Optimization and Evaluation of Reasoning in Probabilistic Description Logic: Towards a Systematic Approach

Pavel Klinov Bijan Parsia

School of Computer Science
University of Manchester

International Semantic Web Conference,
28th October 2008

Next. . .

Uncertainty: ubiquitous and inevitable

Uncertainty comes in various forms

- Probabilistic uncertainty
  - There is a 99% chance that Pavel will catch his flight tomorrow
- Vagueness (fuzziness)
  - Pavel is young to *some* degree
- Ambiguity, subjectivity, incompleteness, granularity, etc-etc

Work on uncertainty in the context of Semantic Web

- W3C Uncertainty Reasoning for WWW Incubator Group
- Probabilistic OWL
  - Multy Entity Bayesian Networks (PR-OWL)
  - P-$\mathcal{SHIQ}$(D) (Pronto, Contra-Buvem-Rufum)
- Fuzzy DL
  - FiRE, fuzzyDL

What we are doing

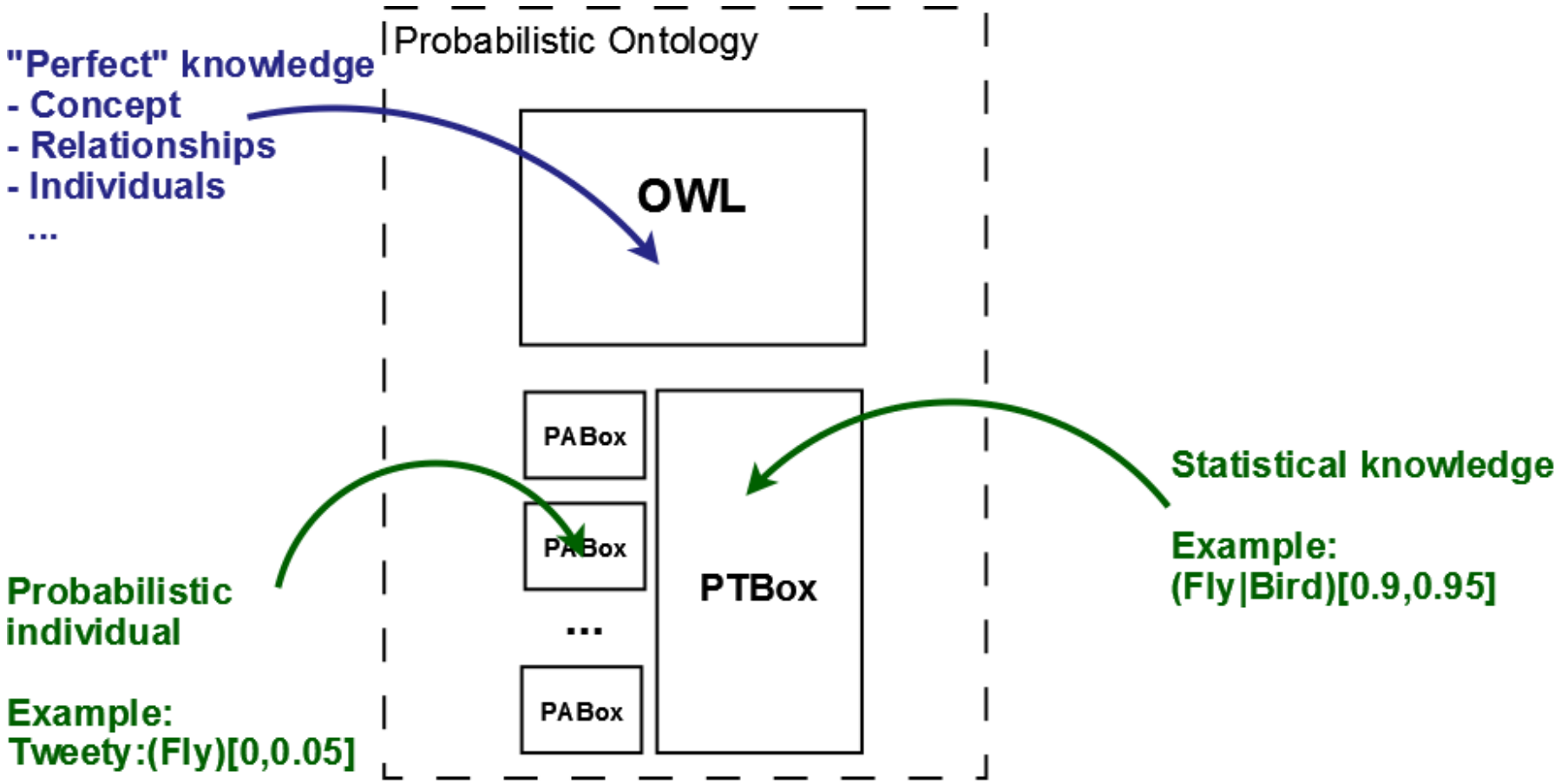
Implementing probabilistic OWL

1. Selected a formalism — P-$\mathcal{SHIQ}$(D)
   - Expressive
   - Nicely integrates with OWL
   - Probabilistically sound
2. Implemented it
   - Reasoner (Pronto)
   - Browser (OWLSight)
3. **Work in progress**
   - **Testing and evaluation (PREVAL-DL)**
   - **Optimization**

