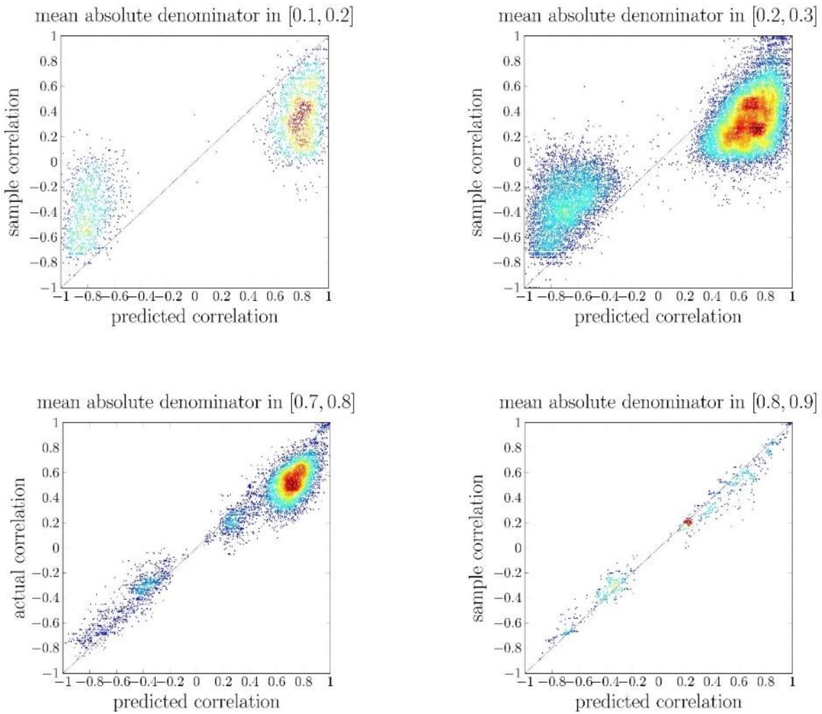
BIAS OF PREDICTIONS



why?

When actual correlation is low, only overestimated sample correlations pass the independence test

Predicted vs sample correlations over all data sets, grouped by mean absolute value of the denominators used in their computations



- Correlation of predicted vs sample correlations is 0.89
- Predictions based on large correlations have reduced bias

VS STATISTICAL MATCHING

Conditional Independence Assumption (CIA)

Non common variables are **independent** given the **common variables**

\+ **Multivariate Normality**

vs.

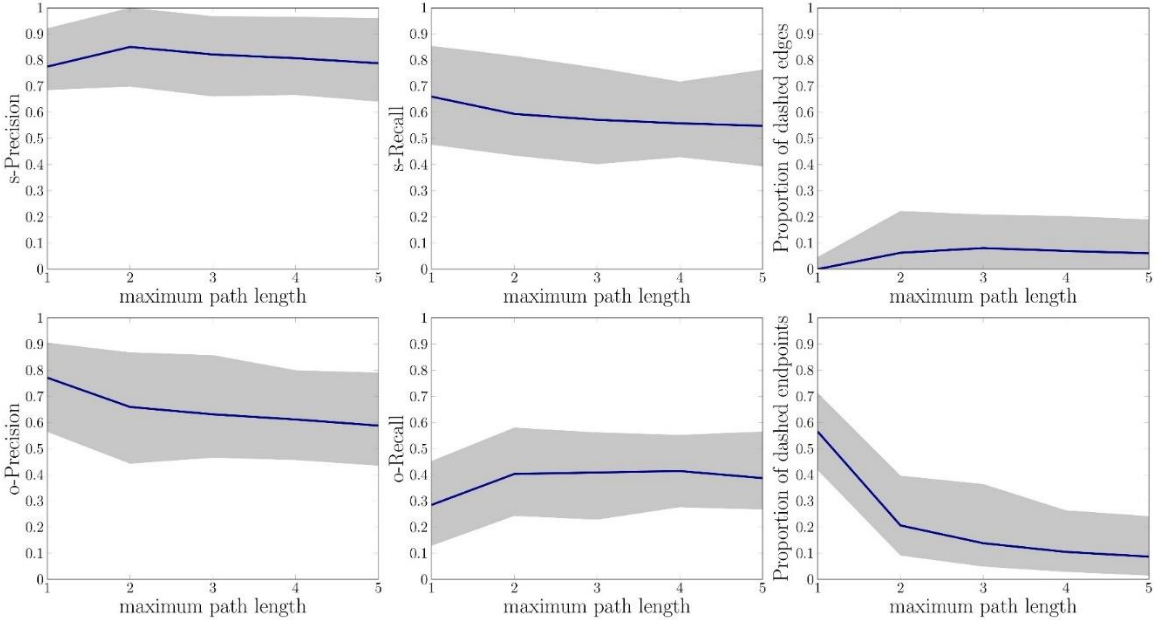
1. **(Causal) Markov**
2. **(Causal) Faithfulness**
3. **Acyclicity**
4. **Multivariate normality**

Data Sets	SMR_{all}	SMR_{XW}	INCA rule
ACPJ	0.05 [0.04;0.05]	0.00 [0.00;0.01]	-
Breast-Cancer	0.55 [0.55;0.55]	0.25 [0.24;0.25]	0.88 [0.87;0.90]
C&C	0.99 [0.99;0.99]	0.68 [0.65;0.71]	0.91 [0.91;0.91]
Compactiv	0.97 [0.96;0.98]	0.49 [0.44;0.54]	0.88 [0.83;0.92]
Insurance-C	0.83 [0.82;0.84]	0.47 [0.42;0.51]	0.90 [0.89;0.91]
Lymphoma	0.60 [0.60;0.60]	0.32 [0.31;0.32]	0.50 [0.47;0.52]
Ohsumed	0.02 [0.01;0.03]	0.01 [0.00;0.01]	-
Ovarian	0.62 [0.62;0.63]	0.50 [0.50;0.51]	0.14 [0.14;0.14]
Wine	0.83 [0.74;0.90]	0.58 [0.52;0.64]	0.99 [0.47;1.00]
p53	0.91 [0.91;0.91]	0.45 [0.44;0.45]	0.87 [0.87;0.87]
Mean over data sets	0.64 [0.62;0.65]	0.38 [0.35;0.40]	0.76 [0.68;0.77]
On all predictions	0.73 [0.73;0.73]	0.58 [0.57;0.58]	0.89 [0.89;0.89]

- When predictions are based on only 2 common variables, statistical matching is unreliable
- SM is more successful when the predictions are based on larger sets of common variables.
- INCA rules's predictions are highly correlated with sample estimates (0.89 correlation)

MAXIMUM PATH LENGTH

20 variables, 5 overlapping data sets, 100 samples each



- MPL controls path length.
- Unconstrained MPL corresponds to soundness and completeness.
- MPL>1 does not affect performance, allows more orientations.