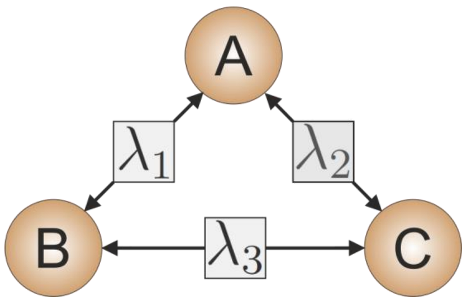
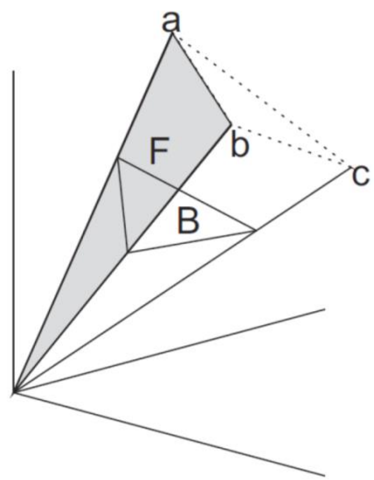
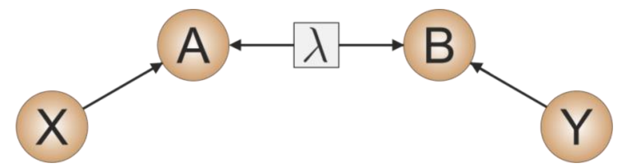
An entropic approach to causal inference with applications to nonlocality[2] and machine learning[1]







Rafael Chaves

CEQIP 2014

Joint work[1] with L. Luft, T. Maciel, D. Gross, D. Janzing and B. Schölkopf

Joint work[2] with C. Majens and D. Gross.

Causal inference
and
quantum non-locality?

Reichenbach‘s principle: no correlation without causation.

Reichenbach‘s principle: no correlation without causation.

„If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“

Reichenbach's principle: no correlation without causation.

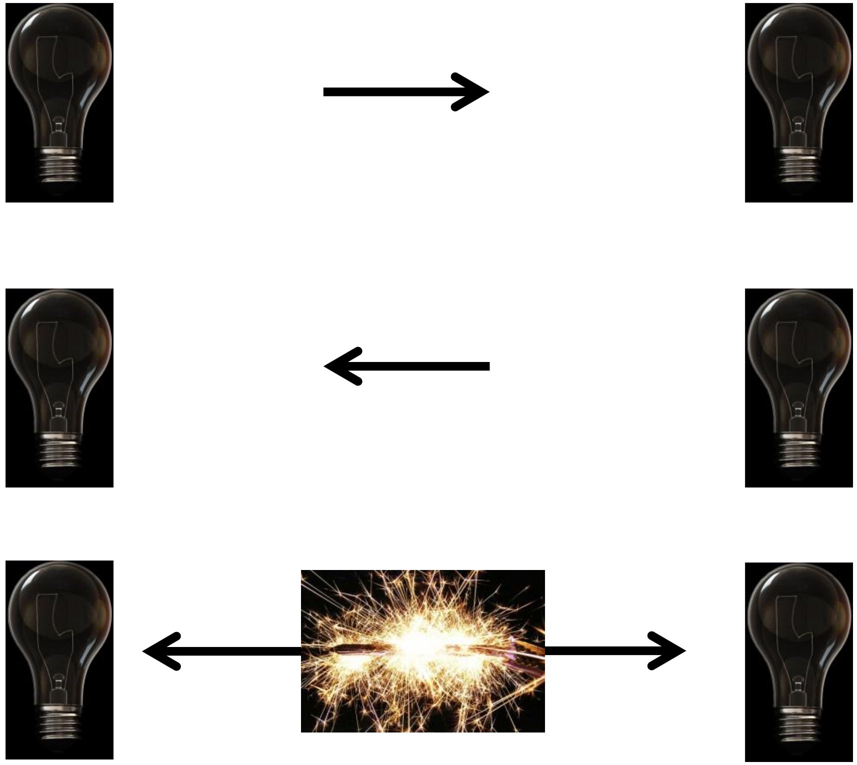
„If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“

Reichenbach‘s principle: no correlation without causation.

„If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“

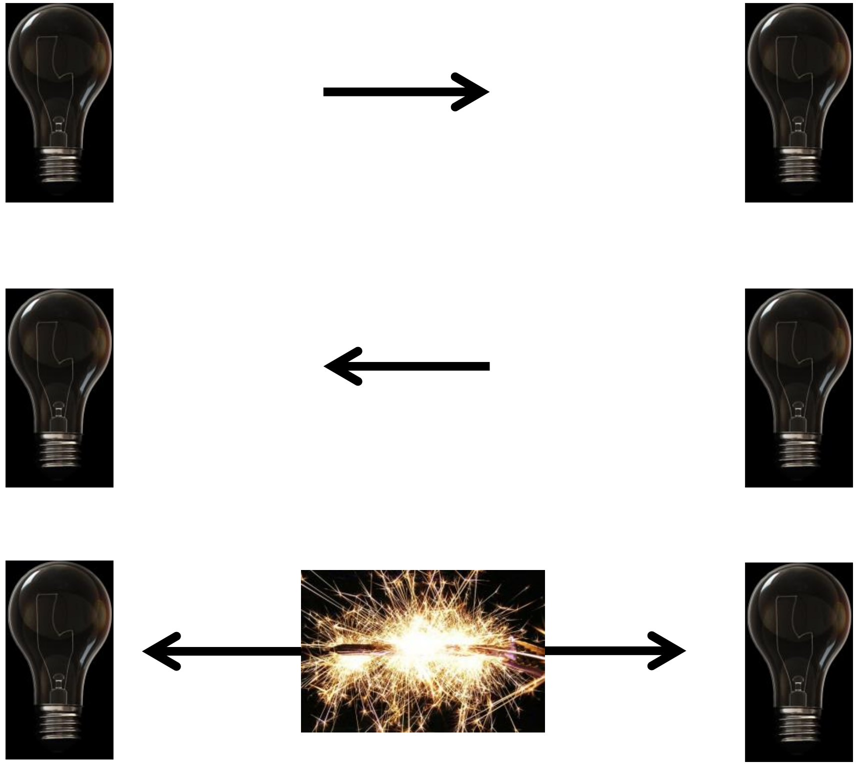
Reichenbach‘s principle: no correlation without causation.

„If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“



Reichenbach's principle: no correlation without causation.

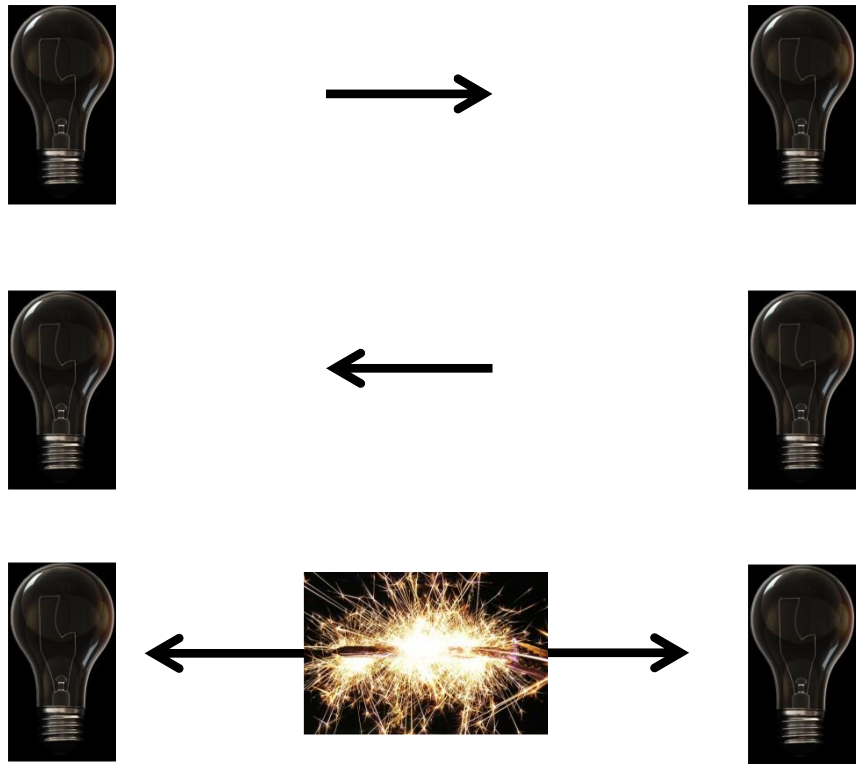
„If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“

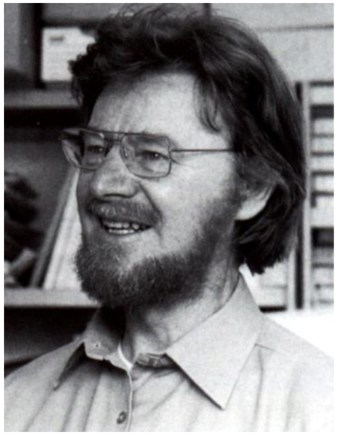


Task: Infer causal relationships from raw (**perhaps marginal**) data

Reichenbach's principle: no correlation without causation.

