# Description Complexity of Unary Structures in First-Order Logic with Links to Entropy

**Reijo Jaakkola**, Antti Kuusisto, Miikka Vilander

Mathematics Research Centre, Tampere University, Finland

February 13, 2025

# Description complexity (in FO)

- Given a $\tau$-structure $\mathfrak{M}$ of size $n$, its description complexity $C(\mathfrak{M})$ is the size of the smallest sentence $\varphi \in \mathrm{FO}[\tau]$ with the following property: for every $\tau$-structure $\mathfrak{N}$ of size $n$ we have that

$$\mathfrak{N} \models \varphi \Leftrightarrow \mathfrak{N} \cong \mathfrak{M}.$$

- **Example:** The clique $G$ of size $n$ is described by the sentence

$$\forall x \forall y (x = y \vee E(x, y)).$$

This sentence has size

$$\underbrace{2}_{\text{quantifiers}} + \underbrace{1}_{\text{disjunction}} + \underbrace{2}_{\text{atomic formulas}} = 5$$

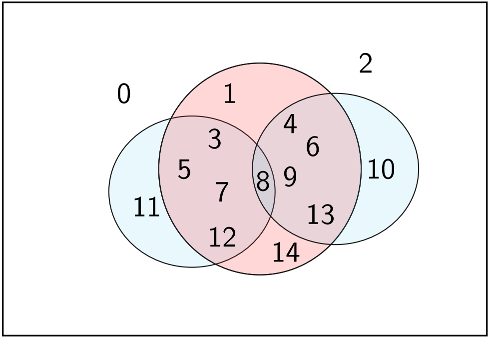
so $C(G) \leq 5$ (and in fact $C(G) = 5$).

# Unary structures

- Unary structure (of size $n$) over $\tau = \{P_1, \ldots, P_k\}$ is a tuple

$$\mathfrak{M} := ([n], P_1^{\mathfrak{M}}, \ldots, P_k^{\mathfrak{M}}),$$

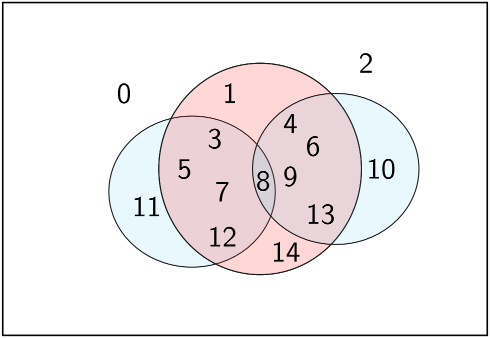
where $[n] := \{0, \ldots, n-1\}$ and $P_1^{\mathfrak{M}}, \ldots, P_k^{\mathfrak{M}} \subseteq [n]$.



# Unary structures

- **Terminology:** a type is $\pi \subseteq \{P_1, \ldots, P_k\}$. A type $\pi$ is realized in $\mathfrak{M}$ if there is $i \in [n]$ such that $i \in P^{\mathfrak{M}}$ iff $P \in \pi$. We write

$$|\pi| := |\{i \in [n] \mid i \text{ realizes } \pi\}|$$



- What is the description complexity of a unary structure?
- The naive formula

$$\bigwedge_{\ell=1}^{2^{|\tau|}} \exists x_1 \ldots \exists x_{|\pi_\ell|} \Big( \bigwedge_{i=1}^{|\pi_\ell|} \pi_\ell(x_i) \land \bigwedge_{j=i+1}^{|\pi_\ell|} x_i \neq x_j \Big),$$

has size $\mathcal{O}(n^2)$.

# Upper bound

## Theorem

*Let $\mathfrak{M}$ be a unary structure. Let $T = \{\pi_1, \ldots, \pi_\ell\}$ be the types realized in $\mathfrak{M}$, enumerated in ascending order of numbers of realizing points. We have*

$$C(\mathfrak{M}) \leq \min(3|\pi_\ell|, 6|\pi_{\ell-1}|) + \mathcal{O}(1).$$

# Lower bound

## Theorem

*Let $\mathfrak{M}$ be a unary structure. Let $T = \{\pi_1, \ldots, \pi_\ell\}$ be the types realized in $\mathfrak{M}$, enumerated in ascending order of numbers of realizing points. Now*

$$C(\mathfrak{M}) \geq 3|\pi_{\ell-1}| - 3.$$

## Proof idea.

Proof via a formula size game for $\mathrm{FO}$-sentences in prenex normal form.

