

Logical Properties of Stable Conditional Independence

Mathias Niepert

Dirk Van Gucht

Multivariate Probability Distribution

Smoker	Gender	Lungcancer	...
Y	M	Y	...
Y	F	Y	...
N	F	N	...
...

- S, G, L random variables; $P(L=Y \mid G=M)$?
- Trivial approach requires representation of an exponential number of probabilities (densities)
- Marginal/conditional probabilities difficult to compute („summing out“)

Conditional Independencies

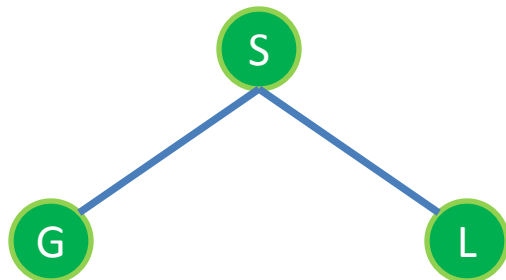
Smoker	Gender	Lungcancer	...
Y	M	Y	...
Y	F	Y	...
N	F	N	...
...

- **Solution:** Use conditional independencies in the data to factorize probability measure
- Naive way: Assume independency among all random variables (Naive Bayes classifier, ...)
- $P(L=Y, G=M) = P(L=Y) P(G=M)$

Graphical Models

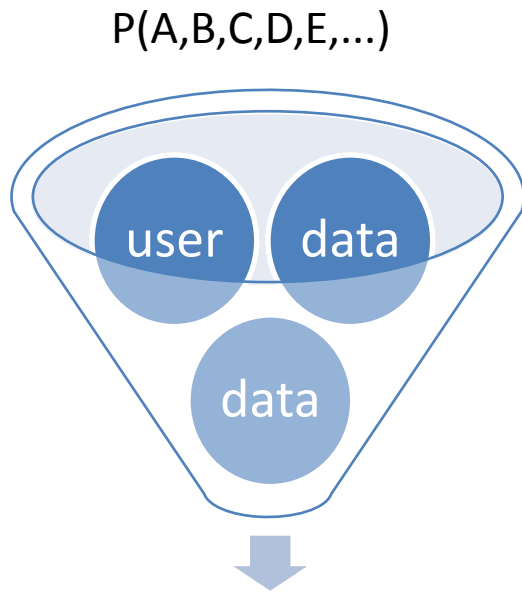
Smoker	Gender	Lungcancer	...
Y	M	Y	...
Y	F	Y	...
N	F	N	...
...

- Better for many applications: Graphical Models
 - Nodes correspond to random variables
 - Edges „encode“ (in-)dependence information



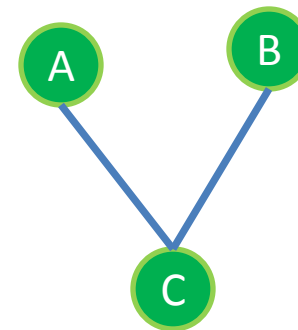
encodes $I(G, L | S)$

Representing and Reasoning about CI



$I(A, B | C), I(E, A | D),$
 $I(B, A | D), \dots$

Representational
Choice



Learning

Yes/No

0.2346

„Query“

$I(E, B | D) ?$



„Query“

$P(A=a, B=b) ?$

Why Alternatives?

- Naive representation of CI structure possibly exponential in number of variables, but also
- Faithfulness problem of graphical models
- Try to find a representation which is
 - Broader (i.e, captures more probability measures) than graphical models but
 - Possibly without the exponential blow-up

What is Stable CI?

- Same as unrestricted CI, except that adding random variables does not create dependency; $I(A,B|C) \rightarrow I(A,B|CD)$
- Early work (Matus 92) called it ascending CI
- Every set of CI statements can be partitioned into stable and unstable part
- Used to reduce size of representation (de Waal & van der Gaag UAI04)

Logical Properties of Stable CI

Axiomatization of stable CI exists (Niepert et al. UAI'08)

$I(A,B|C) \rightarrow I(B,A|C)$ [symmetry]

$I(A,BD|C) \rightarrow I(A,B|C)$ [decomposition]

$I(A,B|CD) \wedge I(A,D|C) \rightarrow I(A,BD|C)$ [contraction]

