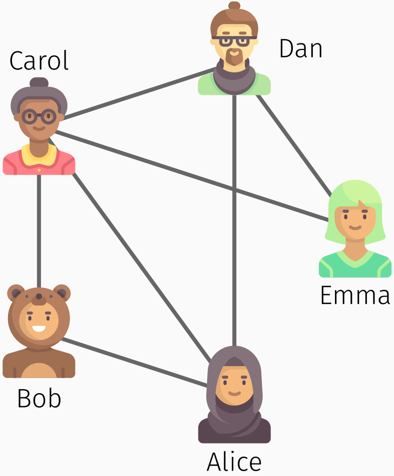
Descriptive Complexity of Learning

PhD Defence

Steffen van Bergerem

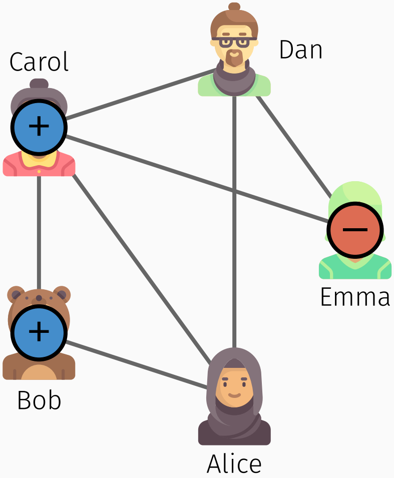
March 10, 2023

Learning from Examples



Icons made by Freepik from Flaticon

Learning from Examples



1

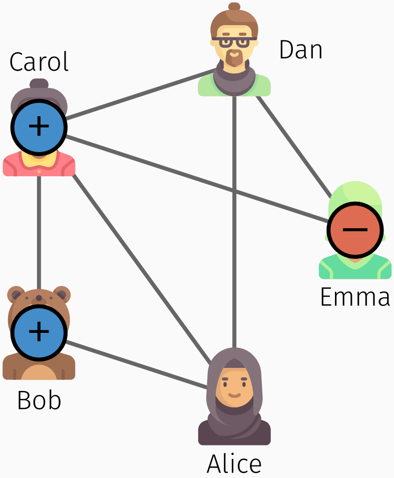
Positive examples

Bob

Carol

Negative examples

Emma



1

Positive examples

Bob

Carol

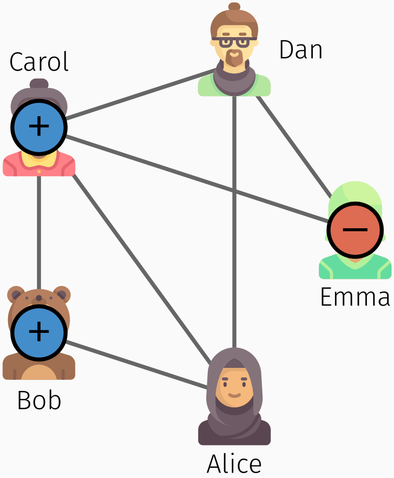
Negative examples

Emma

Possible hypothesis

Alice's friends

Learning from Examples



1

Positive examples

Bob

Carol

Negative examples

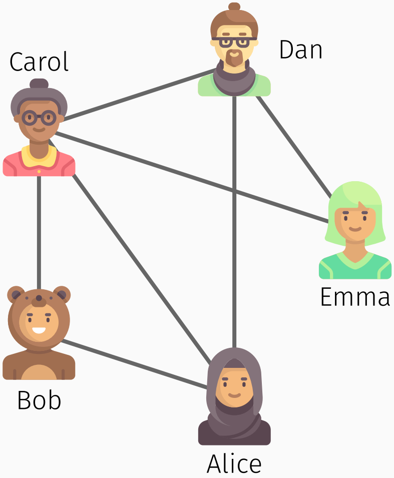
Emma

Possible hypothesis

Alice's friends

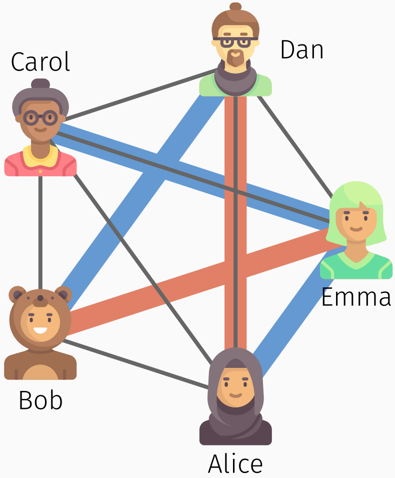
$\varphi(x) = E(x, Alice)$

Learning from Examples



2

Learning from Examples



2

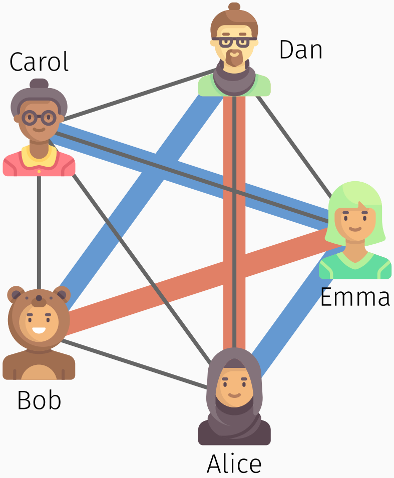
Positive examples

(Alice, Emma)
(Bob, Dan)
(Carol, Emma)

Negative examples

(Alice, Dan)
(Bob, Emma)

Learning from Examples



2

Positive examples

(Alice, Emma)
(Bob, Dan)
(Carol, Emma)

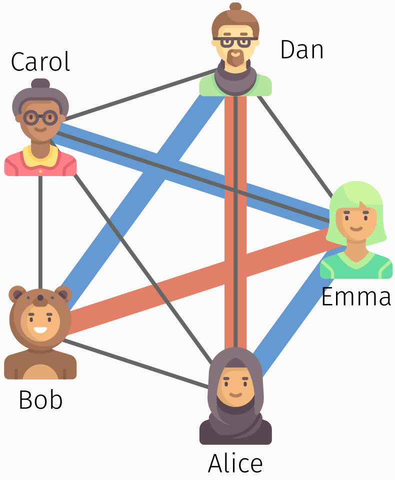
Negative examples

(Alice, Dan)
(Bob, Emma)

Possible hypothesis

having a common friend who is not Carol

Learning from Examples