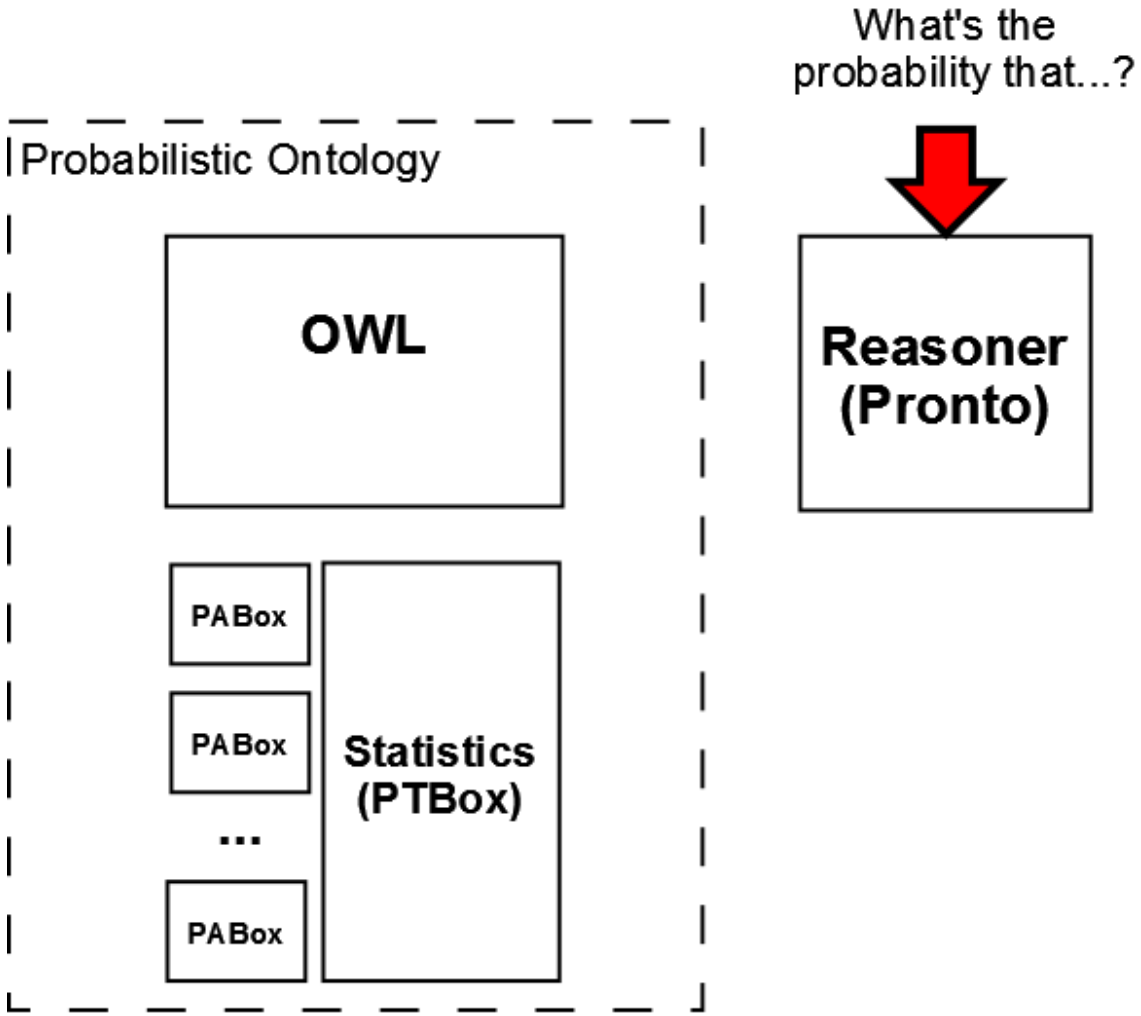
Next...

2. P-$\mathcal{SHIQ}$(D)— probabilistic OWL

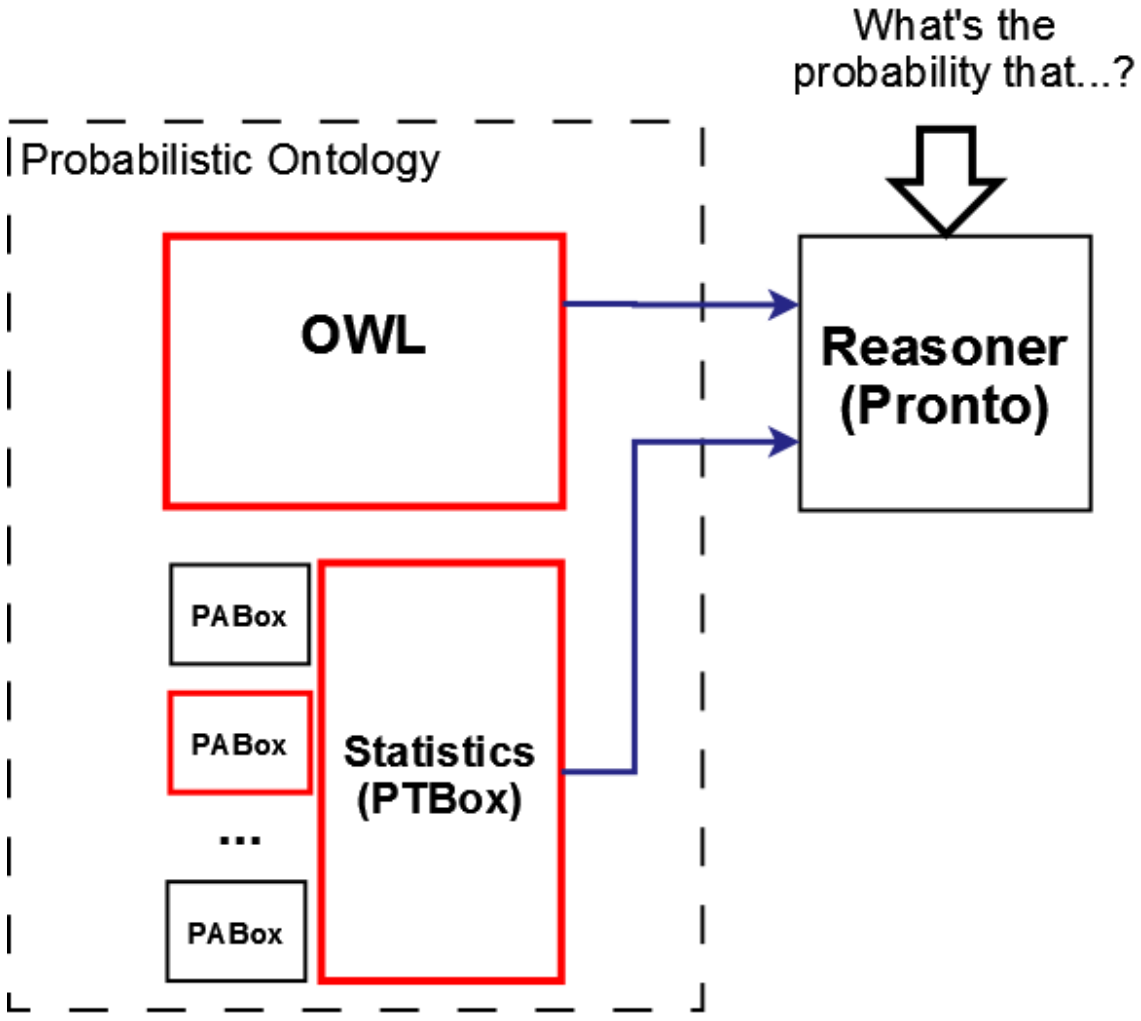
Knowledge Representation and Reasoning in P-$\mathcal{SHIQ}$(D)



