ALESSANDRO ANTONUCCI &
ALESSANDRO FACCHINI

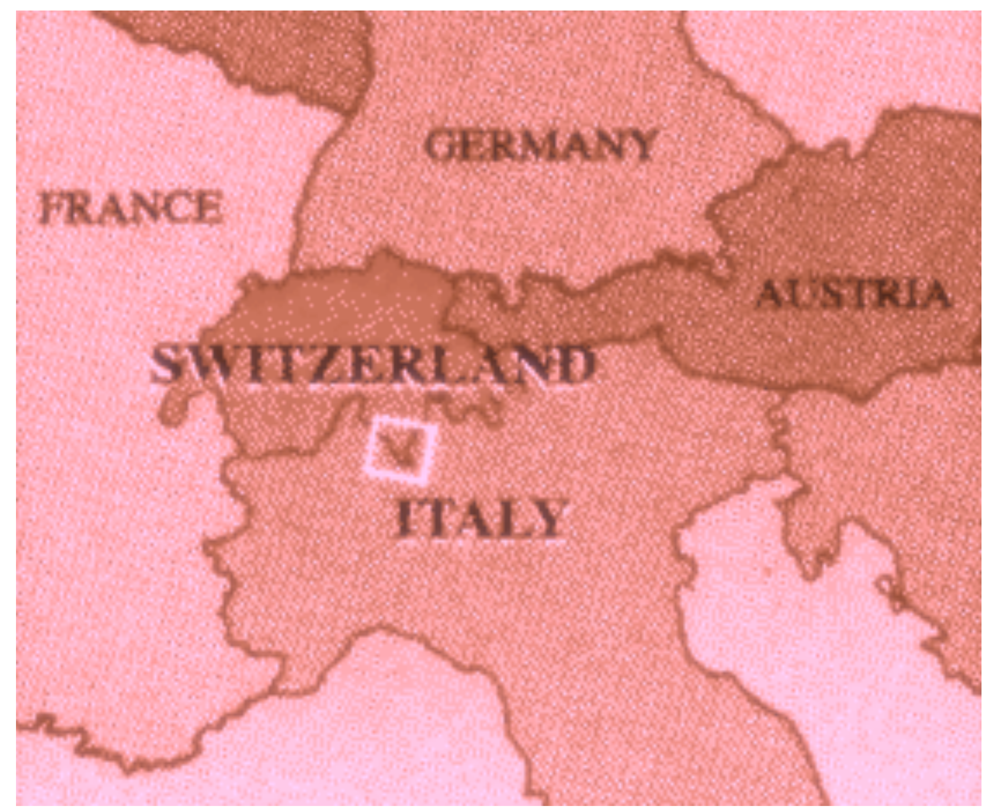
SET-VALUED PROBABILISTIC SENTENTIAL DECISION DIAGRAMS

USI/SUPSI

IDSIA

idsia.ch

Lugano, Ticino, Switzerland



UNSTRUCTURED MACHINE LEARNING (DATA ONLY)

- Graphical models? (E.g.,) Bayesian nets (Pearl, 1984)
- Fast (and deep)? Sum-product nets (Domingos, 2012)
- Robust (and cautious)? Credal nets (Cozman, 2000)
- Deep and robust? Credal SPNs (Mauá et al., 2017)

CN inference harder than BN

(e.g., polytrees vs. binary polytrees),

SPNs/CSPNs less severe transition

STRUCTURED MACHINE LEARNING (DATA + CONSTRAINTS)

- Graphical models? BNs/MRFs with 0-1 potentials (e.g., MLN, Domingos, 2006)
- Deep? PSDDs (Darwiche, 2013)
- Inference in polynomial time wrt the circuit size
- Credal? CSDDs (this paper)
- Two results/algorithms:
  - Marginalisation? Polynomial time wrt the circuit size
  - Conditioning? Polynomial time for singly connected circuits

DATASET (WITH CONSTRAINTS): 100 STUDENTS, 4 SUBJECTS

Logic (L) - Knowledge (K)
Probability (P) - AI (A)

L	K	P	A	#
0	0	1	0	6
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	1	0	0	13
1	1	1	0	8
1	1	1	1	3