2

Positive examples

(Alice, Emma)
(Bob, Dan)
(Carol, Emma)

Negative examples

(Alice, Dan)
(Bob, Emma)

Possible hypothesis

having a common friend who is not Carol

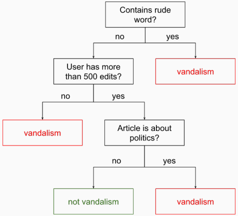
$\varphi(x_1, x_2) = \exists z \ (E(x_1, z) \wedge E(x_2, z) \wedge z \neq \text{Carol})$

Supervised Learning

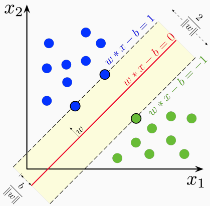
- learn from labelled examples

Supervised Learning

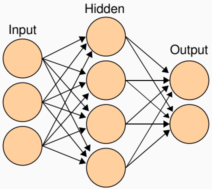
- learn from labelled examples
- algorithms with corresponding hypothesis specifications



decision trees



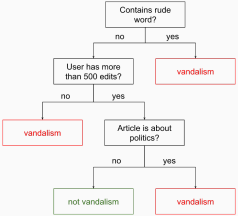
support-vector machines



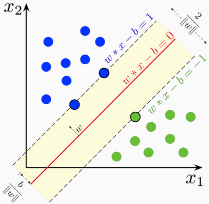
neural networks

Supervised Learning

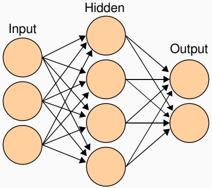
- learn from labelled examples
- algorithms with corresponding hypothesis specifications



decision trees



support-vector machines



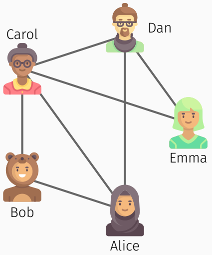
neural networks

- goal of this talk: complexity-theoretic analysis of the problem

→ **Problem: specification of hypotheses**

Logical machine-learning framework

- introduced by Grohe and Turán (2002)
- inputs are labelled tuples from relational structure