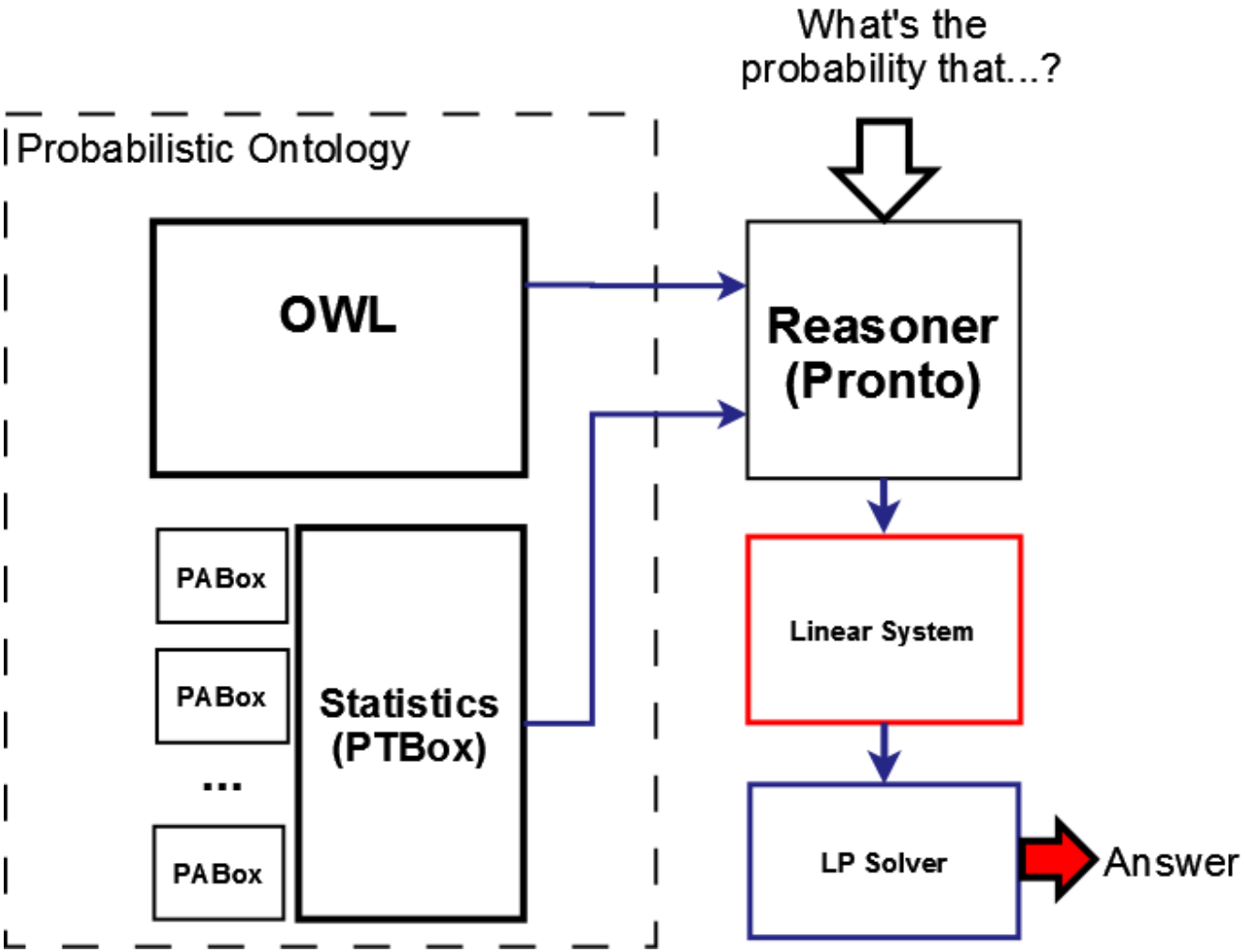
Knowledge Representation and Reasoning in P-$\mathcal{SHIQ}$(D)



Knowledge Representation and Reasoning in P-$\mathcal{SHIQ}$(D)



Knowledge Representation and Reasoning in P-$\mathcal{SHIQ}$(D)



Reasoning in P-$\mathcal{SHIQ}$(D)

Reasoning tasks

- Satisfiability (**PSAT**). Does KB have a probabilistic model?
- Tight entailment (**TEnt**): $KB \models (D|C)[?, ?]$

Defaults and non-monotonicity

- Reasoning is non-monotonic. Tweety is likely to fly *unless* it happens to be a penguin
- Results coincide with standard (monotonic) logical entailment in the absense of *contradictions* (e.g. Sam is a flying penguin)

Here are the problems

Linear systems can be exponentially large

- Size of linear systems $= M \times N$
  - M (number of inequalities) — polynomial in the size of KB
  - N (number of variables) — exponential in the size of KB
- Need *at least* N classical DL SATs just to create the system!