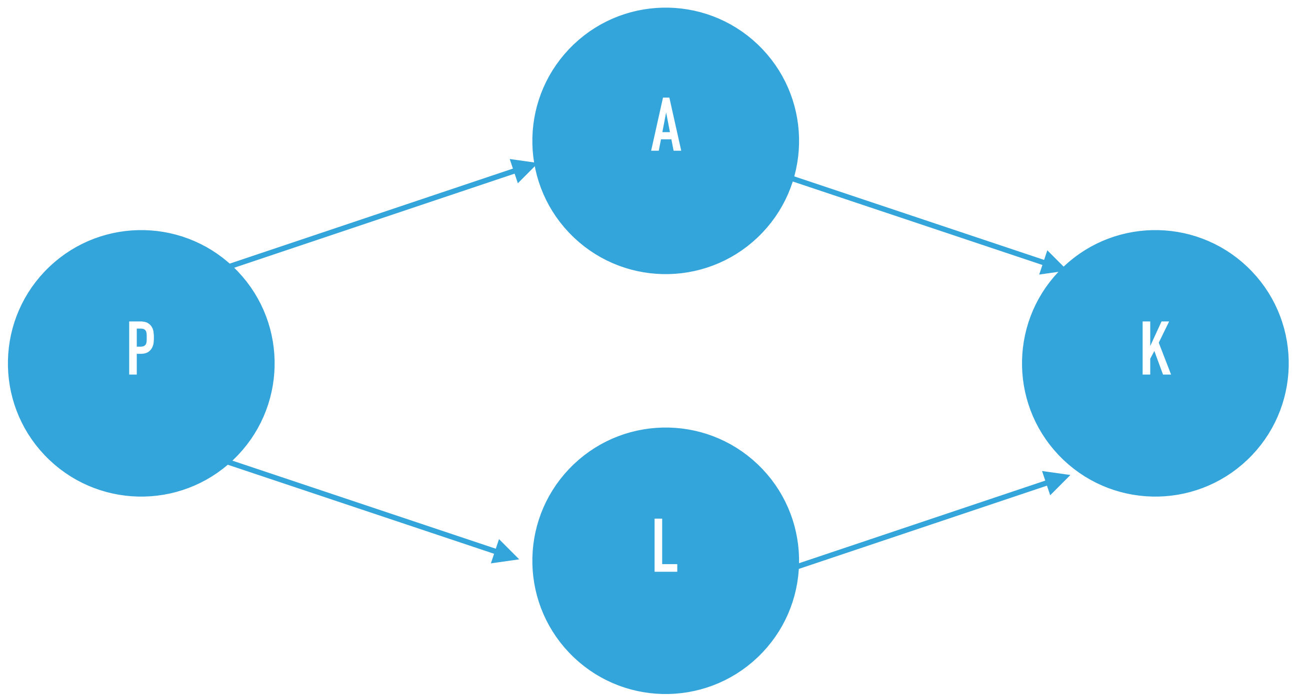
- 16 possible joint states
- 8 observed configurations
- 1 configuration (all zeros) never observed but possible
- 7 impossible configurations
- 3 logical constraints:

$$A \Rightarrow P \qquad K \Rightarrow A \vee L$$

$$P \vee L$$

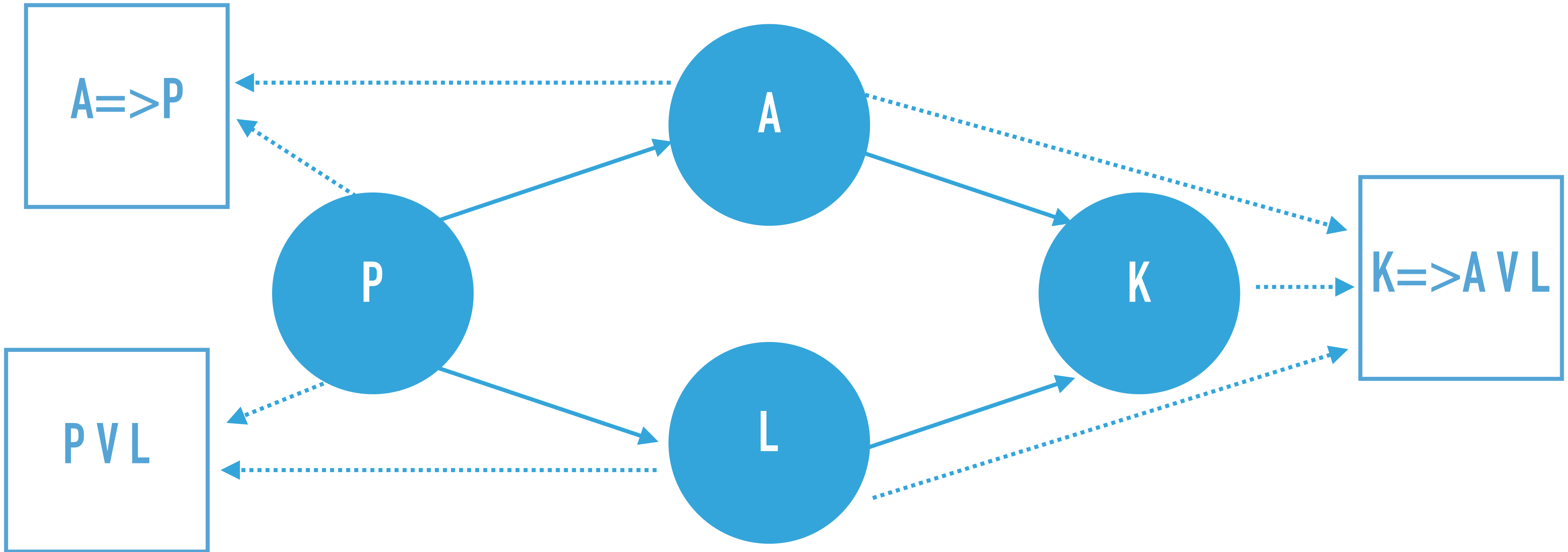
LEARNING BAYESIAN NETS

- Structural learning (e.g., P independent of K given A and L)
- Bayesian learning of parameters $p = \frac{n + 1/2}{N + 1}$
- Joint probability mass function assigning non-zero probability to logically impossible events



LOGICALLY CONSTRAINED BAYESIAN NETS

- Logically impossible events possible?
- Frequentist/maximum likelihood? $p = \frac{n}{N}$
- Zero probability to logically possible events
- Constraints as dummy (and leaf) children? treewidth …