Positive examples

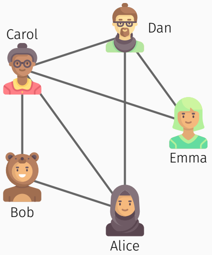
(Alice, Emma)
(Bob, Dan)
(Carol, Emma)

Negative examples

(Alice, Dan)
(Bob, Emma)

Logical machine-learning framework

- introduced by Grohe and Turán (2002)
- inputs are labelled tuples from relational structure



Positive examples

(Alice, Emma)
(Bob, Dan)
(Carol, Emma)

Negative examples

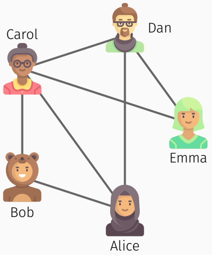
(Alice, Dan)
(Bob, Emma)

- hypotheses are described using logics

$$\varphi(x_1, x_2) \quad = \exists z \; \big(E(x_1, z) \wedge E(x_2, z) \wedge z \neq \text{Carol}\big)$$

Logical machine-learning framework

- introduced by Grohe and Turán (2002)
- inputs are labelled tuples from relational structure



Positive examples

(Alice, Emma)
(Bob, Dan)
(Carol, Emma)

Negative examples

(Alice, Dan)
(Bob, Emma)

- hypotheses are described using logics

$$\varphi(x_1, x_2; y) = \exists z \left(E(x_1, z) \wedge E(x_2, z) \wedge z \neq y \right)$$

parameter: Carol

Algorithmic Meta-Theorems

- **Grohe, Kreutzer, Siebertz (2014)**: properties definable in first-order logic (FO) can be decided in almost linear time on nowhere dense graph classes.

- Algorithmic meta-theorems yield efficient algorithms for certain kinds of logics on certain kinds of structures.

Algorithmic Meta-Theorems

- **Grohe, Kreutzer, Siebertz (2014)**: properties definable in first-order logic (FO) can be decided in almost linear time on nowhere dense graph classes.
- Algorithmic meta-theorems yield efficient algorithms for certain kinds of logics on certain kinds of structures.
- **Goal**: efficient algorithms for learning hypotheses described by certain kinds of logics on certain kinds of structures.

Learning First-Order Logic

Learning Problem

Learning Problem (for fixed $k, \ell, q \in \mathbb{N}$)

Input

- graph G
- labelled examples $T = \big((\bar{v}_1, \lambda_1), \dots, (\bar{v}_m, \lambda_m)\big)$

k-tuples

Output

- FO-formula $\varphi(x_1, \dots, x_k; y_1, \dots, y_\ell)$ with $\mathrm{qr}(\varphi) \leqslant q$
- parameters $w_1, \dots, w_\ell \in V(G)$

with $\quad \mathrm{error}_T(\varphi, \bar{w}) \leqslant \min_{\varphi^*, \bar{w}^*} \mathrm{error}_T(\varphi^*, \bar{w}^*)$

Learning Problem (for fixed $k, \ell, q \in \mathbb{N}$)

Input

- graph G
- labelled examples $T = \big((\bar{v}_1, \lambda_1), \dots, (\bar{v}_m, \lambda_m)\big)$ (k-tuples)
- $\varepsilon > 0$

Output

- FO-formula $\varphi(x_1, \dots, x_k; y_1, \dots, y_\ell)$ with $\mathrm{qr}(\varphi) \leqslant q$
- parameters $w_1, \dots, w_\ell \in V(G)$

with $\quad \mathrm{error}_T(\varphi, \bar{w}) \leqslant \min_{\varphi^*, \bar{w}^*} \mathrm{error}_T(\varphi^*, \bar{w}^*) + \varepsilon$