Linear systems can be exponentially many

- Reasoning algorithm entails results from the most preferred models (as opposed to all models)
- Computation of the most preferred models may lead to exponentially many PSATs

Next...

3. Performance evaluation methodology

Why optimization is difficult

When you have to deal with EXPTIME-Complete problem...

- Theoretical complexity isn't the last word, so:
  - Define *typical* problem instances
  - Make the algorithm to work fast on them
- In order to do that you need to:
  - Have some characteristic test data
  - Understand what makes hard cases hard
  - Ideally be able to estimate hardness beforehand

No condition *was* satisfied for P-$\mathcal{SHIQ}$(D)

- No characteristic ontologies, benchmark suites
- No canonical reasoning problem (like classification in DL)

Our approach

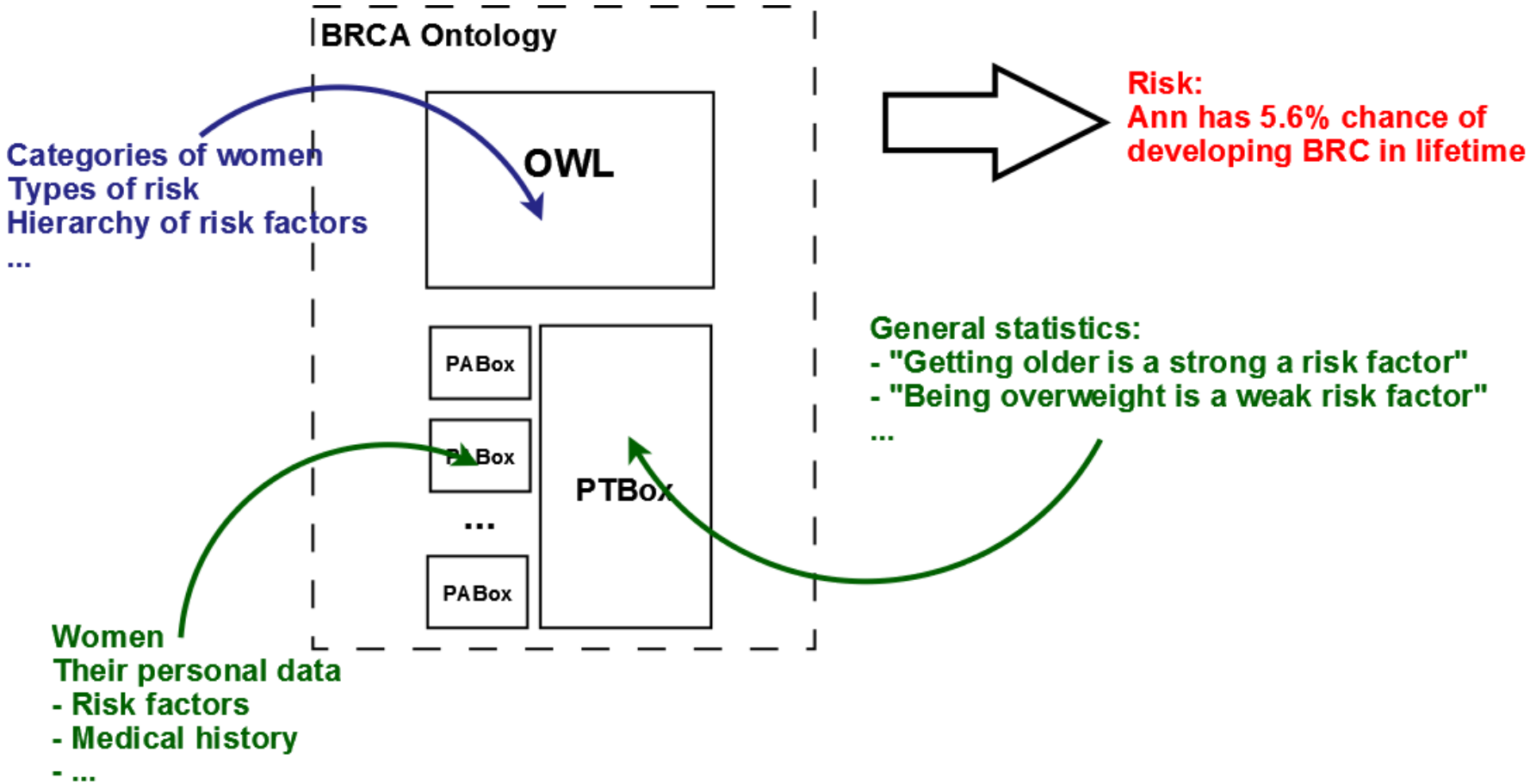
We developed the first realistic P-$\mathcal{SHIQ}$(D) ontology prototype

- It models the Breast Cancer Risk Assessment problem (BRCA)
- It is used as a main dataset for the evaluation
  - It is too hard for the existing P-$\mathcal{SHIQ}$(D) reasoners
  - We believe that it can be "new GALEN"