„If an improbable coincidence has ocurred, there must exist direct influence and/or a common cause.“

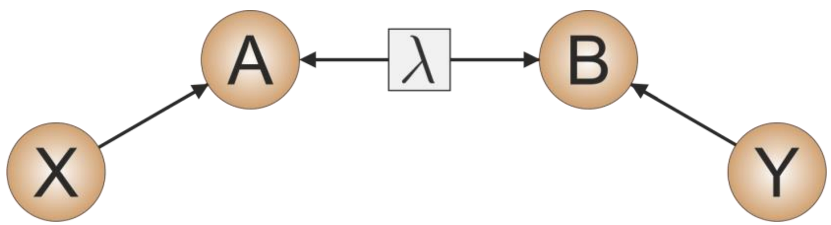




Task: Infer causal relationships from raw (**perhaps marginal**) data

Bell´s theorem

The assumptions (causal relationships) of a LHV model impose constraints on the possible observed distributions. Those can be tested via Bell inequalities, that may be violated by quantum states.



Task: Infer causal relationships from raw (**perhaps marginal**) data.

How to do that?

Main idea

Challenge:

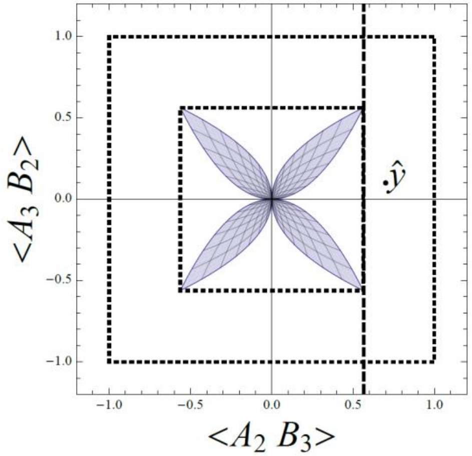
- Describe marginals compatible with DAGs...

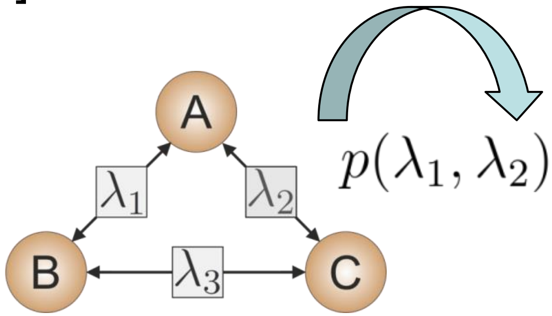
Main idea

Challenge:

- Describe marginals compatible with DAGs...
- ...very difficult, non-convex sets (e.g., quantifier elimination by **[Geiger and Meek, UAI 1999]**)

Picture from **[Steeg and Galstyan, UAI 2011]**





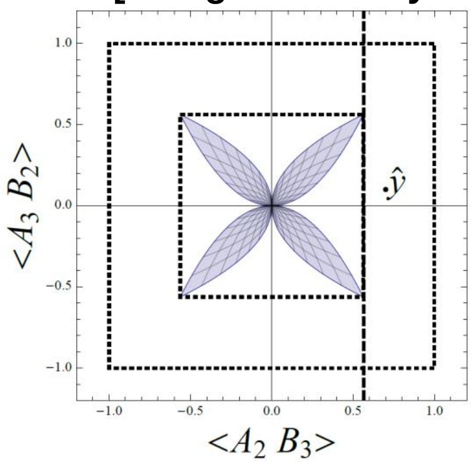
$$p(\lambda_1, \lambda_2) = p(\lambda_1)p(\lambda_2)$$

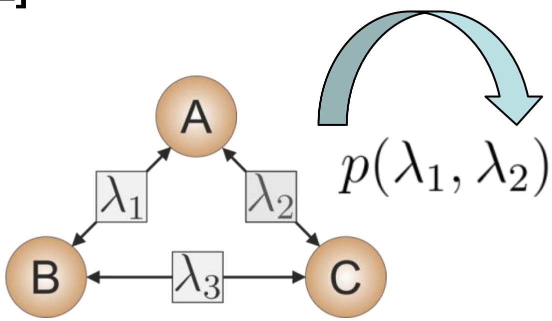
Main idea

Challenge:

- Describe marginals compatible with DAGs...
- ...very difficult, non-convex sets (e.g., quantifier elimination by **[Geiger and Meek, UAI 1999]**)

Picture from **[Steeg and Galstyan, UAI 2011]**





$$p(\lambda_1, \lambda_2) = p(\lambda_1)p(\lambda_2)$$

Our idea

Rely on entropic information!

- Concise characterization as a convex set
- Naturally encodes the causal constraints
- Quantitative and stable tool

Outline

- The entropic approach to Causal Inference
- Applications
- Where to go from here?

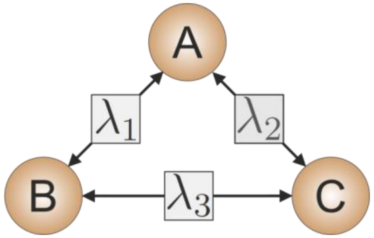
- The entropic approach to Causal Inference
- Applications
- Where to go from here?

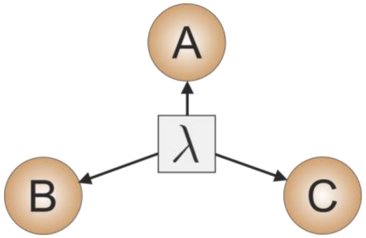
DAGs

- For n variables $X_1, \ldots, X_n$, the causal relationships are encoded in a causal structure, represented by
- a directed acyclic graph (DAG),
- with *i*th variable being a deterministic

  $X_i = f_i(pa_i, u_i)$

  of its parents $\mathbf{pa}_i$ and jointly independent noise variables $\mathbf{u}_i$



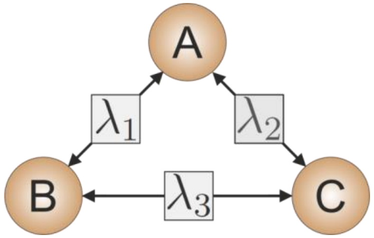


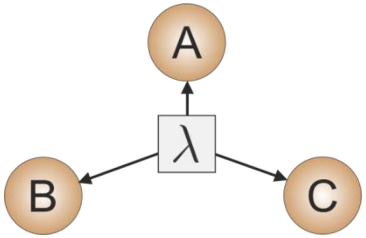
DAGs

- For n variables X_1, ... ,X_n**,** the causal relationships are encoded in a causal structure, represented by
- a directed acyclic graph (DAG),
- with *i*th variable being a deterministic

  $\mathbf{X_i=f_i(pa_i,u_i)}$

  of its parents $\mathbf{pa_i}$ and jointly independent noise variables $\mathbf{u_i}$
- Some of the variables are not observed (latent variables)



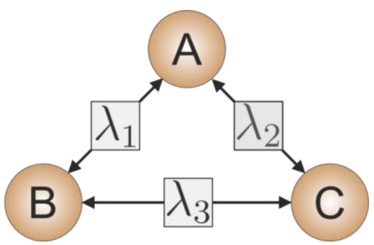


DAGs

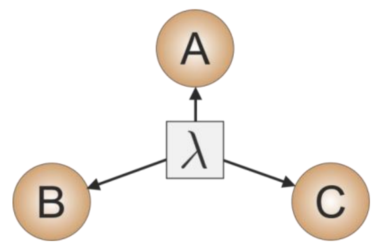
- For n variables $X_1, \ldots, X_n$, the causal relationships are encoded in a causal structure, represented by
- a directed acyclic graph (DAG),
- with *i*th variable being a deterministic

  $$X_i=f_i(pa_i,u_i)$$

  of its parents $\mathbf{pa_i}$ and jointly independent noise variables $\mathbf{u_i}$
- Some of the variables are not observed (latent variables)
- The DAG encodes causal constraints as (conditional) independences