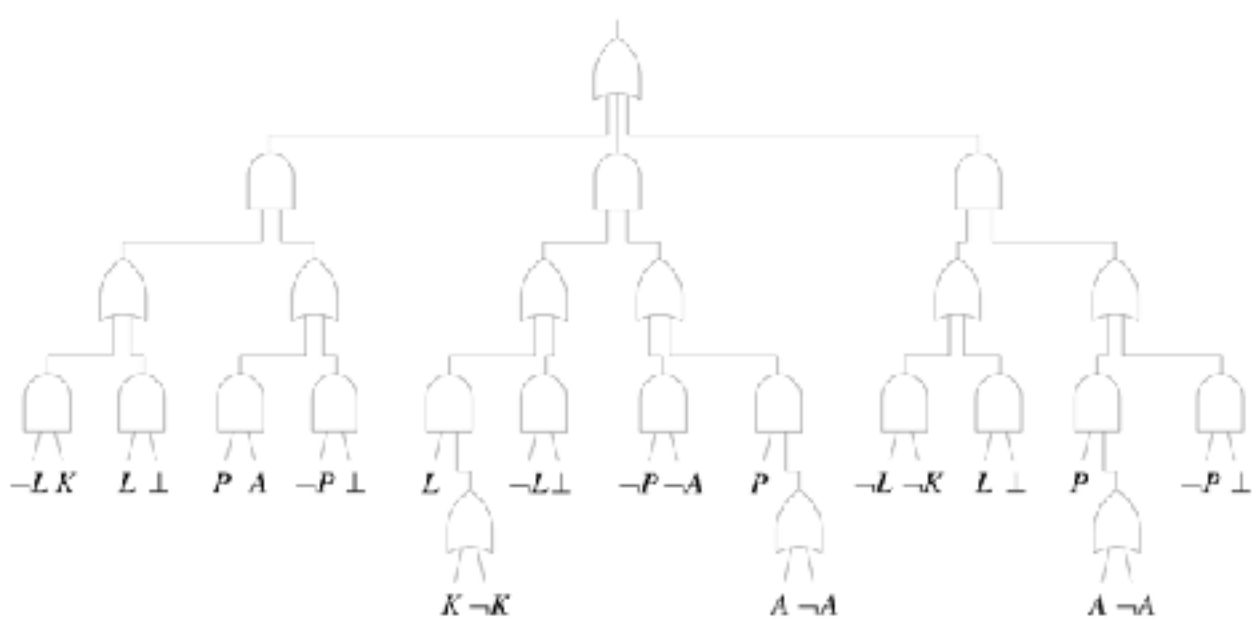


BAYESIAN NETS FROM SMALL DATASETS

- BN parameters are (conditional) probabilities
- p(K=0|L=1,A=1)? three observations with K=1, zero obs with K=0
- Probability zero for the frequentist, p=1/8 for the Bayesian
- Imprecise Probabilities? Convex set of mass functions
- E.g., Walley's IDM (1996) [0,1/4] $\frac{n}{N+1} \leq p \leq \frac{n+1}{N+1}$
- Bayesian nets with interval-valued parameters?
- Credal nets! Harder inferences, but good approx (e.g., Antonucci 2014)
- Small datasets + constraints? Constrained CNs

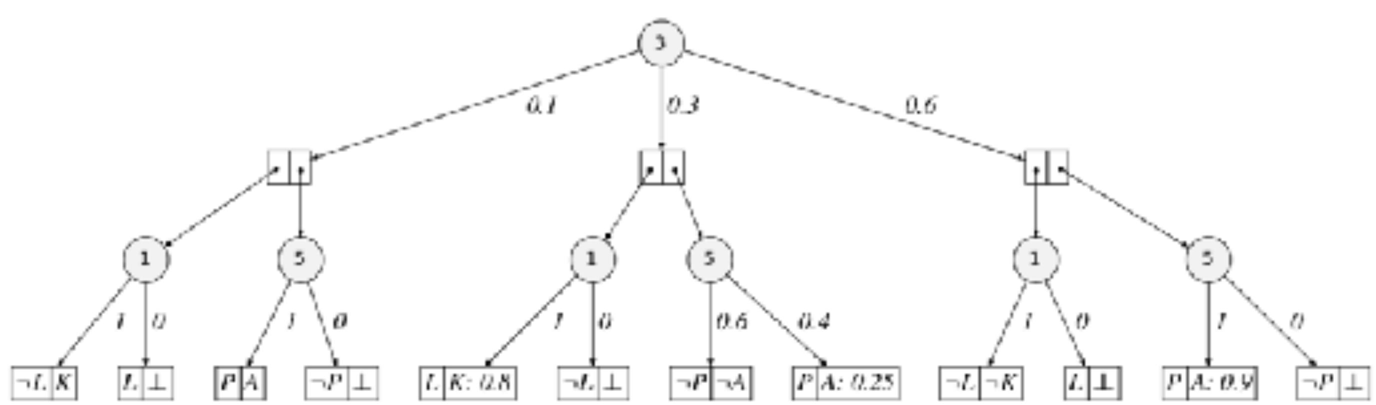
MODELING CONSTRAINTS WITH CIRCUITS

- Sentential Decision Diagrams (SDDs, Darwiche, 2011)
- Logical circuit implementing Boolean formulae
- AND and OR gates alternating
- Single formula corresponds to many SDDs
- Finding smallest SDD hard, but good heuristics (Choi)