We developed a systematic evaluation methodology

1. Proposed a reasoner-independent approach to generate series of P-$\mathcal{SHIQ}$(D) reasoning problems (PSAT, TEnt)
2. Implemented it (PREVAL-DL)
3. Applied to Pronto

Breast Cancer Risk Assessment (BRCA) ontology



Evaluation methodology: sampling

- There are several possible testing and evaluation strategies:
  1. Generate typical problem instances
  2. Generate only hard instances
  3. Re-use existing instances (those that occurred in practice)
- We had no choice (no existing instances, no understanding of what's typical, vague intuitions of what's hard)
- Thus we've picked *random sampling* method

Random sampling

- Each sample is a randomly generated subset of the BRCA ontology
  - Size and number of samples is configurable
  - Size: 10 to 15 probabilistic statements
    Number of samples: 100-200
- From samples to reasoning problem instances:
  - PSAT: sample = problem instance
  - TEnt: also need to generate a reasonable query (i.e. what to infer from the sample)

Evaluation of TEnt

Each TEnt problem instance consists of:

- Fragment of BRCA ontology
  $(BRCA\_in\_10\_years|Age\_over\_50)[0.025, 0.039]$,
  $(BRCA\_in\_lifetime|BRCA1\_gene\_mutation)[0.6, 0.8]$
- Probabilistic individual (i.e. woman) + probabilistic facts
  $(Ann : Age\_over\_50)[1, 1], (Ann : BRCA1\_gene\_mutation)[0.7, 0.8]$
- Query: compute risk of developing breast cancer
  $(Ann : BRCA\_in\_lifetime)[?, ?]$

Evaluation of TEnt

Each TEnt problem instance consists of:

- Fragment of BRCA ontology
  $(BRCA\_in\_10\_years|Age\_over\_50)[0.025, 0.039]$,
  $(BRCA\_in\_lifetime|BRCA1\_gene\_mutation)[0.6, 0.8]$
- Probabilistic individual (i.e. woman) + probabilistic facts
  $(Ann : Age\_over\_50)[1, 1]$, $(Ann : BRCA1\_gene\_mutation)[0.7, 0.8]$
- Query: compute risk of developing breast cancer
  $(Ann : BRCA\_in\_lifetime)[?, ?]$

Evaluation of TEnt

Each TEnt problem instance consists of:

- Fragment of BRCA ontology
  $(BRCA\_in\_10\_years|Age\_over\_50)[0.025, 0.039]$,
  $(BRCA\_in\_lifetime|BRCA1\_gene\_mutation)[0.6, 0.8]$
- Probabilistic individual (i.e. woman) + probabilistic facts
  $(Ann : Age\_over\_50)[1, 1], (Ann : BRCA1\_gene\_mutation)[0.7, 0.8]$
- Query: compute risk of developing breast cancer
  $(Ann : BRCA\_in\_lifetime)[?, ?]$

Evaluation of TEnt

Each TEnt problem instance consists of:

- Fragment of BRCA ontology
  $(BRCA\_in\_10\_years|Age\_over\_50)[0.025, 0.039]$,
  $(BRCA\_in\_lifetime|BRCA1\_gene\_mutation)[0.6, 0.8]$
- Probabilistic individual (i.e. woman) + probabilistic facts
  $(Ann : Age\_over\_50)[1, 1], (Ann : BRCA1\_gene\_mutation)[0.7, 0.8]$
- Query: compute risk of developing breast cancer
  $(Ann : BRCA\_in\_lifetime)[?, ?]$

Complexity metrics

What is hard and what is easy?

- In addition to measuring performance we would also like to:
  - Understand why some samples are much harder than others
  - Be able to predict (i.e. estimate beforehand) such hardness
  - This might lead to new optimization strategies
- Complexity metric — function that estimates the complexity of a KB *without* reasoning

Simple "connectivity" metric

- Linear systems get bigger when classes in conditional constraints are unrelated
- Our metric measures this "connectivity" — the total number of subsumptions between classes in conditional constraints

Complexity metrics

What is hard and what is easy?

- In addition to measuring performance we would also like to:
  - Understand why some samples are much harder than others
  - Be able to predict (i.e. estimate beforehand) such hardness
  - This might lead to new optimization strategies
- Complexity metric — function that estimates the complexity of a KB *without* reasoning

