Outline

- Part 1: Motivation
- Part 2: Probabilistic Databases
- Part 3: Weighted Model Counting
- Part 4: Lifted Inference for WFOMC

- Part 5: Completeness of Lifted Inference
- Part 6: Query Compilation
- Part 7: Symmetric Lifted Inference Complexity
- Part 8: Open-World Probabilistic Databases
- Part 9: Discussion & Conclusions

Complexity over Symmetric DBs

Recall: in a symmetric DB all ground facts have the same probability

- We can apply new rules that exploit symmetries
- Dichotomy into PTIME / #P-hard no longer applies
- Lower bounds on query compilation no loner apply

Symmetric WFOMC

No database!

Def. A *weighted vocabulary* is (**R**, **w**), where

- **R** = (R_1, R_2, …, R_k) = relational vocabulary
- **w** = (w_1, w_2, …, w_k) = weights

Fix domain of size n;

- Implicit weights: $w(t) = w_i, \ \forall t \in [n]^{|arity(R_i)|}$

Complexity of symmetric WFOMC(Q,n): fixed Q, input n

Examples

$Q = \forall x \exists y\, R(x,y)$

Computable in PTIME in n

Examples

$Q = \forall x \exists y\, R(x,y)$

$\mathrm{FOMC}(Q,n) = (2^n-1)^n$ $\mathrm{WOMC}(Q,n) = ((1+w_R)^n-1)^n$

Computable in PTIME in n

Examples

Q = ∀x∃y R(x,y)

FOMC(Q,n) = $(2^n-1)^n$ WOMC(Q,n) = $((1+w_R)^n-1)^n$

Q = ∃x∃y [R(x) ∧ S(x,y) ∧T(y)]

$$\text{FOMC}(Q, n) = \sum_{i=0,n} \sum_{j=0,n} \binom{n}{i} \binom{n}{j} 2^{n^2 - ij} (2^{ij} - 1)$$

Computable in PTIME in n

Examples

$Q = \forall x \exists y\ R(x,y)$

$\text{FOMC}(Q,n) = (2^n-1)^n$ $\text{WOMC}(Q,n) = ((1+w_R)^n-1)^n$

$Q = \exists x \exists y\ [R(x) \land S(x,y) \land T(y)]$

$$\text{FOMC}(Q,n) = \sum_{i=0,n} \sum_{j=0,n} \binom{n}{i}\binom{n}{j} 2^{n^2-ij}\ (2^{ij}-1)$$

$$\text{WFOMC}(Q,n) =$$

$$\sum_{i=0,n} \sum_{j=0,n} \binom{n}{i}\binom{n}{j} {w_R}^i {w_T}^j (1+w_S)^{n-ij}\ ((1+w_S)^{ij}-1)$$

Computable in PTIME in n

Hardness is Hard

Triangle = $\exists x \exists y \exists z\, [R(x,y) \wedge S(y,z) \wedge T(z,x)]$

Complexity of FOMC(Triangle, n) = open problem

Hardness is Hard

Triangle = $\exists x \exists y \exists z\ [R(x,y) \wedge S(y,z) \wedge T(z,x)]$

It is hard to prove that Triangle is hard!

- The input = just one number n, runtime = $f(n)$
- In unary: n = 111…11, runtime = f(size of input)
- FOMC(Q, n) in $\#P_1$
- Unlikely #P-hard [Valiant'79]

Complexity of FOMC(Triangle, n) = open problem

The Class $\#P_1$

- $\#P_1$ = functions in #P over a unary input alphabet
  Also called *tally problems*
- Valiant [1979]: *there exists* $\#P_1$ complete problems
- Bertoni, Goldwurm, Sabadini [1991]:
  there exists a CFG s.t. counting # strings of a given length is $\#P_1$ complete
- What about a natural problem?
  - Goldsmith: "no natural combinatorial problems known to be $\#P_1$ complete"

The Logic FO^k

FO^k = FO restricted to k variables

- Note: may reuse variables!
- "The graph has a path of length 10":

$$\exists x \exists y (R(x,y) \wedge \exists x (R(y,x) \wedge \exists y (R(x,y) \wedge \exists x (R(y,x) \ldots)))$$