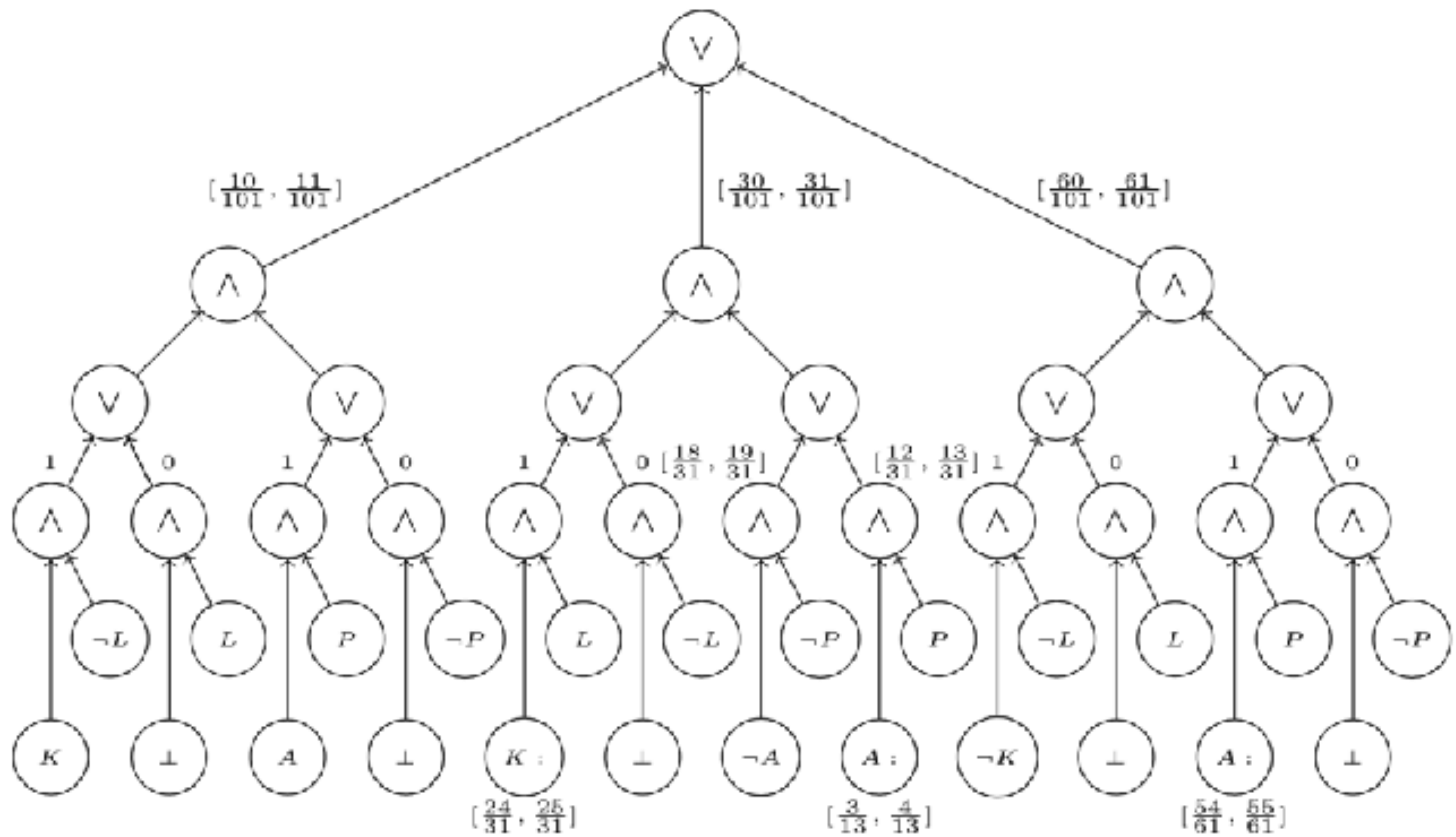
MODELING DATA + CONSTRAINTS WITH CIRCUITS

- Probabilistic Sentential Decision Diagrams (PSDDs)
- SDD with probability mass functions on the OR gates
- Parameters are conditional distributions
- Define a factorised joint mass function by context-specific independence. Fast inference by propagation.
- Impossible events are really impossible!