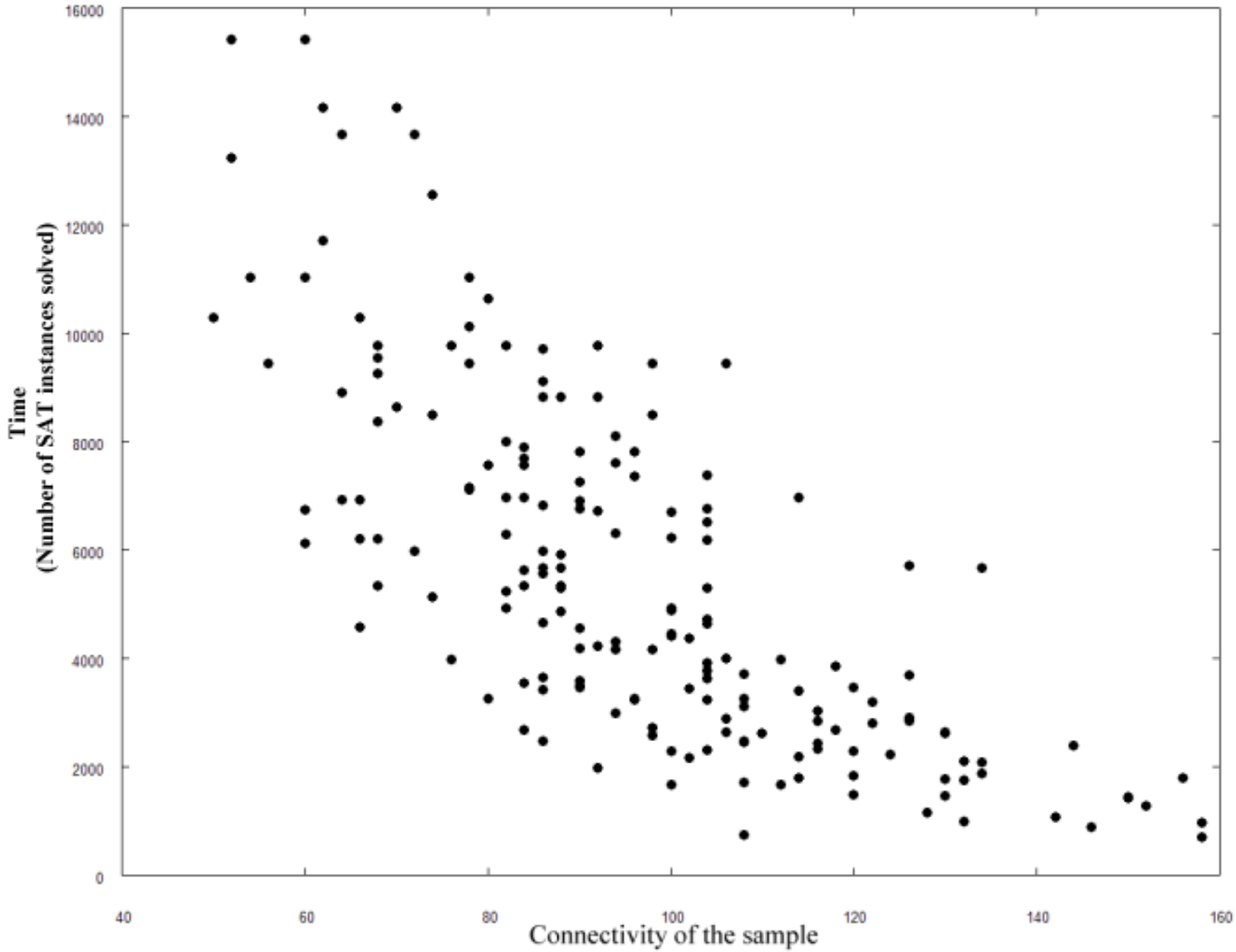
Simple "connectivity" metric

- Linear systems get bigger when classes in conditional constraints are unrelated
- Our metric measures this "connectivity" — the total number of subsumptions between classes in conditional constraints

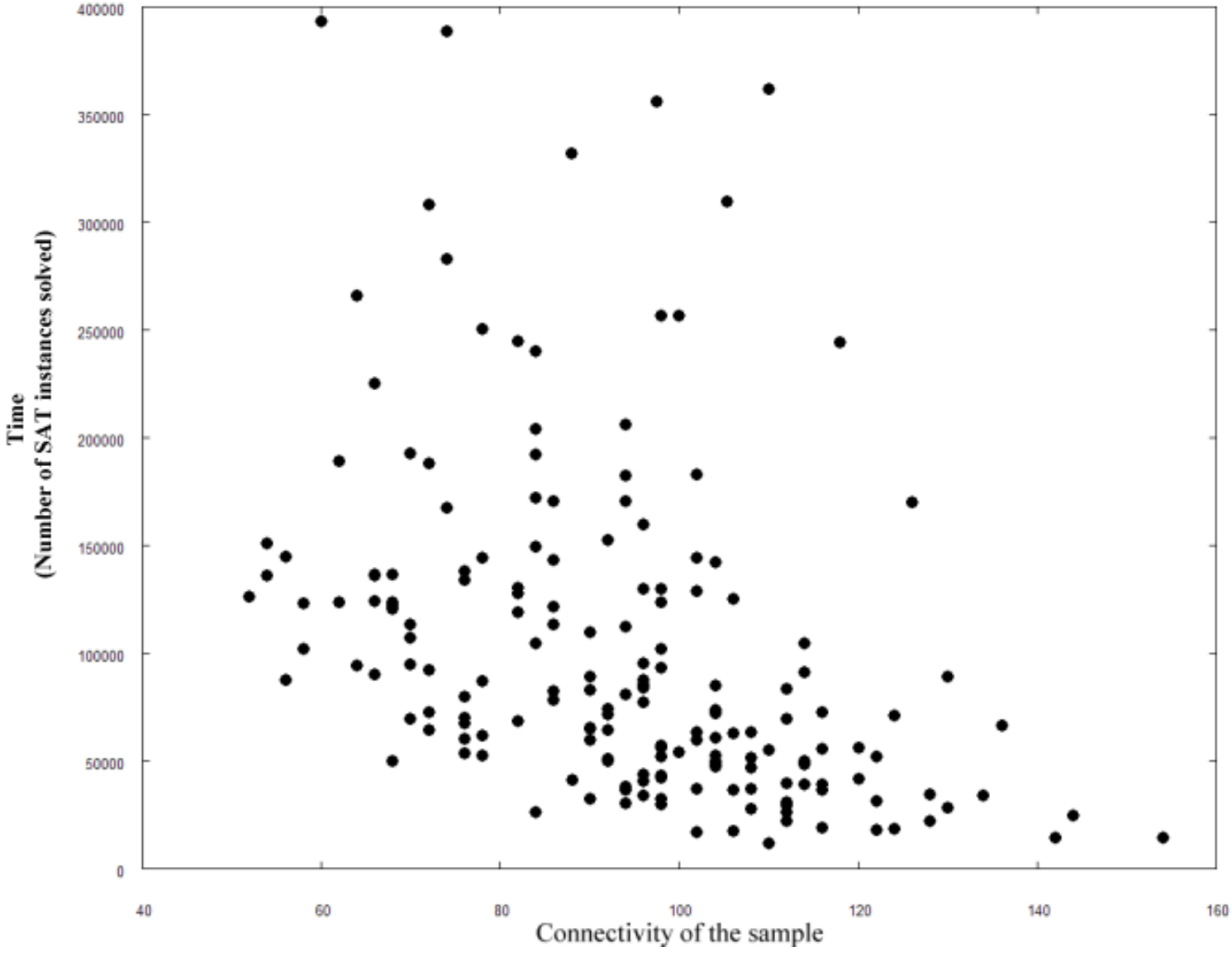
Next...