Learning Problem (for fixed $k, \ell, q \in \mathbb{N}$)

Input

- graph G
- labelled examples $T = \big((\bar{v}_1, \lambda_1), \dots, (\bar{v}_m, \lambda_m)\big)$ (k-tuples)
- $\varepsilon > 0$

Output

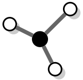
- FO-formula $\varphi(x_1, \dots, x_k; y_1, \dots, y_\ell)$ with $\mathrm{qr}(\varphi) \leqslant q$
- parameters $w_1, \dots, w_\ell \in V(G)$

with $$\mathrm{error}_T(\varphi, \bar{w}) \leqslant \min_{\varphi^*, \bar{w}^*} \mathrm{error}_T(\varphi^*, \bar{w}^*) + \varepsilon$$

Statistical learning theory: Empirical Risk Minimisation Problem

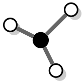
Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time (in the size of the structure).



Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time (in the size of the structure).

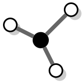


- FO in general not in sublinear time

(v. B., LICS 2019)

Theorem (Grohe and Ritzert, 2017)

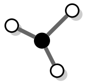
Concepts definable in FO on structures of small degree can be learned in sublinear time (in the size of the structure).



- FO in general not in sublinear time
  (v. B., LICS 2019)
- unary MSO-formulas on strings in sublinear time
  (Grohe, Löding, and Ritzert, 2017)
- unary MSO-formulas on trees in sublinear time
  (Grienenberger and Ritzert, 2019)

Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time (in the size of the structure).



- FO in general not in sublinear time
  (v. B., LICS 2019)
- unary MSO-formulas on strings in sublinear time
  (Grohe, Löding, and Ritzert, 2017)
- unary MSO-formulas on trees in sublinear time
  (Grienenberger and Ritzert, 2019)

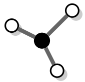
1 more expressive logics

- FO with counting (v. B., LICS 2019)
- FO with weight aggregation
  (v. B. and Schweikardt, CSL 2021)

Learning Concepts Definable in Logics

Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time (in the size of the structure).



- FO in general not in sublinear time
  (v. B., LICS 2019)
- unary MSO-formulas on strings in sublinear time
  (Grohe, Löding, and Ritzert, 2017)
- unary MSO-formulas on trees in sublinear time
  (Grienenberger and Ritzert, 2019)

1 more expressive logics

- FO with counting (v. B., LICS 2019)
- FO with weight aggregation
  (v. B. and Schweikardt, CSL 2021)

2 more complex structures
(v. B., Grohe, and Ritzert, PODS 2022)

- parameterised complexity
- nowhere dense structures

Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time.

Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time.

We only consider formulas of limited quantifier rank.

1: **for all** formulas φ **do**
2: **for all** parameters $\bar{w}$ **do**
3: compute $\mathrm{error}_T(\varphi, \bar{w})$
4: **return** hypothesis $\varphi, \bar{w}$ with smallest number of errors

Learning First-Order Logic

Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time.

We only consider formulas of limited quantifier rank.

constantly many

for all normal forms of formulas φ **do**
  for all parameters $\bar{w}$ **do**
    compute $\mathrm{error}_T(\varphi, \bar{w})$
return hypothesis $\varphi, \bar{w}$ with smallest number of errors

1) Gaifman normal form

Learning First-Order Logic

Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time.

We only consider formulas of limited quantifier rank.

constantly many

small degree,
small neighbourhood
(polylogarithmic)

1: **for all** normal forms of formulas φ **do**
2: **for all** parameters $\bar{w}$ in a certain neighbourhood **do**
3: compute $\mathrm{error}_T(\varphi, \bar{w})$
4: **return** hypothesis $\varphi, \bar{w}$ with smallest number of errors

1) Gaifman normal form
2) Gaifman locality and Feferman-Vaught decompositions

Learning FO with Counting

Concepts definable in FOCN
on structures of degree at most $(\log \log n)^{c}$
can be learned in sublinear time.