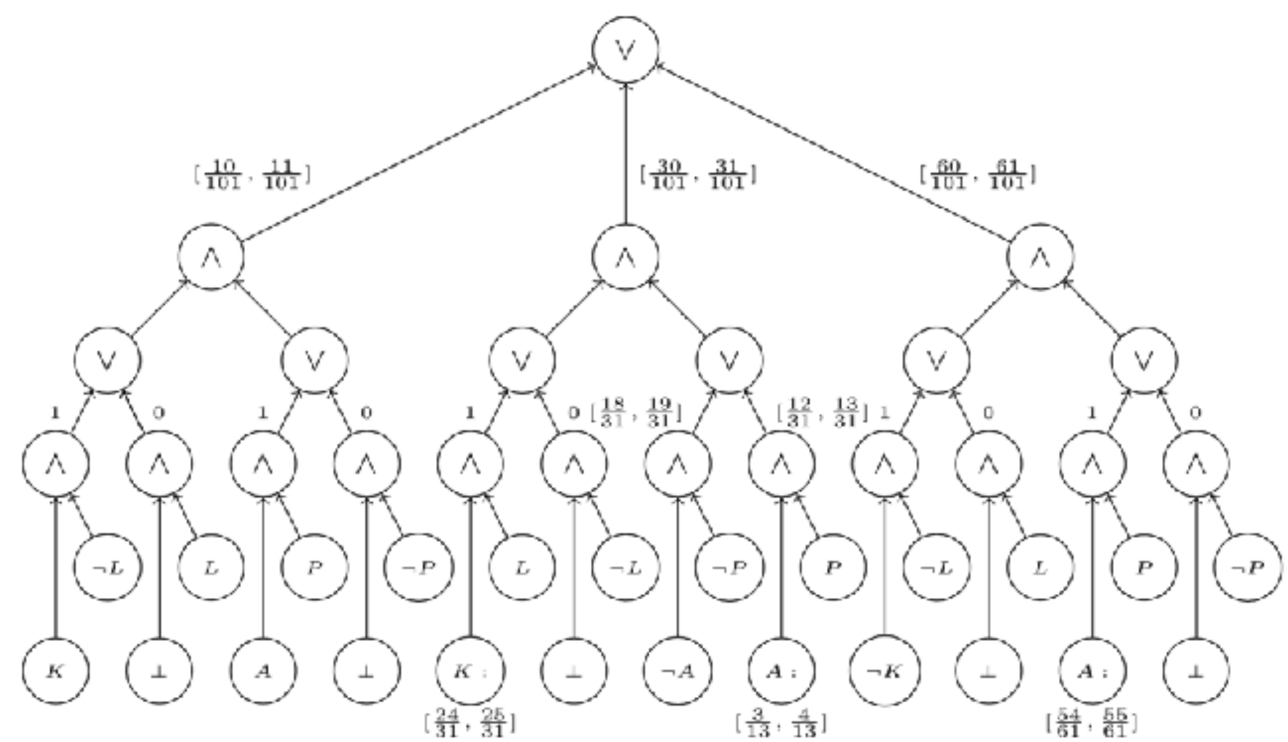
MODELLING SMALL DATA + CONSTRAINTS WITH CIRCUITS

- Credal Sentential Decision Diagrams (CSDDs)
- SDD with interval-probabilities on the OR gates
- CSDD semantics? Collection of consistent PSDDs
- CSDD inference? Lower/upper bounds wrt consistent PSDDs