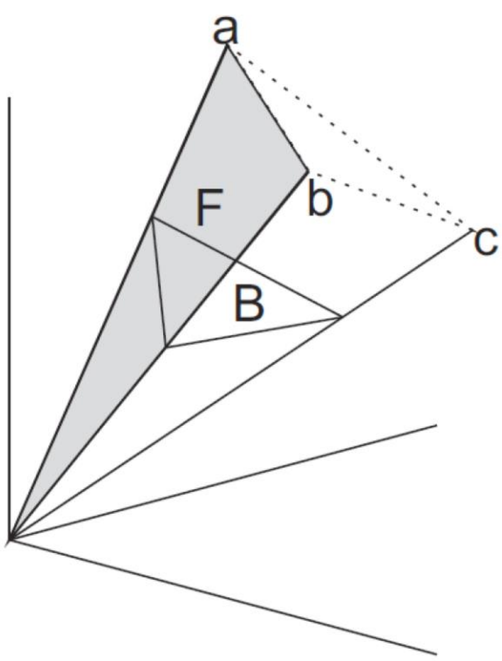
$$I(\lambda_1 : \lambda_2) = 0$$
$$I(B : C|\lambda_3) = 0$$
...



$$I(B : C|\lambda) = 0$$
...

Entropic cone

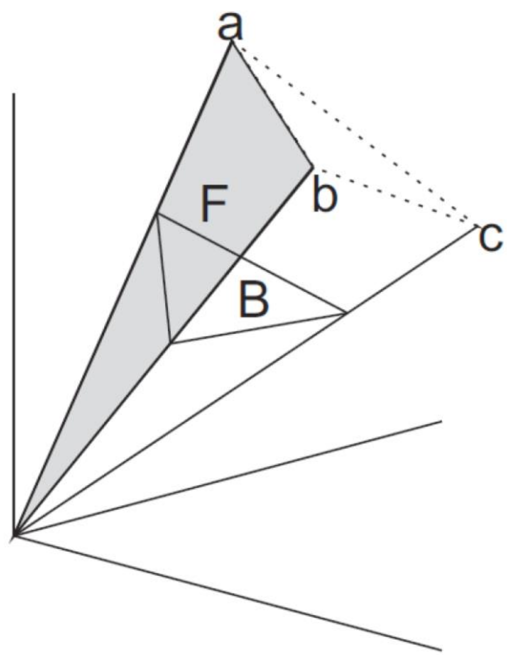
Step 1/3: *Unconstrained, global object*



- Entropic vector $V \in \mathbb{R}^{2^n}$:each entry is the entropy $S(X_S)$ indexed by subset $S \subset \{1, \ldots, n\}$
- Defines a convex cone
- Structure not fully understood, but...
- ...contained in Shannon Cone, defined by subadditivity and monotonicity (polymatroidal axioms)
- ***Shannon cone***: Γ_n

Entropic cone

Step 1/3: *Unconstrained, global object*



- Entropic vector $V \in \mathbb{R}^{2^n}$:each entry is the entropy $S(X_S)$ indexed by subset $S \subset \{1, \dots, n\}$
- Defines a convex cone
- Structure not fully understood, but...
- ...contained in Shannon Cone, defined by subadditivity and monotonicity (polymatroidal axioms)
- ***Shannon cone:*** Γ_n

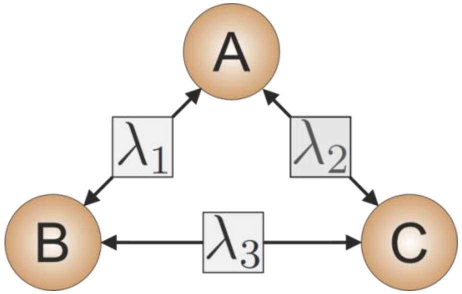
Example: *3 variables* $\rightarrow$ $v = (H(\emptyset), H(A), H(B), H(C), H(A,B), H(A,C), H(B,C), H(A,B,C))$

polymatroidal axioms $\longrightarrow$

$$H(A,B,C) + H(A) \leq H(A,B) + H(A,C) \qquad (I(B:C|A) \geq 0)$$
$$H(A,B) \leq H(A) + H(B) \qquad (I(A:B) \geq 0)$$
$$H(A,B) \leq H(A,B,C)$$

Entropic cone

Step 2/3: *Choose candidate structure and add causal constraints*



- Piece of cake! Conditional independences are naturally embedded in mutual informations

$$\begin{aligned} p(\lambda_1, \lambda_2) &= p(\lambda_1)p(\lambda_2) \\ p(A, B|\lambda_1) &= p(A|\lambda_1)p(B|\lambda_1) \end{aligned} \longrightarrow \begin{aligned} I(\lambda_1 : \lambda_2) &= 0 \\ I(A : B|\lambda_1) &= 0 \end{aligned}$$

- We can even relax (stable!)

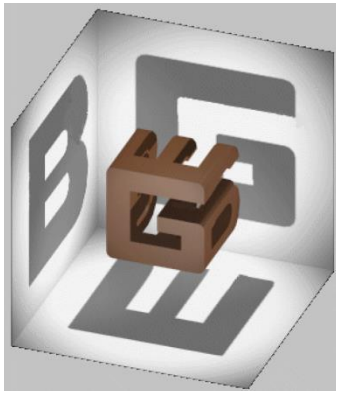
$$\begin{aligned} I(\lambda_1 : \lambda_2) &\leq \epsilon_1 \\ I(A : B|\lambda_1) &\leq \epsilon_2 \end{aligned}$$

- *C: cone of constraints*

- New global cone $\Gamma_n \cap C$ of entropies subject to causal structure

Entropic cone

Step 3/3: *Marginalize to $\mathcal{M}$*



- $\mathcal{M} \subset 2^{\{1,\dots,n\}}$: set of jointly observables

- **Geometrically trivial**:

  just restrict $\Gamma_n \cap C$ to observable coordinates

- **Algorithmically costly**: $\Gamma_n \cap C$ represented in terms of inequalities (use, eg, Fourier-Motzkin elimination)

Final result: description of marginal, causal entropic cone $(\Gamma_n \cap C)_{|\mathcal{M}}$ in terms of „entropic Bell inequalities“

[T. Fritz and RC, IEEE Trans. Inf. Th. 59, 803 (2013)]

[RC, L. Luft, D. Gross, NJP 16, 043001 (2014)]

- The entropic approach to Causal Inference
- Applications
  - Classical
  - Quantum
- Where to go from here?

Classical

RC, L. Luft, T. Maciel, D. Gross, D. Janzing, B. Schölkopf.
To appear in *Conference on Uncertainty in Artificial Intelligence 2014*

Common ancestors problem

- Can the correlations between n variables be explained by common ancestors connecting at most M of them? **[Steudel and Ay, arXiv:1010.5720]**

Entropic approach: Polymatroidal axioms + Causal Structure + Marginalization

Common ancestors problem

- Can the correlations between n variables be explained by common ancestors connecting at most M of them? **[Steudel and Ay, arXiv:1010.5720]**

Entropic approach: Polymatroidal axioms + Causal Structure + Marginalization

$$\sum_{i=2,\ldots,n} I(X_1 : X_i) \leq (M-1)H(X_1)$$

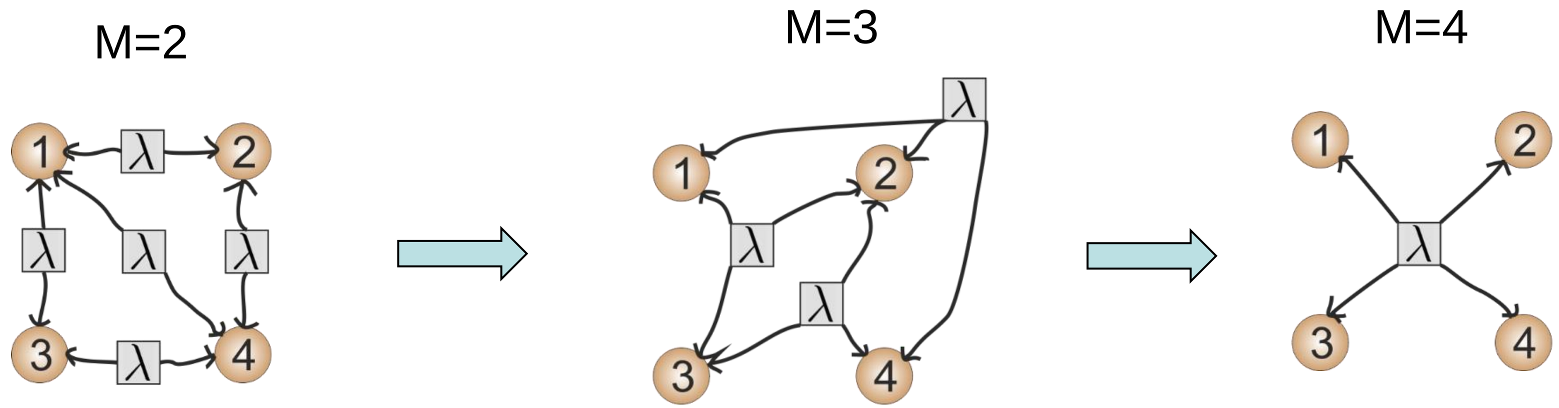
Common ancestors problem

- Can the correlations between n variables be explained by common ancestors connecting at most M of them? **[Steudel and Ay, arXiv:1010.5720]**