1. Show that the prefix needs to have $|\pi_{\ell-1}|$ quantifiers.
2. Show that the quantifier-free part needs to contain $|\pi_{\ell-1}| - 1$ atomic formulas.

$\square$

# Bounded quantifier depth

- We use $\mathrm{FO}_d$ to denote the set of sentences of $\mathrm{FO}$ which have quantifier rank $\leq d$.
- Given a unary structure $\mathfrak{M}$ of size $n$ we define

$$[\mathfrak{M}]_d := \{\mathfrak{N} \mid \mathfrak{N} \text{ has size } n \text{ and } \mathfrak{M} \equiv_d \mathfrak{N}\}.$$

- $C_d(\mathfrak{M})$ is the size of the shortest formula in $\mathrm{FO}_d$ which defines $[\mathfrak{M}]_d$.
- Form of lossy compression.

# Bounded quantifier depth

**Theorem**

*Let $\mathfrak{M}$ be a unary structure. Let $T = \{\pi_1, \ldots, \pi_\ell\}$ be the types realized in $\mathfrak{M}$, enumerated in ascending order of numbers of realizing points. Let $r$ be the largest index for which $|\pi_r| < d$.*

1. $C_d(\mathfrak{M}) \leq 3d + 3|\pi_r| + \mathcal{O}(1)$.
2. *If* $|\pi_{\ell-1}| < d$, *then* $C_d(\mathfrak{M}) \leq 6|\pi_{\ell-1}| + \mathcal{O}(1)$.

# Expected description complexity

- Let

$$\mathbb{E}_n[C] := \frac{1}{2^{|\tau|n}} \sum_{\mathfrak{M}} C(\mathfrak{M})$$

be the expected description complexity of a random $\tau$-model of size $n$.

**Theorem**

$$\mathbb{E}_n[C] \sim \frac{3n}{2^{|\tau|}}, \ \textit{as } n \to \infty$$

**Proof.**

With high probability, the types in a random unary structure of size $n$ are realized roughly the same number of times. For such structures our formula size bounds match up to a sublinear additive term. □

# Entropy

- Intuitively $C(\mathfrak{M})$ measures the "randomness" of $\mathfrak{M}$.
- Another way to measure the randomness of $\mathfrak{M}$ is its entropy.
- For $\mathfrak{M}$ its Boltzmann entropy is

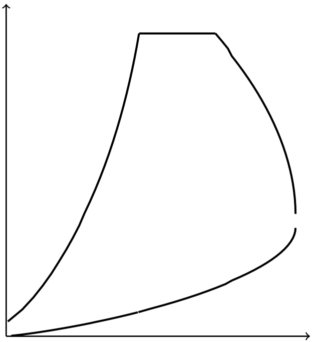
$$H_B(\mathfrak{M}) := \log(|\{\mathfrak{N} \mid \mathfrak{M} \cong \mathfrak{N}\}|),$$

  while its Shannon entropy $H_S(\mathfrak{M})$ looks at the type distribution.
- For large $n$ we have

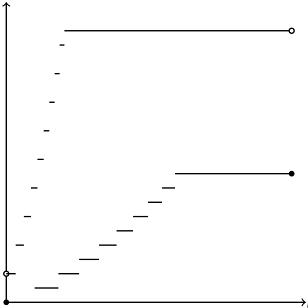
$$\frac{1}{n} H_B(\mathfrak{M}) \sim H_S(\mathfrak{M}),$$

  where $\mathfrak{M}$, $\mathfrak{N}$ have size $n$.

# Links to entropy




(a)




(b)

Figure: Figure 1a on the left shows an area that encapsulates all combinations of Shannon entropy and $\mathrm{FO}$-description complexity for the values $|\tau| = 2$ and $n = 1000$. Figure 1b concerns the case of $\mathrm{FO}_d$ and shows bounds on description complexity in terms of Boltzmann entropy for values $|\tau| = 2$, $n = 100$ and $d = 10$.

# Conclusions

- Almost sharp bounds on the description complexity of unary structures.
- Some connections between description complexity and entropy.
  - Surprisingly, in the case of full $\mathrm{FO}$ the hardest structures to describe do not have maximal entropy.
  - In the case of $\mathrm{FO}_d$ there seems to be a monotone connection. This has been established in a simpler context in [Jaakkola et. al., 2023].
- **Some future directions:**
  1. Sharper bounds on the description complexity of unary structures.
  2. Description complexity of graphs.

# Thanks!