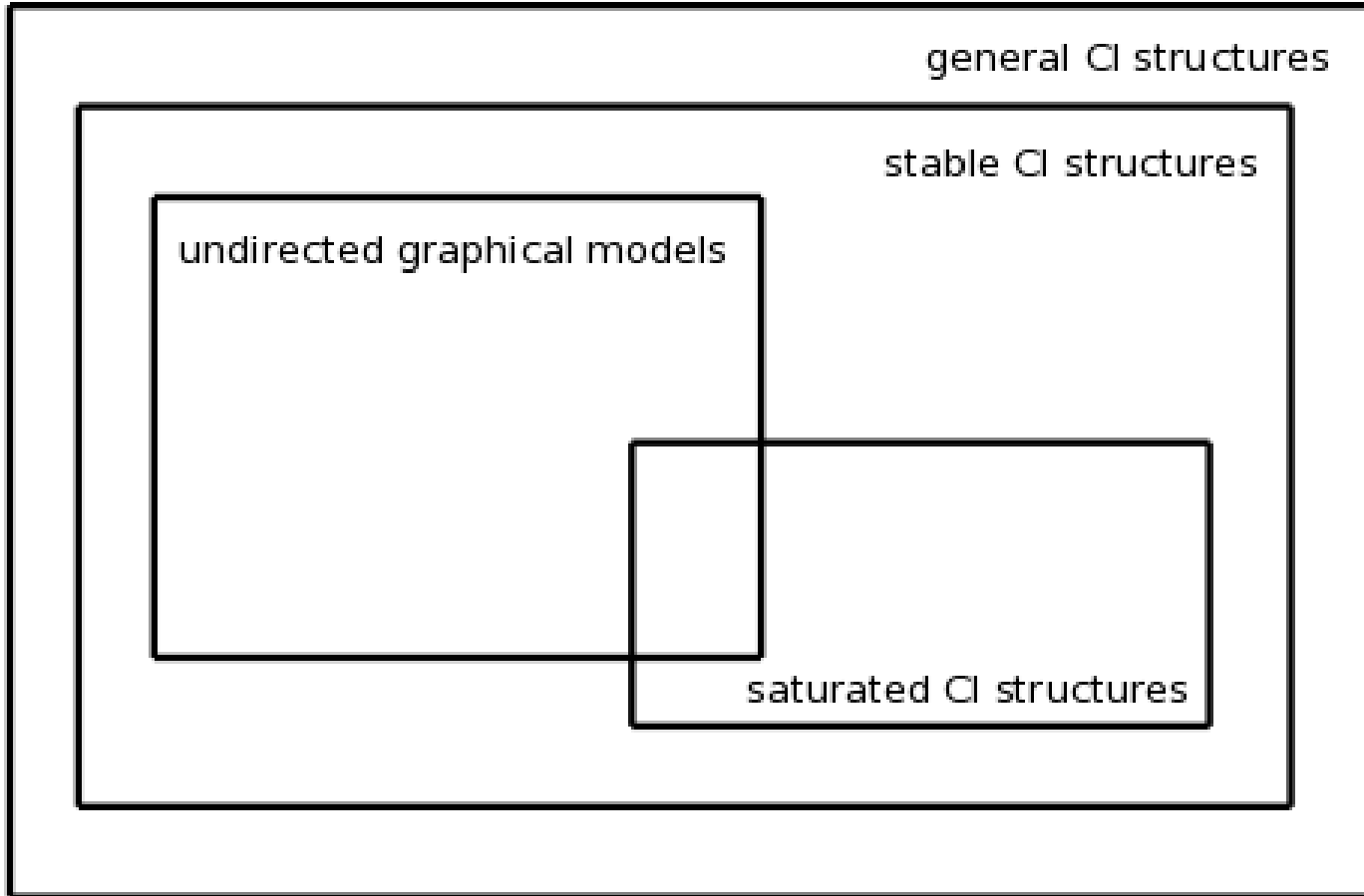
$I(A,B|C) \rightarrow I(A,B|CD)$ [strong union]

$I(A,B|C), I(D,E|AC), I(D,E|BC) \rightarrow I(D,E|C)$
[strong contraction]

Logical Properties of Stable CI

- Generalization of undirected graphical models (follows from axiomatization)
- Number of stable CI structures grows at least double-exponentially with the number of variables (not true for Graphical models)
- For 4 variables there are **64** UG, 18478 general, and at least **4221** stable CI structures

Logical Properties of Stable CI



Logical Properties of Stable CI

- Problem: How can we achieve „lossless compression“ of representation of stable CI?
- Solution in the computational complexity of the implication problem
- Implication problem for stable CI:
 - Given a set of stable CI statements C and a CI statement c . Decide if $C \models c$
 - If so, c is a stable CI statement!