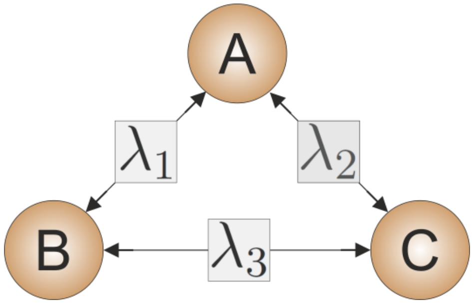
Entropic approach: Polymatroidal axioms + Causal Structure + Marginalization

$$\sum_{i=2,\dots,n} I(X_1 : X_i) \leq (M-1)H(X_1)$$

A hierarchy of causal relationship tests....



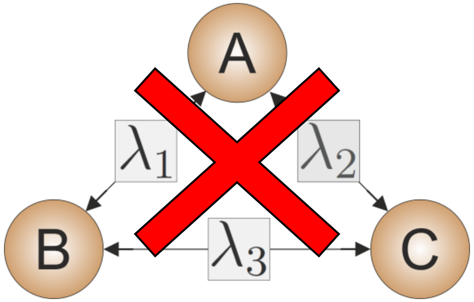
Quantifying causal influences



$$\mathcal{B} = I(A:B) + I(A:C) - H(A) \leq 0$$

$$p(A=a, B=b, C=c) = \begin{cases} 1/2 & , a=b=c \\ 0 & , \text{otherwise} \end{cases}$$

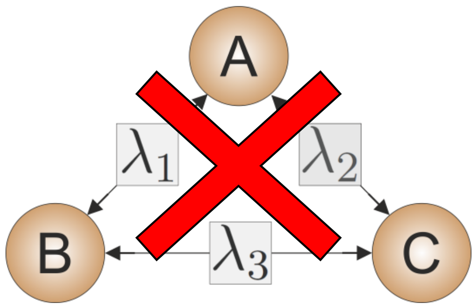
Quantifying causal influences



$$\mathcal{B} = I(A:B) + I(A:C) - H(A) \leq 0$$

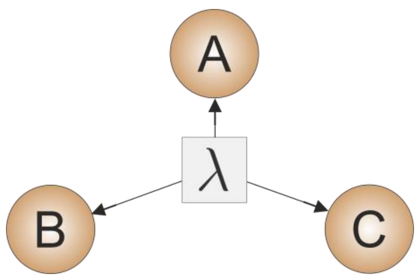
$$p(A=a, B=b, C=c) = \begin{cases} 1/2 & , a=b=c \\ 0 & , \text{otherwise} \end{cases}$$

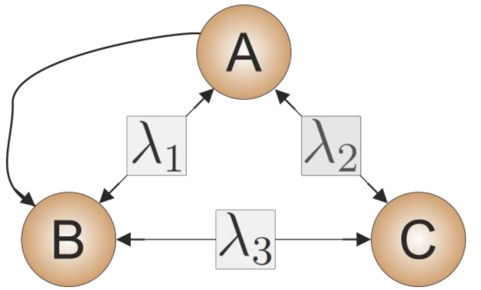
Quantifying causal influences



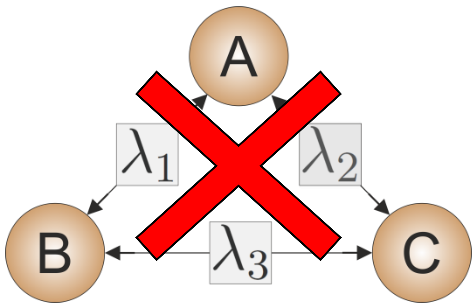
$$\mathcal{B} = I(A:B) + I(A:C) - H(A) \leq 0$$

$$p(A=a, B=b, C=c) = \begin{cases} 1/2 & , a=b=c \\ 0 & , \text{otherwise} \end{cases}$$



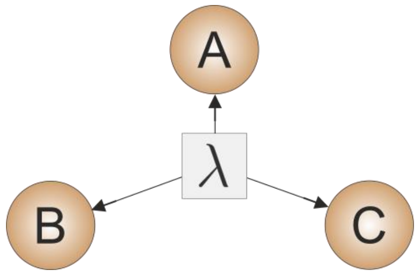


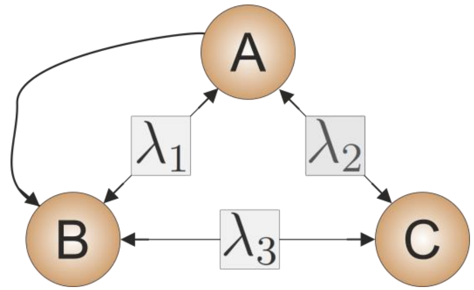
Quantifying causal influences



$$\mathcal{B} = I(A:B) + I(A:C) - H(A) \leq 0$$

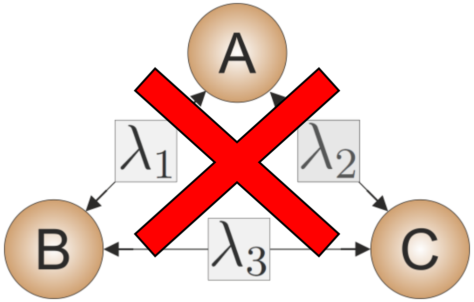
$$p(A=a, B=b, C=c) = \begin{cases} 1/2 & , a=b=c \\ 0 & , \text{otherwise} \end{cases}$$





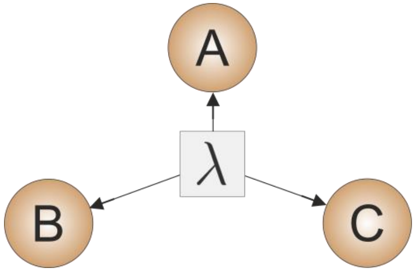
$$I(A:B|pa_B) \leq \mathcal{C}_{A \to B}$$

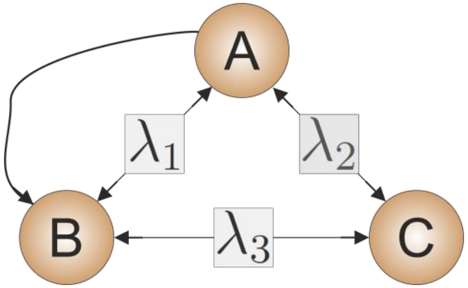
Quantifying causal influences



$$\mathcal{B} = I(A:B) + I(A:C) - H(A) \leq 0$$

$$p(A=a, B=b, C=c) = \begin{cases} 1/2 & , a=b=c \\ 0 & , \text{otherwise} \end{cases}$$





$$I(A:B|pa_B) \leq \mathcal{C}_{A\to B}$$

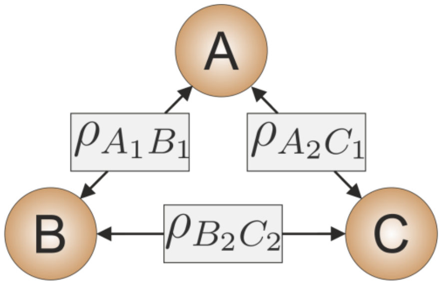
$$\mathcal{B} \leq I(A:B|pa_B) \leq \mathcal{C}_{A\to B}$$

Quantum

RC, C. Majens and D. Gross. In preparation.

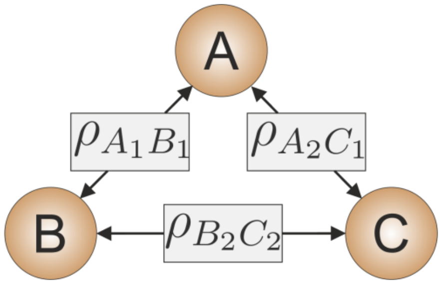
Classical/Quantum Bayesian networks?