Introduced by Kuske and Schweikardt

Terms

- $t_{\text{degree}}(x) = \#(y).\big(E(x,y)\big)$
- $t_{3-\text{walks}}(x,y) = \#(z_1,z_2).\big(E(x,z_1) \wedge E(z_1,z_2) \wedge E(z_2,y)\big)$

Formulas

First-Order Logic with Counting (FOCN)

Introduced by Kuske and Schweikardt

Terms

- $t_{\text{degree}}(x) = \#(y).\big(E(x,y)\big)$
- $t_{3-\text{walks}}(x,y) = \#(z_1,z_2).\big(E(x,z_1) \wedge E(z_1,z_2) \wedge E(z_2,y)\big)$

Formulas

- $\varphi_{\text{regular}} = \exists\kappa\,\forall x\,\big(t_{\text{degree}}(x) = \kappa\big)$
- $\psi(x,y) = \Big(t_{3-\text{walks}}(x,y) = t_{\text{degree}}(x) \cdot t_{\text{degree}}(y)\Big)$

Theorem (v. B., LICS 2019)

Concepts definable in FOCN on structures of bounded degree can be learned in sublinear time.

Theorem (v. B., LICS 2019)

Concepts definable in FOCN on structures of bounded degree can be learned in sublinear time.

No Gaifman normal form for FOCN, only Hanf normal form.

Theorem (v. B., LICS 2019)

Concepts definable in FOCN on structures of bounded degree can be learned in sublinear time.

No Gaifman normal form for FOCN, only Hanf normal form.

Theorem (v. B., 2022)

Concepts definable in FOCN on structures of degree at most $(\log \log n)^c$ can be learned in sublinear time.

Theorem (v. B., LICS 2019)

Concepts definable in FOCN on structures of bounded degree can be learned in sublinear time.

No Gaifman normal form for FOCN, only Hanf normal form.

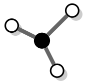
Theorem (v. B., 2022)

Concepts definable in FOCN on structures of degree at most $(\log\log n)^c$ can be learned in sublinear time.

- number of formulas not constant any more
- uses efficient isomorphism test (Grohe, Neuen, Schweitzer 2018)

Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time (in the size of the structure).



- FO in general not in sublinear time
  (v. B., LICS 2019)
- unary MSO-formulas on strings in sublinear time
  (Grohe, Löding, and Ritzert, 2017)
- unary MSO-formulas on trees in sublinear time
  (Grienenberger and Ritzert, 2019)

1 more expressive logics

- FO with counting (v. B., LICS 2019)
- FO with weight aggregation
  (v. B. and Schweikardt, CSL 2021)

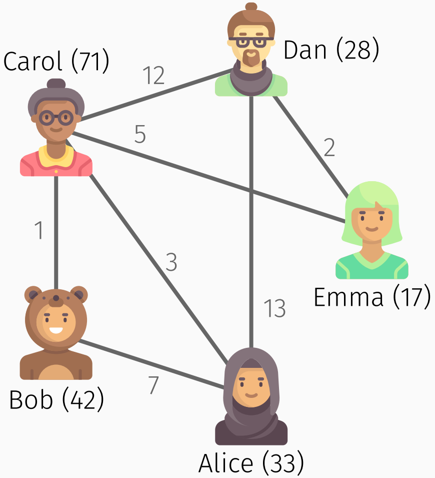
2 more complex structures
(v. B., Grohe, and Ritzert, PODS 2022)

- parameterised complexity
- nowhere dense structures

Learning FO with Weight Aggregation

Concepts definable in FOWA_1
on weighted structures of small degree
can be learned in sublinear time.