Computational Complexity

- Polynomial time reduction from variant of 3-CNF to the implication problem for stable CI exists
→ Implication problem coNP-complete

In Summary:

Property	Stable CI
Complete finite axiomatization	Yes
Implication Problem	coNP-complete
Perfect models	Yes (but not for binary measures)

Computational Complexity

- Polynomial time reduction from variant of 3-CNF to the implication problem for stable CI exists
→ Implication problem coNP-complete

In Summary:

Property	Stable CI
Complete finite axiomatization	Yes
Implication Problem	coNP-complete
Perfect models	Yes (but not for binary measures)

Logical Properties of Stable CI

- Mapping g from set of CI statements to propositional formula exists (linear time)
- SAT solvers can be used to decide the implication problem

irredundant-subset ($\mathcal{C} : \text{set}$) $\mathcal{C}' : \text{set}$

$\mathcal{C}' := \mathcal{C}$

for each $c \in \mathcal{C}'$

begin

if $g(\mathcal{C}' - \{c\}) \wedge \neg g(c)$ not satisfiable

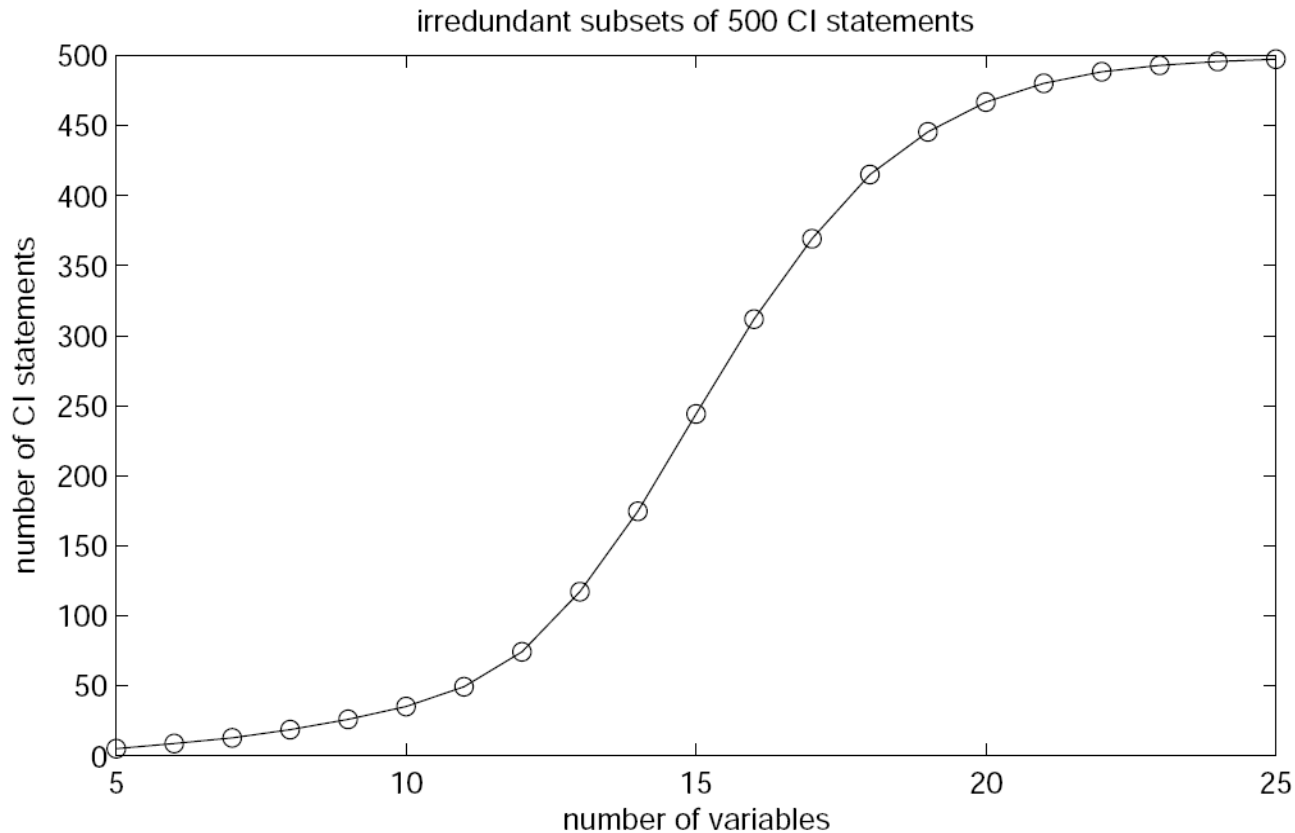
then $\mathcal{C}' := \mathcal{C}' - \{c\}$

end

return \mathcal{C}'