- Entropic description of Bayesian networks with quantum „hidden" variables



Classical/Quantum Bayesian networks?

- Entropic description of Bayesian networks with quantum „hidden“ variables



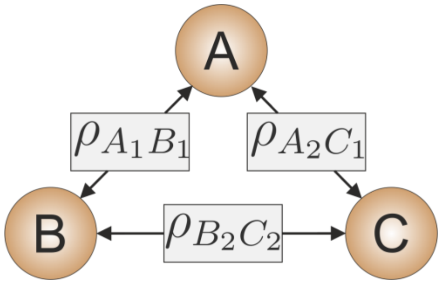
- Quantum variables respect strong subadditivity but not monotonicity

$$H(\rho_{A_1}|\rho_{B_1}) \geq 0$$

$$H(A|\rho_{B_1C_1}) \geq 0$$

Classical/Quantum Bayesian networks?

➢ Entropic description of Bayesian networks with quantum „hidden" variables



- Quantum variables respect strong subadditivity but not monotonicity

$$H(\rho_{A_1}|\rho_{B_1}) \geq 0$$

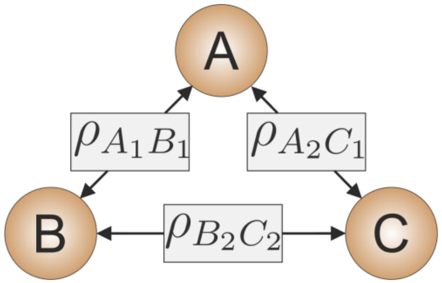
$$H(A|\rho_{B_1C_1}) \geq 0$$

- Measurements disturb/destroy the quantum system

$$H(\rho_{A_1A_2}, A)$$

Classical/Quantum Bayesian networks?

- Entropic description of Bayesian networks with quantum „hidden“ variables



- Quantum variables respect strong subadditivity but not monotonicity

$$H(\rho_{A_1}|\rho_{B_1}) \geq 0$$

$$H(A|\rho_{B_1C_1}) \geq 0$$

- Measurements disturb/destroy the quantum system

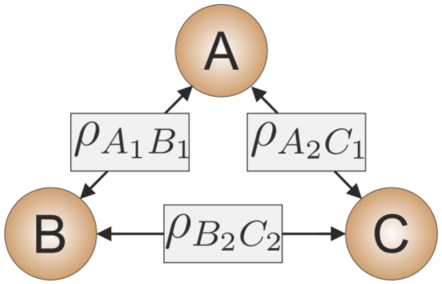
$$H(\rho_{A_1A_2}, A)$$

- We need a rule mapping the quantum states to classical variables

$$I(A:B) \leq I(A_1A_2:B_1B_2)$$

Classical/Quantum Bayesian networks?

- Entropic description of Bayesian networks with quantum „hidden“ variables



- Quantum variables respect strong subadditivity but not monotonicity

$$✗\quad H(\rho_{A_1}|\rho_{B_1}) \geq 0$$

$$✓\quad H(A|\rho_{B_1C_1}) \geq 0$$

- Measurements disturb/destroy the quantum system

$$✗\quad H(\rho_{A_1A_2}, A)$$

- We need a rule mapping the quantum states to classical variables

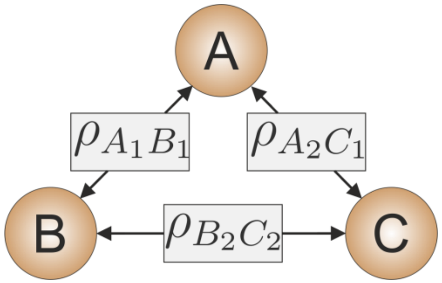
$$✓\quad I(A:B) \leq I(A_1A_2 : B_1B_2)$$

- Set of conditional independencies are fulfilled

$$✓\quad I(A:B|\rho_{A_1B_1}) = 0$$

Classical/Quantum Bayesian networks?

- Entropic description of Bayesian networks with quantum „hidden“ variables



- Quantum variables respect strong subadditivity but not monotonicity

$$H(\rho_{A_1}|\rho_{B_1}) \geq 0$$
$$H(A|\rho_{B_1C_1}) \geq 0$$

- Measurements disturb/destroy the quantum system

$$H(\rho_{A_1A_2}, A)$$

- We need a rule mapping the quantum states to classical variables

$$I(A:B) \leq I(A_1A_2:B_1B_2)$$

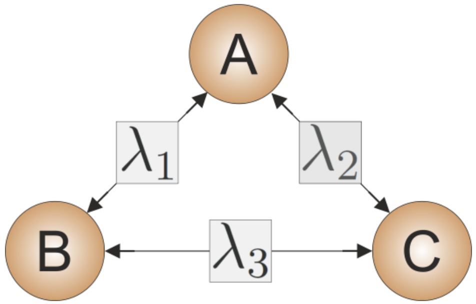
- Set of conditional independencies are fulfilled

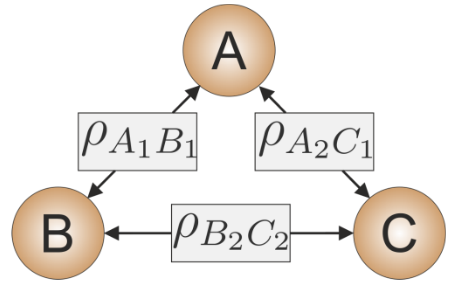
$$I(A:B|\rho_{A_1B_1}) = 0$$

- For classical variable the entropy H corresponds to the Shannon entropy

Quantum common ancestor networks

- Marginal entropic cones coincide!





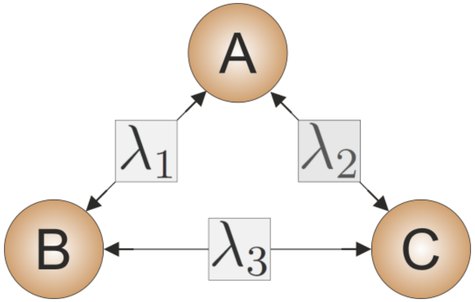
$$I(A:B)+I(A:C)\leq H(A)$$

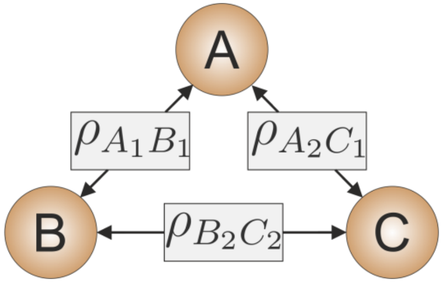
$$I(A:B:C)+I(A:B)+I(A:C)+I(B:C)\leq H(A,B)$$

$$I(A:B:C)+I(A:B)+I(A:C)+I(B:C)\leq \frac{1}{2}(H(A)+H(B)+H(C))$$

Quantum common ancestor networks

- Marginal entropic cones coincide!





$$I(A:B)+I(A:C)\leq H(A)$$

$$I(A:B:C)+I(A:B)+I(A:C)+I(B:C)\leq H(A,B)$$

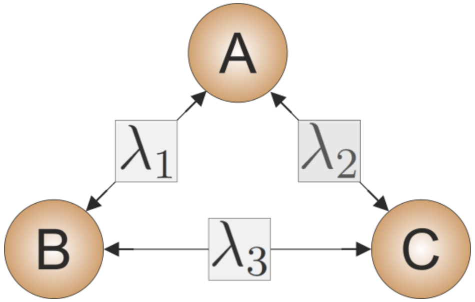
$$I(A:B:C)+I(A:B)+I(A:C)+I(B:C)\leq \tfrac{1}{2}(H(A)+H(B)+H(C))$$

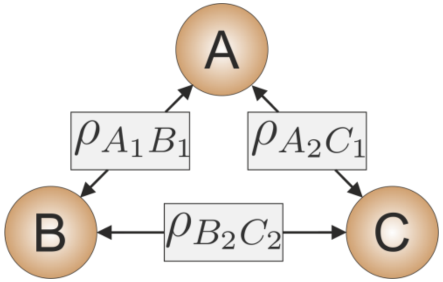
- For arbitrary quantum common ancestor DAGs the monogamy relation holds

$$\sum_{i=2,\ldots,n} I(X_1:X_i)\leq (M-1)H(X_1)$$

Quantum common ancestor networks

- Marginal entropic cones coincide!