4 Pronto evaluation results

PSAT evaluation results



Naive TEnt evaluation results

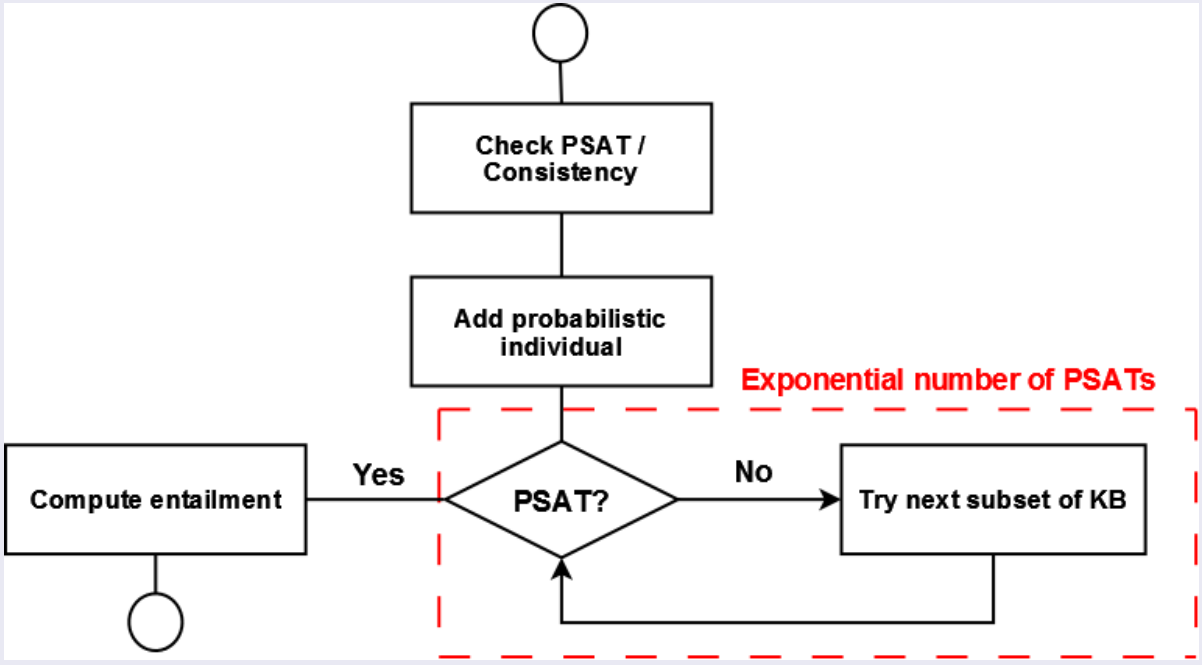


Discussion

TEnt results seem strange?

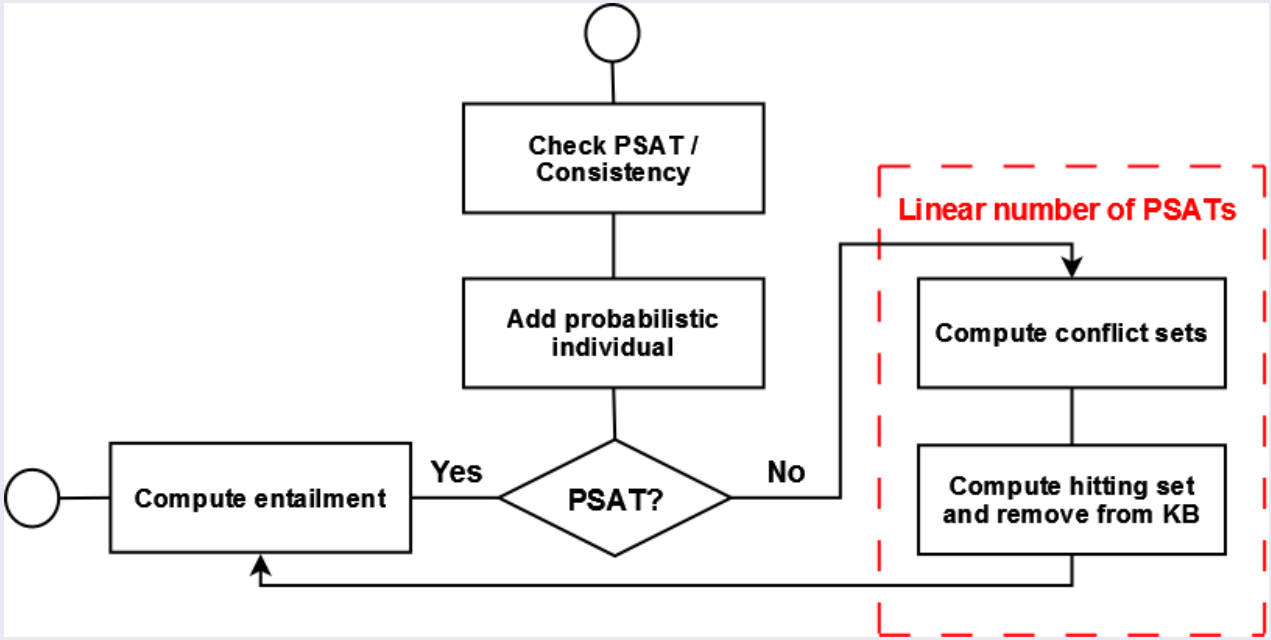
- There are many "outliers"
- The metric is much less predictable