Irredundant Representation

- Experiments
 - Initially 500 randomly generated CI statements

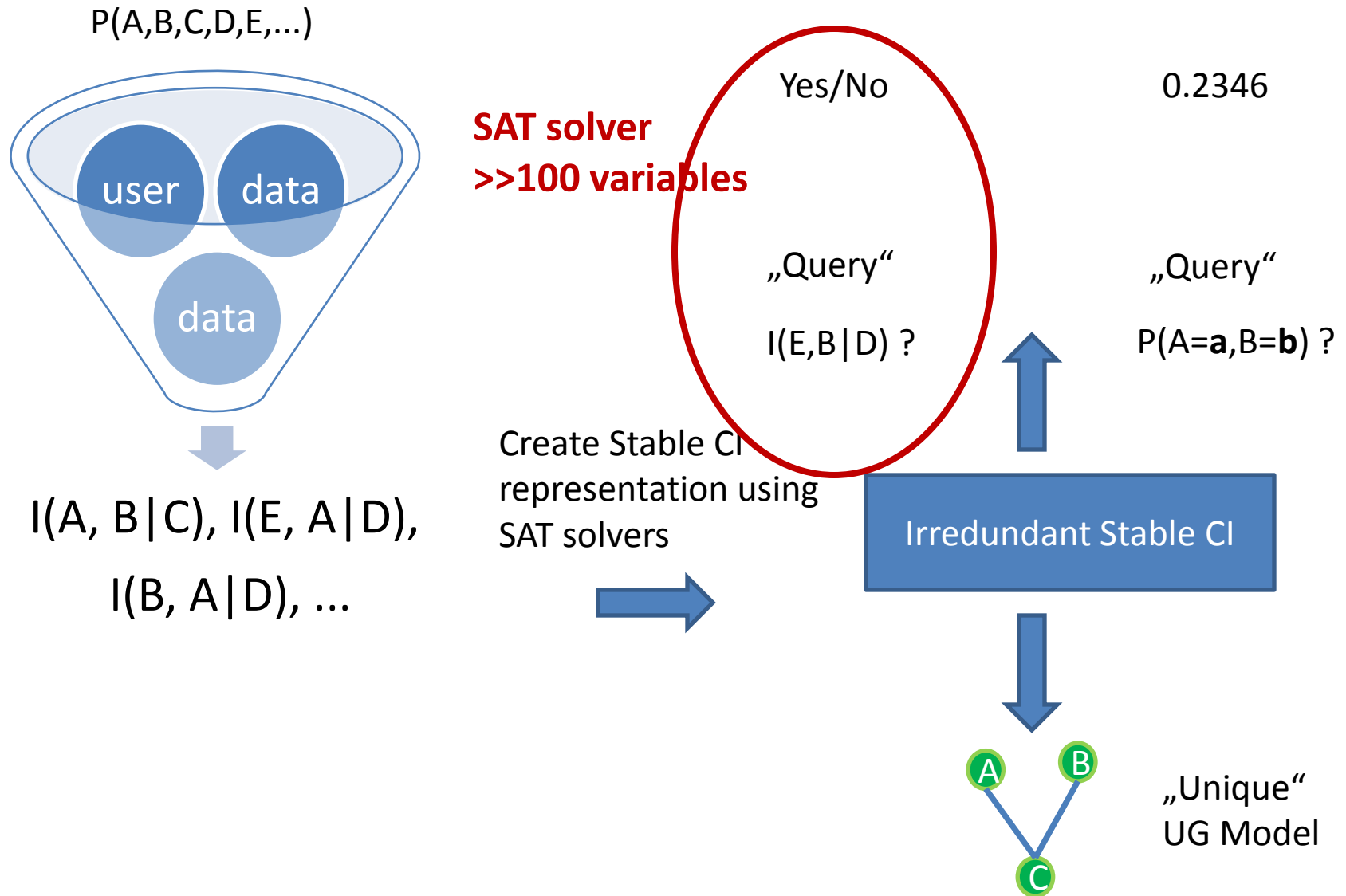


Irredundant Representation

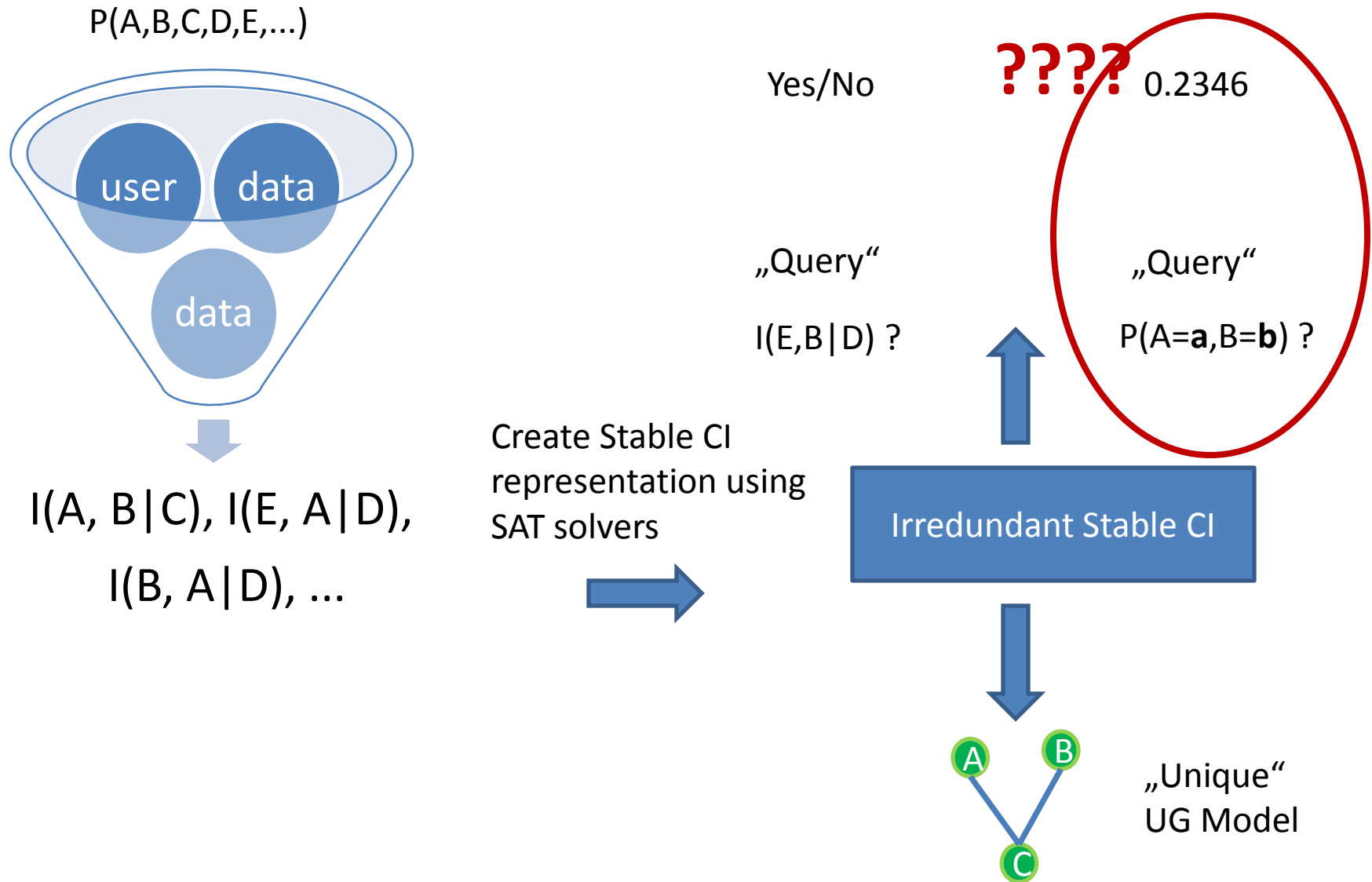
- Experiments (cont.)
 - 100,000 antecedents
 - Randomly generated
 - Time (in ms) to decide the implication problem

Variables	50	100	200	300	400
Time [ms]	740	1523	3362	5627	7076

Representing and Reasoning about CI



Representing and Reasoning about CI



Stable Conditional Independence

- Alternative for representing and reasoning about conditional independence
- Complete, finite axiomatization exists
- Generalization of UG models
- Correspondence to propositional logic
- SAT solver can decide the implication problem

Thank you!