$$I(A:B)+I(A:C)\leq H(A)$$

$$I(A:B:C)+I(A:B)+I(A:C)+I(B:C)\leq H(A,B)$$

$$I(A:B:C)+I(A:B)+I(A:C)+I(B:C)\leq \tfrac{1}{2}(H(A)+H(B)+H(C))$$

- For arbitrary quantum common ancestor DAGs the monogamy relation holds

$$\sum_{i=2,\ldots,n} I(X_1:X_i)\leq (M-1)H(X_1)$$

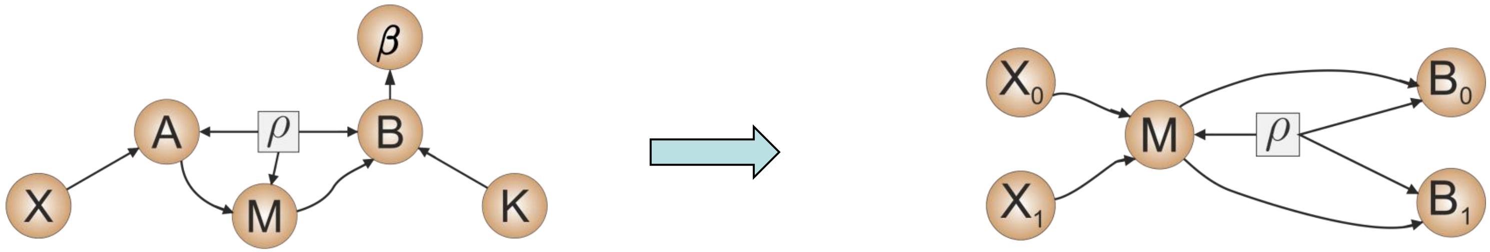
- Are these also valid for GPTs?

Case n=3, M=2 in **[J.Henson, R. Lal, M. F. Pusey arXiv:1405.2572]**

Not difficult to generalize the proof for any n and M=2

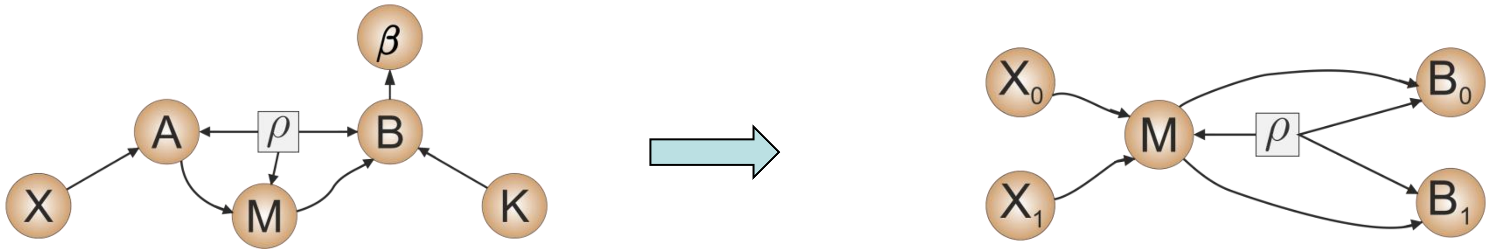
Information causality

[Nature 461, 1101 (2009)]



Information causality

[Nature 461, 1101 (2009)]



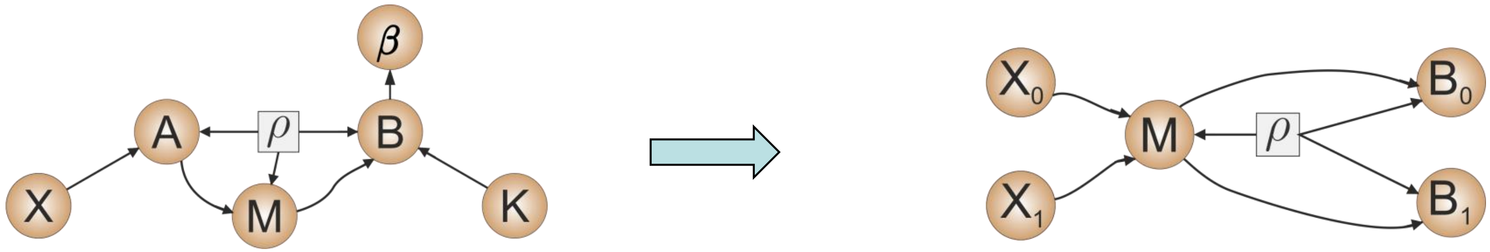
- Restricting to the marginal information: {X_0, B_0}, {X_1, B_1}, {X_0, X_1}, {M}

$$I(X_0 : B_0) + I(X_1 : B_1) \leq H(M) + I(X_0 : X_1)$$

Same as ineq as derived in **[Al-Safi and Short,PRA 84, 042323 (2011)]**

Information causality

[Nature 461, 1101 (2009)]



- Restricting to the marginal information: $\{X_0, B_0\}$, $\{X_1, B_1\}$, $\{X_0, X_1\}$, $\{M\}$

$$I(X_0 : B_0) + I(X_1 : B_1) \leq H(M) + I(X_0 : X_1)$$

Same as ineq as derived in **[Al-Safi and Short,PRA 84, 042323 (2011)]**

- Restricting to the marginal information: $\{X_0, X_1, B_0\}$, $\{X_0, X_1, B_1\}$, $\{M\}$

$$I(X_0 : B_0) + I(X_1 : B_1) + I(X_0 : X_1|B_1) \leq H(M) + I(X_0 : X_1)$$

Information causality

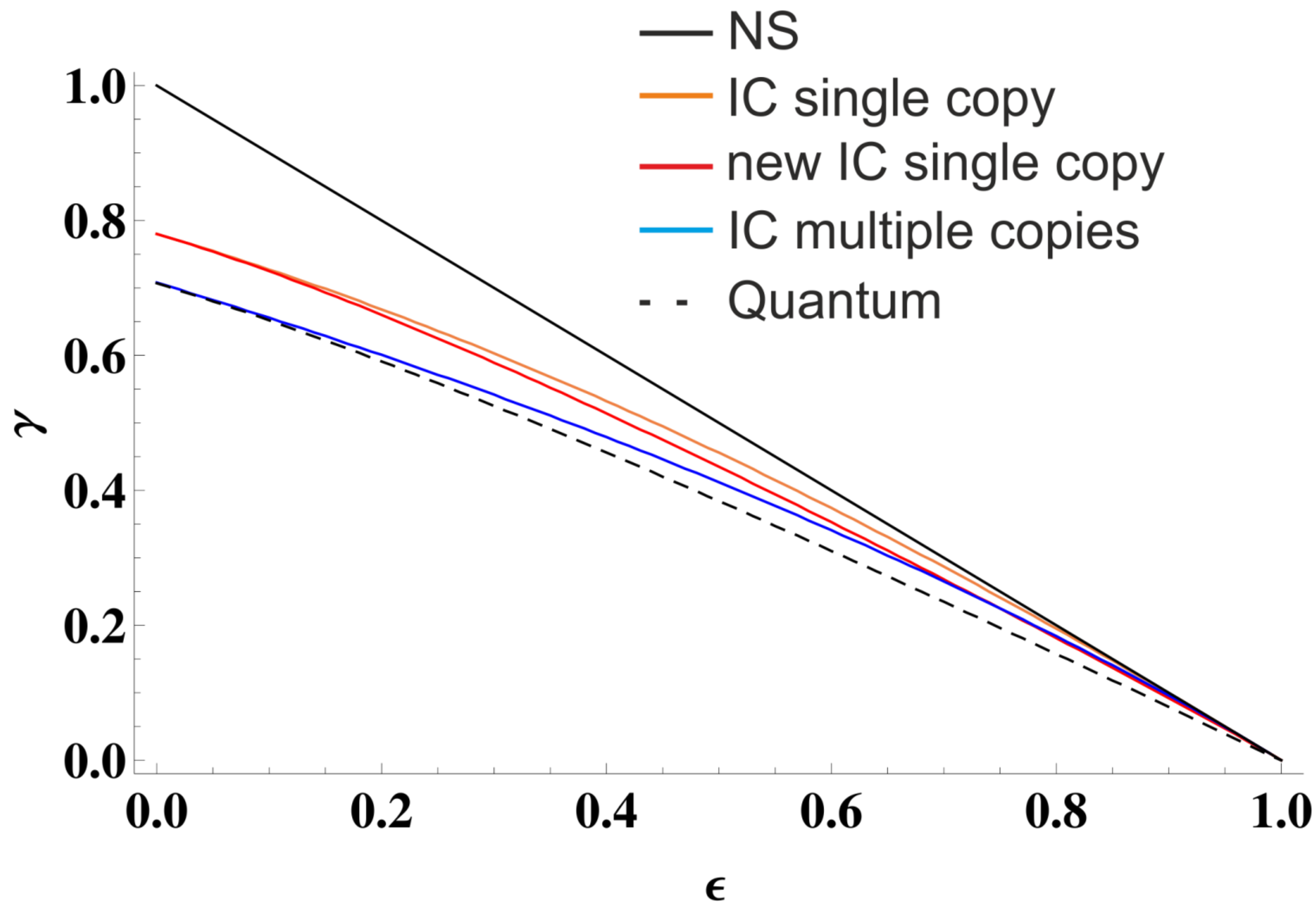
$$I(X_0 : B_0) + I(X_1 : B_1) \leq H(M) + I(X_0 : X_1)$$