First-Order Logic with Weight Aggregation (FOWA)



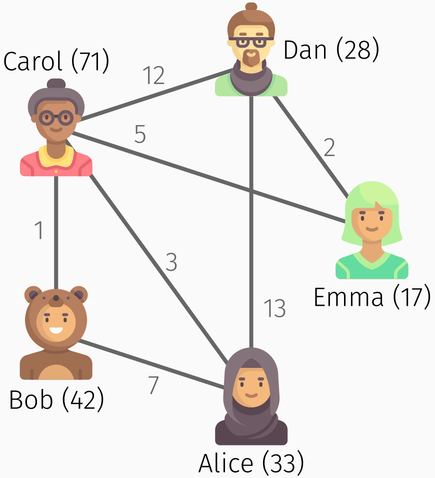
Introduced in (v. B. and Schweikardt, CSL 2021)

Terms

- $t(x) = \sum_y a(y) \cdot \ell(x, y).E(x, y)$

Formulas

First-Order Logic with Weight Aggregation (FOWA)



Introduced in (v. B. and Schweikardt, CSL 2021)

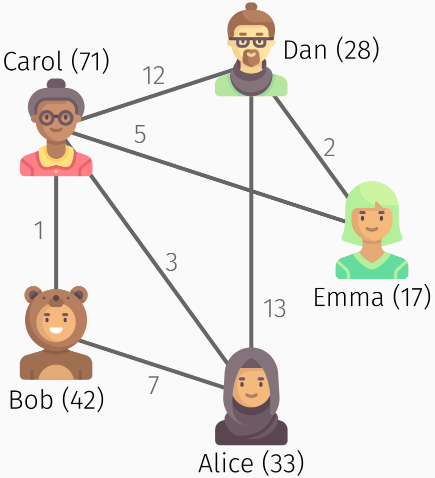
Terms

- $t(x) = \sum_y a(y) \cdot \ell(x, y).E(x, y)$

Formulas

- $\varphi_1(x) = \left(\sum_y a(y).E(x, y) \leqslant \sum_y a(y).\neg E(x, y)\right)$

First-Order Logic with Weight Aggregation (FOWA)



Introduced in (v. B. and Schweikardt, CSL 2021)

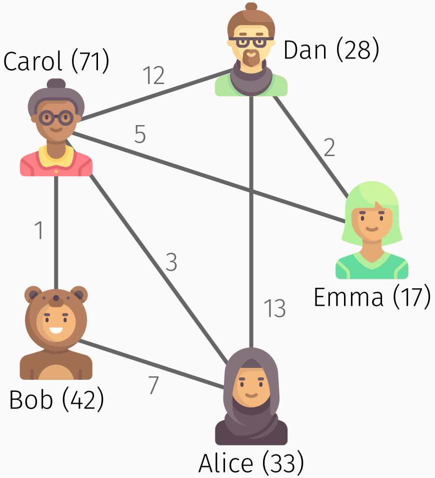
Terms

- $t(x) = \sum_y a(y) \cdot \ell(x, y).E(x, y)$

Formulas

- $\varphi_1(x) = \left(\sum_y a(y).E(x, y) \leqslant \sum_y a(y).\neg E(x, y)\right)$
- $\varphi_2(x) = \exists y \left(\#(z).(E(x, z) \wedge E(y, z)) \geqslant 2\right)$

First-Order Logic with Weight Aggregation (FOWA)



Introduced in (v. B. and Schweikardt, CSL 2021)

Terms

- $t(x) = \sum_y a(y) \cdot \ell(x, y).E(x, y)$

Formulas

- $\varphi_1(x) = \left(\sum_y a(y).E(x, y) \leqslant \sum_y a(y).\neg E(x, y)\right)$
- $\varphi_2(x) = \exists y \left(\#(z).(E(x, z) \wedge E(y, z)) \;\geqslant\; 2\right)$

FOWA$_1$

- subformulas comparing terms may only have at most one free variable

Theorem (v. B. and Schweikardt, CSL 2021)

Concepts definable in FOWA_1 on weighted structures of small degree can be learned in sublinear time with quasilinear-time precomputation.

Theorem (v. B. and Schweikardt, CSL 2021)

Concepts definable in $FOWA_1$ on weighted structures of small degree can be learned in sublinear time with quasilinear-time precomputation.