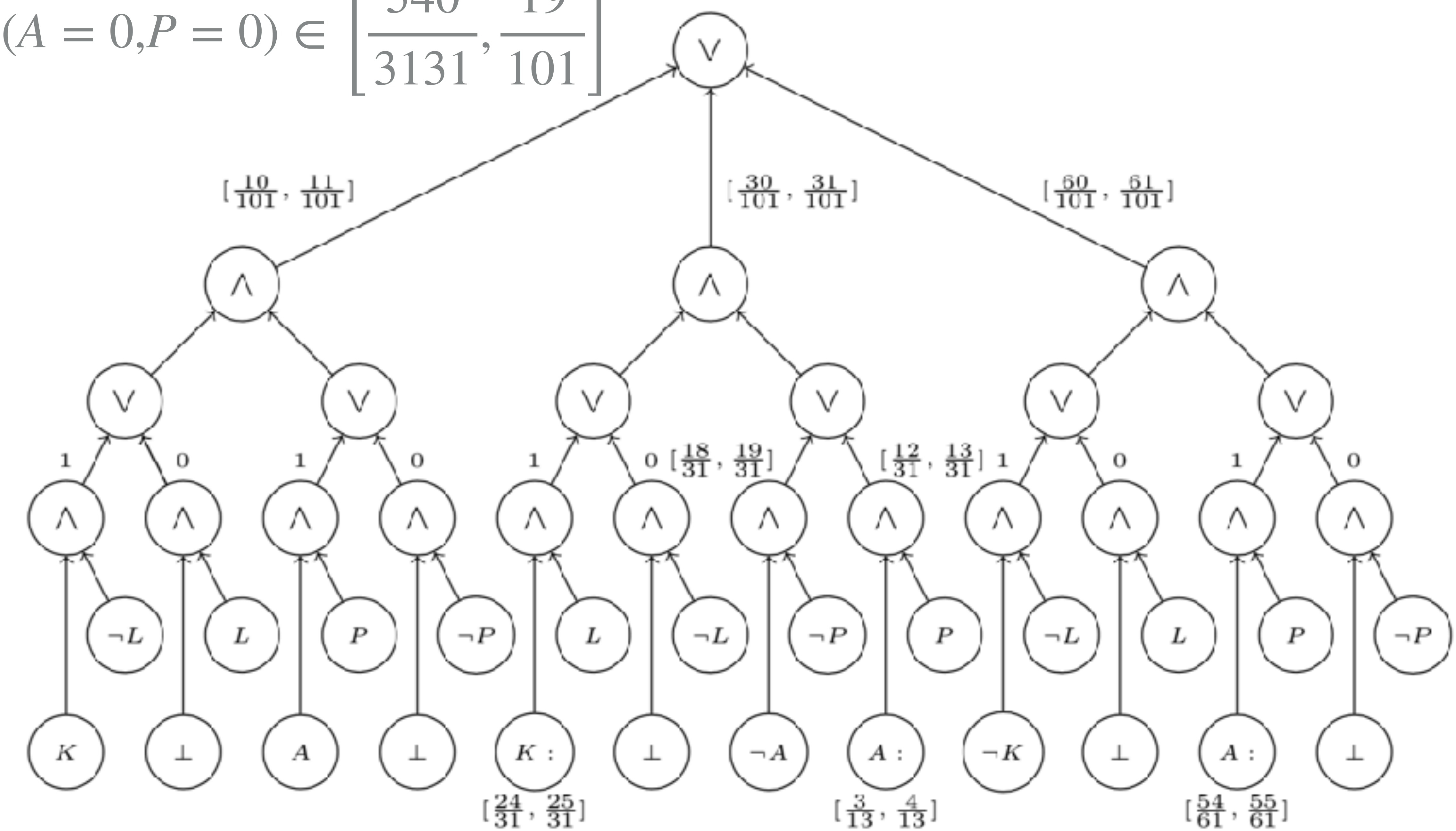
PROPAGATION IN IMPRECISELY ANNOTATED CIRCUITS

- In terms of structure PSDDs = SPNs
- CSPNs + algs by Mauá et al.
- Extending CSPNs algorithms to CSDDs
- Marginalisation? Bottom-up propagation of intervals
- Conditioning? Same with local LP on the OR gates (needs singly connected topology)



PROPAGATION IN IMPRECISELY ANNOTATED CIRCUITS

$$p(A = 0, P = 0) \in \left[\frac{540}{3131}, \frac{19}{101}\right]$$



CONDITIONING BY MARGINALIZATION

BNs and PSDDs (exact)

$$p(A = 0 \,|\, P = 0) = \frac{p(A = 0, P = 0)}{p(P = 0)}$$

CNs and CSDDs (outer approx)

$$\underline{p}(A = 0 \,|\, P = 0) \geq \frac{\underline{p}(A = 0, P = 0)}{\overline{p}(P = 0)}$$

CNs and CSDDs (exact)

$$\underline{p}(A = 0 \,|\, P = 0) > \mu \Leftrightarrow \min_{p} \sum_{a} [I(a) - \mu]\, p(a, P = 0) > 0$$

CONCLUSIONS

- PSDDs as a sound tool for fast structured ML
- CSDDs as a new tool for sensitivity analysis in PSDD
- Robust marginalisation with the same (poly) complexity
- Robust conditioning with the same (poly) complexity for formulae based on singly connected SDDs
- Multiply connected? “Open” the loop (higher complexity)

(LOTS OF) THING TO DO

- Application to "credal" ML with structured spaces (multilabel classification, preference learning, logistics, …)
- Use CSDDs to cope with missing data
- Complexity results for CSDDs
- Extending Choi's library pyPSDD to CSDDs
- Hybrid (structured/unstructured) models
- Structural learning (trade-off small SDD / likelihood)
- …

The 32nd International FLAIRS Conference

Lido Beach Resort

Sarasota, Florida, USA

May 19-22, 2019

Paper submission deadline: November 19, 2018

(see all Important Dates)

The FLORIDA ARTIFICIAL INTELLIGENCE RESEARCH SOCIETY

In cooperation with

Association for the Advancement of Artificial Intelligence