$$I(X_0 : B_0) + I(X_1 : B_1) + I(X_0 : X_1|B_1) \leq H(M) + I(X_0 : X_1)$$

Information causality

$$I(X_0 : B_0) + I(X_1 : B_1) \leq H(M) + I(X_0 : X_1)$$

$$I(X_0 : B_0) + I(X_1 : B_1) + I(X_0 : X_1|B_1) \leq H(M) + I(X_0 : X_1)$$

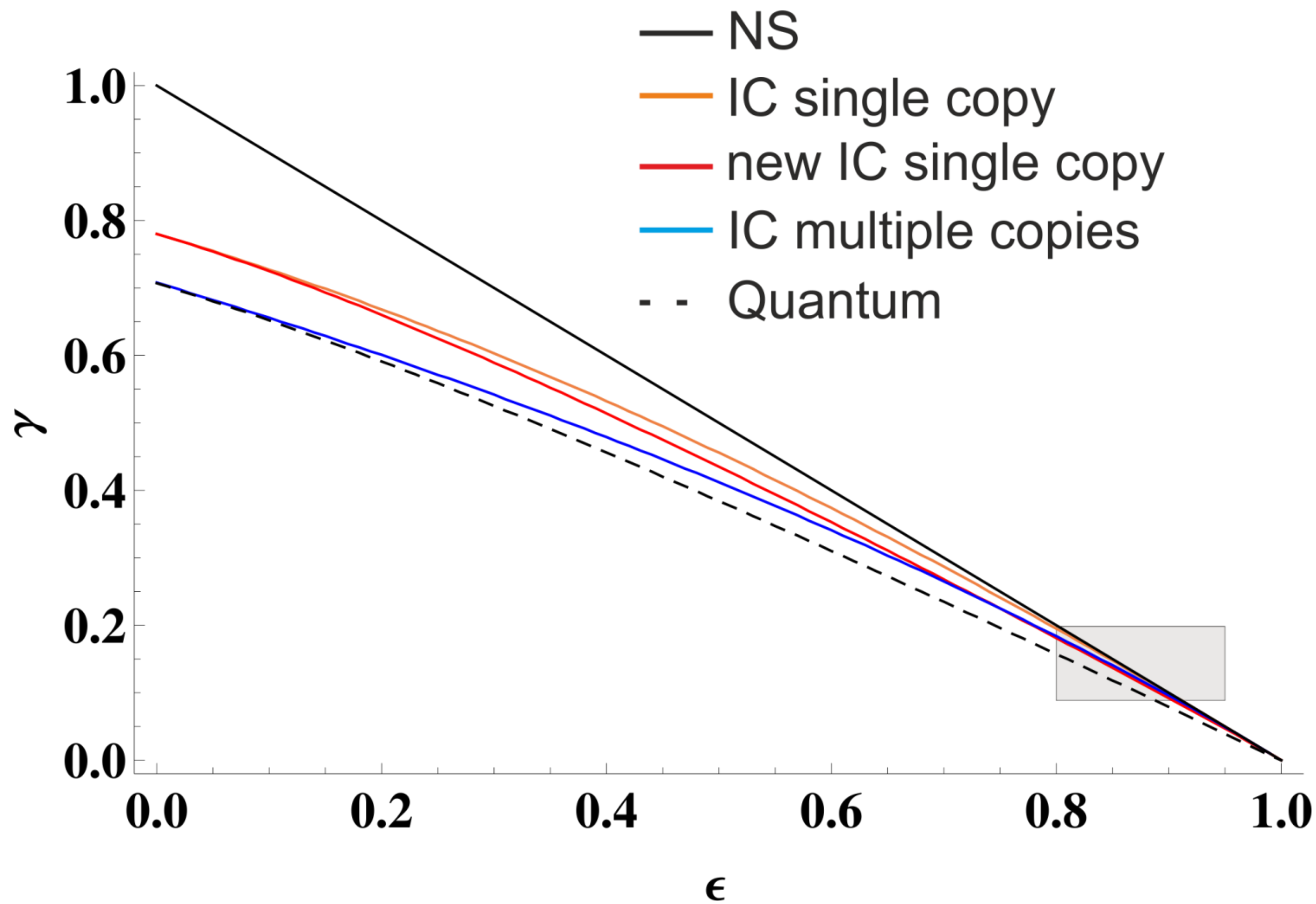


$$p(a, b|x, y) = \gamma P_{PR} + \epsilon P_{det} + (1 - \gamma - \epsilon) P_{white}$$

Information causality

$$I(X_0 : B_0) + I(X_1 : B_1) \leq H(M) + I(X_0 : X_1)$$

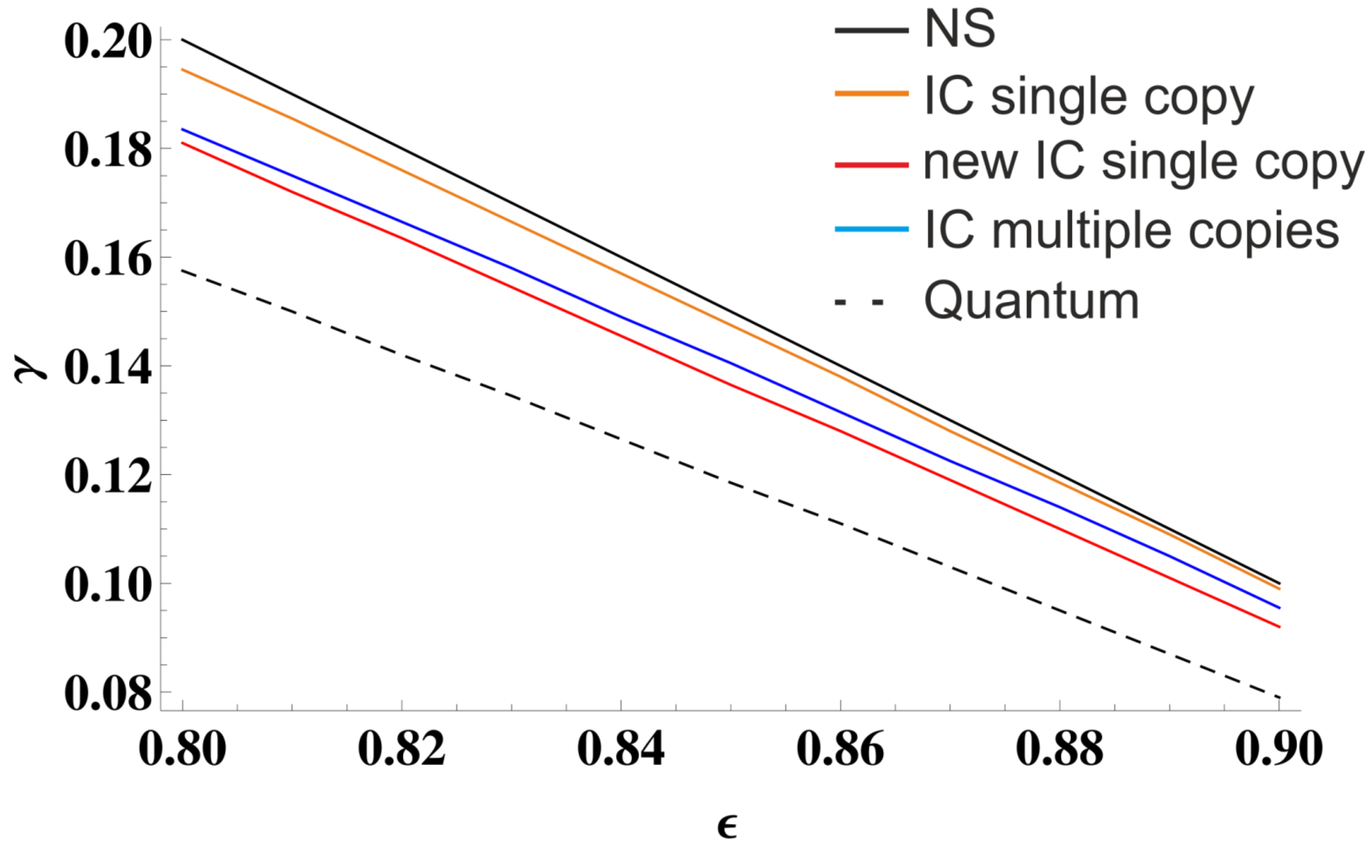
$$I(X_0 : B_0) + I(X_1 : B_1) + I(X_0 : X_1|B_1) \leq H(M) + I(X_0 : X_1)$$