Theorem (v. B. and Schweikardt, CSL 2021)

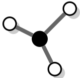
Concepts definable in FOWA_1 on weighted structures of small degree can be learned in sublinear time with quasilinear-time precomputation.

1. **Localisation Theorem for FOWA_1**
   enrich structure, then $\mathrm{FOWA}_1 \to \mathrm{FOW}_1$
2. **Gaifman normal form for FOW_1**
3. **Feferman-Vaught decompositions for FOW_1**
4. **techniques similar to FO**

Learning Concepts Definable in Logics

Theorem (Grohe and Ritzert, 2017)

Concepts definable in FO on structures of small degree can be learned in sublinear time (in the size of the structure).



- FO in general not in sublinear time
  (v. B., LICS 2019)
- unary MSO-formulas on strings in sublinear time
  (Grohe, Löding, and Ritzert, 2017)
- unary MSO-formulas on trees in sublinear time
  (Grienenberger and Ritzert, 2019)

1. more expressive logics
   - FO with counting (v. B., LICS 2019)
   - FO with weight aggregation (v. B. and Schweikardt, CSL 2021)
2. more complex structures (v. B., Grohe, and Ritzert, PODS 2022)
   - parameterised complexity
   - nowhere dense structures

Parameterised Complexity of Learning

In general, learning FO is hard.

But on nowhere dense classes,
learning FO is fixed-parameter tractable.

Input

- graph G
- $k \in \mathbb{N}$

Parameter

k

Output

“Yes” if and only if G has a k-clique

Parameterised Clique Problem

Input

- graph G
- $k \in \mathbb{N}$

Parameter

k

Output

"Yes" if and only if G has a k-clique

Parameterised Complexity

- XP: running time $n^{f(k)}$
- fixed-parameter tractable (FPT): running time $f(k) \cdot p(n)$

Learning Problem (for fixed $k, \ell, q \in \mathbb{N}$)

Input

- graph G
- labelled examples $T = \big((\bar{v}_1, \lambda_1), \dots, (\bar{v}_m, \lambda_m)\big)$
- $\varepsilon > 0$

Output

- FO-formula $\varphi(x_1, \dots, x_k; y_1, \dots, y_\ell)$ with $\mathrm{qr}(\varphi) \leqslant q$
- parameters $w_1, \dots, w_\ell \in V(G)$

with $$\mathrm{error}_T(\varphi, \bar{w}) \leqslant \min_{\varphi^*, \bar{w}^*} \mathrm{error}_T(\varphi^*, \bar{w}^*) + \varepsilon$$

Input

- graph G
- labelled examples $T = \big((\bar{v}_1, \lambda_1), \ldots, (\bar{v}_m, \lambda_m)\big)$
- $\varepsilon > 0, \quad k, \ell, q \in \mathbb{N}$

Output

- FO-formula $\varphi(x_1, \ldots, x_k; y_1, \ldots, y_\ell)$ with $\mathrm{qr}(\varphi) \leqslant q$
- parameters $w_1, \ldots, w_\ell \in V(G)$

with $\quad \mathrm{error}_T(\varphi, \bar{w}) \leqslant \min_{\varphi^*, \bar{w}^*} \mathrm{error}_T(\varphi^*, \bar{w}^*) + \varepsilon$

Input

- graph G
- labelled examples $T = \big((\bar{v}_1, \lambda_1), \dots, (\bar{v}_m, \lambda_m)\big)$
- $\varepsilon > 0, \quad k, \ell, q \in \mathbb{N}$

Parameter

$k + \ell + q + 1/\varepsilon$

Output

- FO-formula $\varphi(x_1, \dots, x_k; y_1, \dots, y_\ell)$ with $\mathrm{qr}(\varphi) \leqslant q$
- parameters $w_1, \dots, w_\ell \in V(G)$

with $\quad \mathrm{error}_T(\varphi, \bar{w}) \leqslant \min_{\varphi^*, \bar{w}^*} \mathrm{error}_T(\varphi^*, \bar{w}^*) + \varepsilon$