Naive TEnt algorithm



TEnt optimization

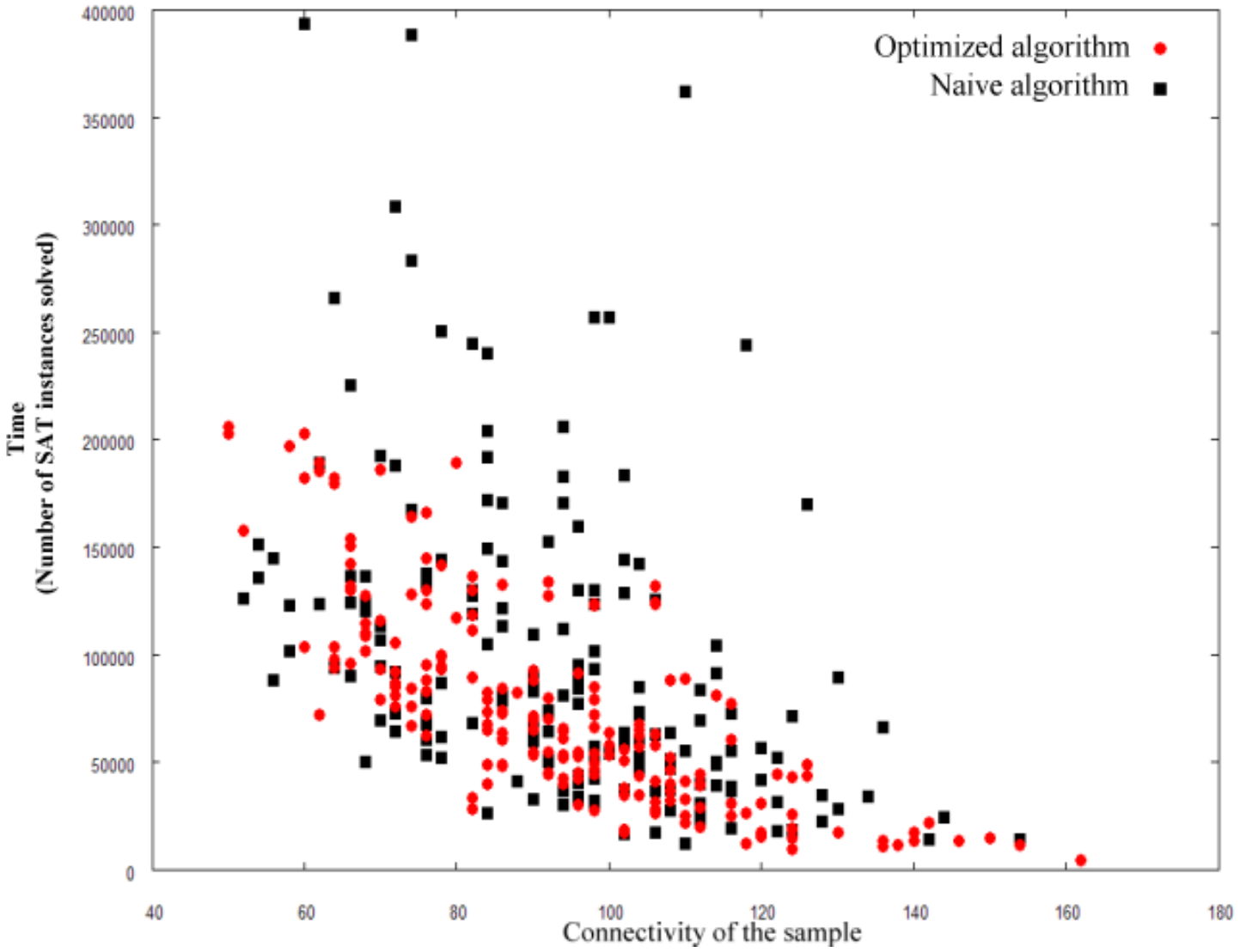
Fast TEnt algorithm



Advantages

- Fast computation of preferred models
- Faster, more robust and predictable

Comparison of two TEnt algorithms



Next...

5 Summary, contributions and future directions

Contributions

PREVAL-DL: Evaluation suite for P-$\mathcal{SHIQ}$(D) reasoners

- PREVAL-DL is an open source Java library that generates series of reasoning problems to evaluate reasoning performance
- Useful for:
  1. Performance comparison (across different tools and formalisms)
  2. Understanding complexity
  3. Evaluation of optimization strategies

Contributions

Optimizations

- Implemented and evaluated for:
  - Tight probabilistic entailment algorithm
- Being investigated for:
  - Probabilistic satisfiability algorithm

Complexity metric

- The experiments validated our intuitions about the connectivity metric

Thanks for your attention!

Some references

- Pronto — P-$\mathcal{SHIQ}$(D) reasoner: http://pellet.owldl.com/pronto
- PREVAL-DL — benchmark suite for P-$\mathcal{SHIQ}$(D) reasoners: http://www2.cs.man.ac.uk/klinovp/projects/prevaldl
- BRCA ontology — prototype of P-$\mathcal{SHIQ}$(D) ontology for the BRCA problem: http://www2.cs.man.ac.uk/klinovp/pronto/brc/cancer_cc.owl

Using OWL but struggling with uncertainty? We may help!

- We welcome any input regarding probabilistic modeling, desired reasoning services, etc.
- Contact us: pklinov@cs.man.ac.uk, bparsia@cs.man.ac.uk