Information causality

$$I(X_0 : B_0) + I(X_1 : B_1) \leq H(M) + I(X_0 : X_1)$$

$$I(X_0 : B_0) + I(X_1 : B_1) + I(X_0 : X_1|B_1) \leq H(M) + I(X_0 : X_1)$$

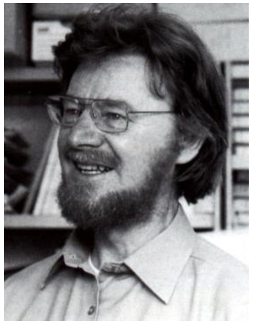


- The new inequality witness the non quantumness of distributions that are not detected by the original one.

- The entropic approach to Causal Inference
- Applications
- Where to go from here?

What we know…

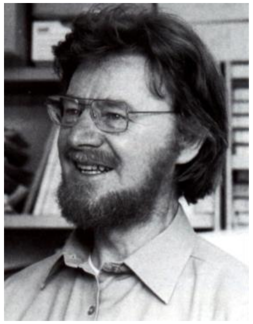
- Entropies allow for a non-trivial, quantitative and operational discrimination between causal relationships...
-  ... both in classical and quantum problems



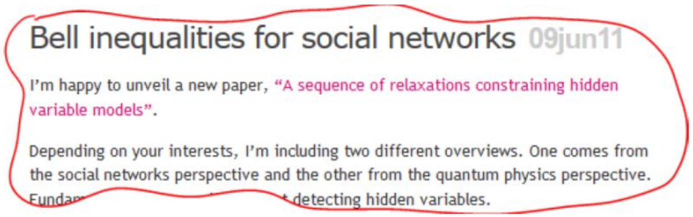
What we know…

- Entropies allow for a non-trivial, quantitative and operational discrimination between causal relationships...

-  ... both in classical and quantum problems



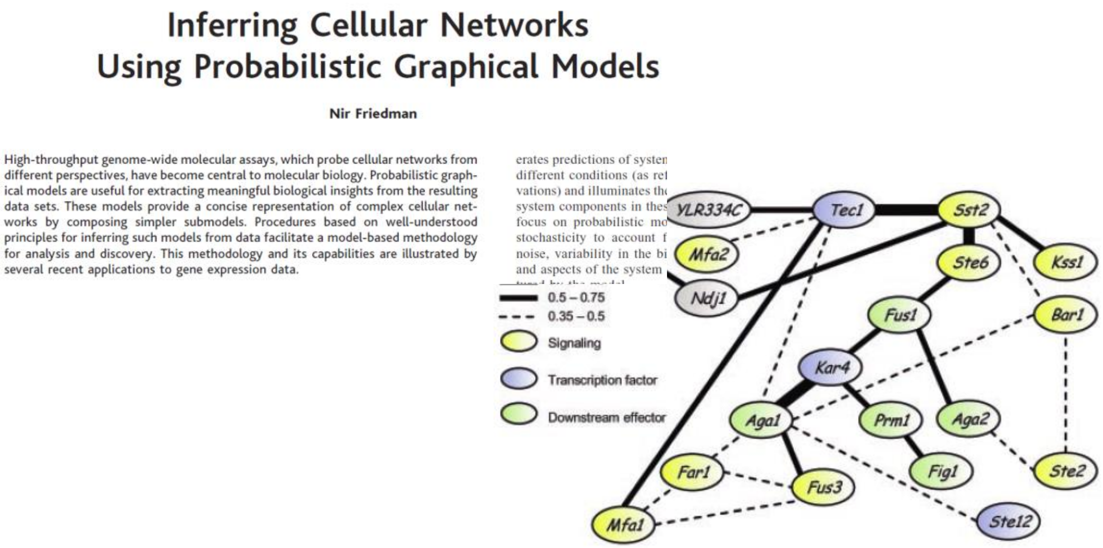
…and what we would like to know



Bell inequalities for social networks 09jun11

I'm happy to unveil a new paper, "A sequence of relaxations constraining hidden variable models".

Depending on your interests, I'm including two different overviews. One comes from the social networks perspective and the other from the quantum physics perspective. Funda… …detecting hidden variables.



Inferring Cellular Networks Using Probabilistic Graphical Models

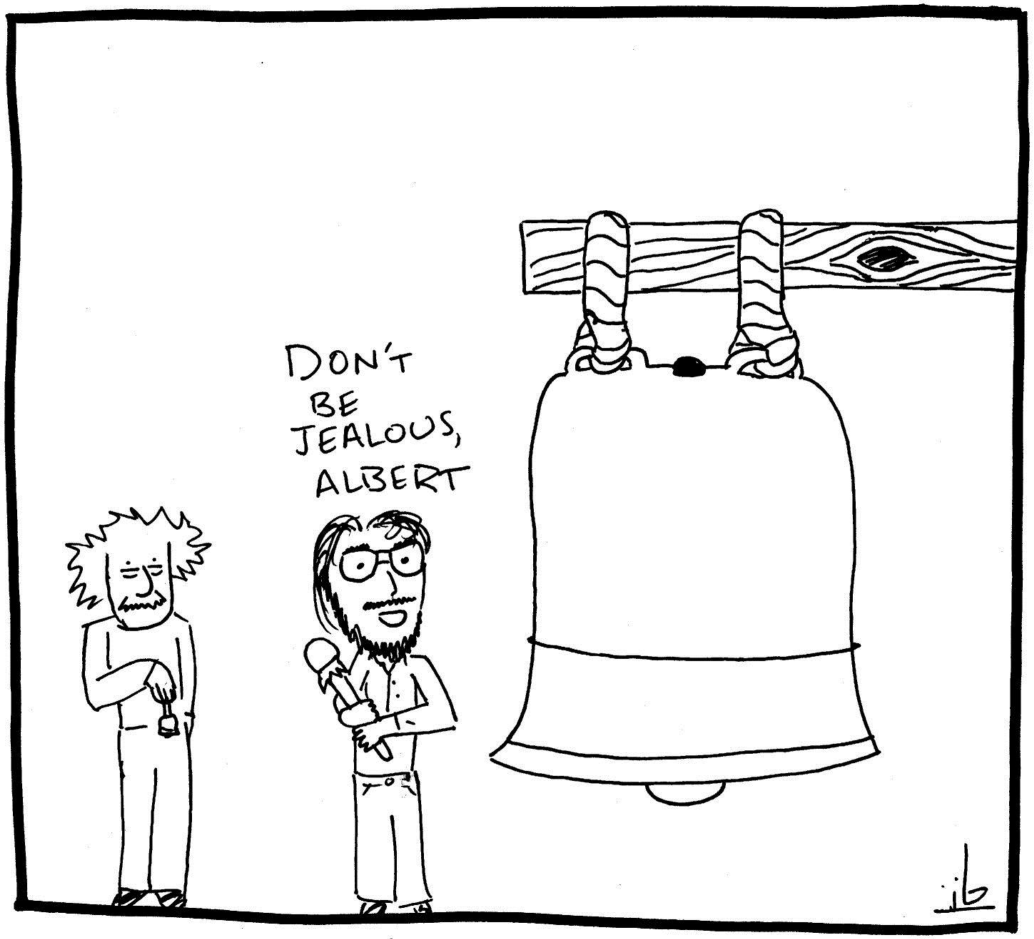
Nir Friedman

High-throughput genome-wide molecular assays, which probe cellular networks from different perspectives, have become central to molecular biology. Probabilistic graphical models are useful for extracting meaningful biological insights from the resulting data sets. These models provide a concise representation of complex cellular networks by composing simpler submodels. Procedures based on well-understood principles for inferring such models from data facilitate a model-based methodology for analysis and discovery. This methodology and its capabilities are illustrated by several recent applications to gene expression data.

- Beyond Bell's theorem?

- New information principles?

Thanks!



BELL'S INEQUALITY