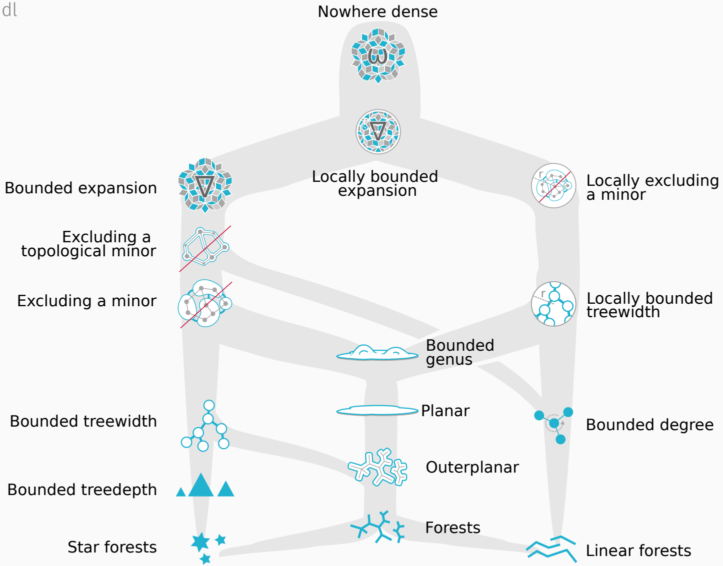
Theorem (v. B., Grohe, and Ritzert, PODS 2022)

Learning concepts definable in FO is at least as hard as the model-checking problem for FO.

Under a common complexity-theoretic assumption (p-Clique $\notin$ FPT), model checking is not fixed-parameter tractable.

Classes with a Tractable Model-Checking Problem

Illustration by Felix Reidl



Theorem (v. B., Grohe, and Ritzert, PODS 2022)

For every nowhere dense class $\mathcal{C}$ of graphs, learning concepts definable in FO is fixed-parameter tractable on $\mathcal{C}$.

Proof uses game characterisation of nowhere dense classes.

Theorem (v. B., Grohe, and Ritzert, PODS 2022)

For every nowhere dense class $\mathcal{C}$ of graphs, learning concepts definable in FO is fixed-parameter tractable on $\mathcal{C}$.

Proof uses game characterisation of nowhere dense classes.

Theorem (v. B., 2022)

For every nowhere dense class $\mathcal{C}$ of structures, learning concepts definable in FO is fixed-parameter tractable on $\mathcal{C}$.

Conclusion

Theorem (v. B., LICS 2019; v. B., 2022)

Concepts definable in FOCN on structures of degree at most $(\log\log n)^c$ can be learned in sublinear time.

Theorem (v. B. and Schweikardt, CSL 2021)

Concepts definable in $FOWA_1$ on weighted structures of small degree can be learned in sublinear time with quasilinear-time precomputation.

Theorem (v. B., Grohe, and Ritzert, PODS 2022; v. B., 2022)

For every nowhere dense class $\mathcal{C}$ of structures, learning concepts definable in FO is fixed-parameter tractable on $\mathcal{C}$.

Discussion

Theorem (v. B., Grohe, and Ritzert, PODS 2022; v. B., 2022)

For every nowhere dense class $\mathcal{C}$, learning concepts definable in FO is fixed-parameter tractable on $\mathcal{C}$.

FPT learnability results for monadic second-order logic (MSO) on classes of bounded treewidth and classes of bounded cliquewidth (Runde, 2022)

Discussion

Theorem (v. B. and Schweikardt, CSL 2021)

Concepts definable in FOWA_1 on weighted structures of small degree can be learned in sublinear time with quasilinear-time precomputation.

Theorem (v. B., Grohe, and Ritzert, PODS 2022; v. B., 2022)

For every nowhere dense class $\mathcal{C}$ of structures, learning concepts definable in FO is fixed-parameter tractable on $\mathcal{C}$.

Future Research

- FPT learnability for FOWA_1?
- non-Boolean classification problems?
- FPT learnability on dense structures?