What is known about FO^k

- Satisfiability is decidable for FO^2
- Satisfiability is undecidable for FO^k, $k \geq 3$

Results for Symmetric Inference

[Van den Broeck’14 ,Beame’15]

Results for Symmetric Inference

Theorem
There exists Q in FO^3 s.t. FOMC(Q, n) is $\#P_1$ hard
There exists CQ Q s.t. WFOMC(Q, n) is $\#P_1$ hard

[Van den Broeck'14 ,Beame'15]

Results for Symmetric Inference

Theorem
There exists Q in FO^3 s.t. FOMC(Q, n) is $\#P_1$ hard
There exists CQ Q s.t. WFOMC(Q, n) is $\#P_1$ hard

Theorem WFOMC(Q, n) is in PTIME
- For any Q in FO^2
- For any gamma-acyclic Q

[Van den Broeck'14 ,Beame'15]

Results for Symmetric Inference

Theorem
There exists Q in FO^3 s.t. FOMC(Q, n) is $\#P_1$ hard
There exists CQ Q s.t. WFOMC(Q, n) is $\#P_1$ hard

Theorem WFOMC(Q, n) is in PTIME

- For any Q in FO^2
- For any gamma-acyclic Q

Corresponding decision problem = the spectrum problem
Data complexity: { Spec(Q) | Q in FO} = NP_1 [Fagin'74]
Combined complexity: NP-complete for FO^2, PSPACE-complete for FO

[Van den Broeck'14 ,Beame'15]

(Non-)Application: 0/1 Laws

Def. $\mu_n(Q)$ = fraction of structures over a domain of size n that are models of Q

$$\mu_n(Q) = FOMC(Q, n) \; / \; FOMC(TRUE, n)$$

Theorem. [Fagin'76]
For all Q in FO (w/o constants) $\lim_{n \to \infty} \mu_n(Q) = 0$ or 1

Example: $Q = \forall x \exists y\, R(x,y)$;

$$FOMC(Q,n) = (2^n-1)^n$$

$$\mu_n(Q) = (2^n-1)^n / 2^{n^2} \to 1$$

(Non-)Application: 0/1 Laws

How does one proof the 0/1 law?

- Attempt: find explicit formula $\mu_n(Q)$, compute limit.
- Fails! because $\mu_n(Q)$ is $\#P_1$-hard in general! Very unlikely to admit a simple closed form formula
- Fagin's proof: beautiful argument involving infinite models, the compactness theorem, and completeness of a theory with a categorical model

Discussion

Fagin 1974

THEOREM 6. *Assume that* $\mathbb{A} \subseteq \mathrm{Fin}(S)$, *and that* $\mathbb{A}$ *is closed under isomorphism.*

1. *If* $S \neq \emptyset$, *then* $\mathbb{A}$ *is an* S-*spectrum iff* $E(\mathbb{A}) \in NP$.
2. *If* $S = \emptyset$, *then* $\mathbb{A}$ *is a spectrum iff* $E(\mathbb{A}) \in NP_1$.

Here: S is a vocabulary, S-spectrum of Q = set of structures that satisfy Q

$\#P_1$ corresponds to {FOMC(Q,n) | Q in FO }

Discussion

Fagin 1974

THEOREM 6. *Assume that* $A \subseteq \mathrm{Fin}(S)$, *and that* A *is closed under isomorphism,*

1. *If* $S \neq \emptyset$, *then* A *is an* S*-spectrum iff* $E(A) \in NP$.
2. *If* $S = \emptyset$, *then* A *is a spectrum iff* $E(A) \in NP_1$.

Here: S is a vocabulary, S-spectrum of Q = set of structures that satisfy Q

Restated:

1. $NP = \exists SO$ — Fagin's classic result
2. $NP_1 = \exists SO(\text{empty-vocabulary})$ — less well known

$\#P_1$ corresponds to {FOMC(Q,n) | Q in FO }

Summary

Exploiting symmetries gives us more power:

- Some queries that are hard over asymmetric databases become easy over symmetric ones: e.g. FO^2 is in PTIME

Limitations:

- Proving hardness is very hard
- Real data is never completely symmetric