Fundamentos lógicos de bases de datos
(Logical foundations of databases)

Diego Figueira
Gabriele Puppis

CNRS  LaBRI
A different perspective: a coarser view on expressiveness...
A different perspective: a coarser view on expressiveness...

How do FO properties distribute among ALL structures?

Or equally, what percentage of graphs verify a given FO sentence?
\[ \mu_n(P) = \text{“the probability that a graph with } n \text{ nodes satisfies property } P” \]
$\mu_n(P) = \text{“the probability that a graph with } n \text{ nodes satisfies property } P\text{”}$

Uniform distribution

( each pair of nodes has an edge with probability $\frac{1}{2}$ )
\( \mu_n(P) = \text{“the probability that a graph with } n \text{ nodes satisfies property } P \text{”} \)

\[ C_n = \{ \text{graphs with } n \text{ nodes} \} \]

\[ \mu_n(P) = \frac{| \{ G \in C_n \mid G \models P \} |}{| C_n |} = \frac{2^{n^2}}{2^{n^2}} \]
\( \mu_n(P) = \) “the probability that a graph with \( n \) nodes satisfies property \( P \)”

\[ C_n = \{ \text{graphs with } n \text{ nodes} \} \]

\[ \mu_n(P) = \frac{|\{G \in C_n | G \models P\}|}{|C_n|} = \frac{2^{n^2}}{2^{3^2}} \]

E.g. for \( P = \) “the graph is complete”

\[ \mu_3(P) = \frac{1}{|C_3|} = \frac{1}{2^{3^2}} \]
0-1 Law

\[ \mu_n(P) = \text{“the probability that a graph with } n \text{ nodes satisfies property } P \text{”} \]

\[ C_n = \{ \text{graphs with } n \text{ nodes} \} \]

\[ \mu_n(P) = \frac{|\{G \in C_n \mid G \models P\}|}{|C_n|} = \frac{2^{n^2}}{2^n} \]

\[ \mu_\infty(P) = \lim_{n \to \infty} \mu_n(P) \]

Uniform distribution
( each pair of nodes has an edge with probability \( \frac{1}{2} \))

E.g. for \( P = \text{“the graph is complete”} \)

\[ \mu_3(P) = \frac{1}{|C_3|} = \frac{1}{2^{3^2}} \]
Theorem. [Glebskii et al. ’69, Fagin ’76]

For every *FO sentence* $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.
Theorem. [Glebskii et al. ’69, Fagin ’76]

For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Examples:

- $\phi = “there is a triangle”$
  \[ \mu_3(\phi) = \frac{1}{|C_3|} \quad \mu_{3n}(\phi) \geq 1 - \left(1 - \frac{1}{|C_3|}\right)^n \to 1 \]
0-1 Law

Theorem. [Glebskii et al. ’69, Fagin ’76]

For every FO sentence \( \phi \), \( \mu_\infty(\phi) \) is either 0 or 1.

Examples:

- \( \phi = “\text{there is a triangle}” \)
  
  \[ \mu_3(\phi) = \frac{1}{|C_3|} \quad \mu_{3n}(\phi) \geq 1 - (1 - \frac{1}{|C_3|})^n \to 1 \]

- \( \phi = “\text{there no 5-clique}” \)

  \[ \mu_\infty(\phi) = 0 \]
0-1 Law

Theorem. [Glebskii et al. ’69, Fagin ’76]

For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Examples:

- $\phi =$ “there is a triangle”
  \[ \mu_3(\phi) = \frac{1}{|C_3|} \quad \mu_{3n}(\phi) \geq 1 - (1 - \frac{1}{|C_3|})^n \rightarrow 1 \]

- $\phi =$ “there no 5-clique”
  \[ \mu_\infty(\phi) = 0 \]

- $\phi_H =$ “there is an occurrence of $H$ as induced sub-graph”
  \[ \mu_\infty(\phi_H) = 1 \]
0-1 Law

Theorem. [Glebskii et al. ’69, Fagin ’76]
For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Examples:

- $\phi =$ “there is a triangle”
  
  $\mu_3(\phi) = \frac{1}{|C_3|}$
  
  $\mu_3n(\phi) \geq 1 - (1 - \frac{1}{|C_3|})^n \to 1$

- $\phi =$ “there no 5-clique”

- $\phi_H =$ “there is an occurrence of $H$ as induced sub-graph”
  
  $\mu_\infty(\phi_H) = 1$

- $\phi =$ “even number of edges”

- $\phi =$ “even number of nodes”

Your turn!
Theorem. \[\text{[Glebskii et al. ’69, Fagin ’76]}\]

For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Examples:

- $\phi =$ “there is a triangle”
  \[\mu_3(\phi) = \frac{1}{|C_3|} \quad \mu_{3n}(\phi) \geq 1 - (1 - \frac{1}{|C_3|})^n \to 1\]
- $\phi =$ “there no 5-clique”
  \[\mu_\infty(\phi) = 0\]
- $\phi_H =$ “there is an occurrence of H as induced sub-graph”
  \[\mu_\infty(\phi_H) = 1\]
- $\phi =$ “even number of edges”
  \[\mu_\infty(\phi) = \frac{1}{2}\]
- $\phi =$ “even number of nodes”
  $\mu_\infty(\phi)$ not even defined

Your turn!
Theorem. [Glebskii et al. ’69, Fagin ’76]

For every $FO$ sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Examples:

• $\phi =$ “there is a triangle”
  \[ \mu_3(\phi) = \frac{1}{|C_3|}, \quad \mu_{3n}(\phi) \geq 1 - (1 - \frac{1}{|C_3|})^n \rightarrow 1 \]

• $\phi =$ “there no 5-clique”
  \[ \mu_\infty(\phi) = 0 \]

• $\phi_H =$ “there is an occurrence of $H$ as induced sub-graph”
  \[ \mu_\infty(\phi_H) = 1 \]

• $\phi =$ “even number of edges”

• $\phi =$ “even number of nodes”

• $\phi =$ “more edges than nodes”
  \[ \mu_\infty(\phi) = \frac{1}{2} \]

Your turn!

• $\phi =$ “not even defined”
  \[ \mu_\infty(\phi) \text{ not even defined} \]

• $\phi =$ “yet not FO-definable!”
  \[ \mu_\infty(\phi) = 1 \]
0-1 Law

For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Let $k =$ quantifier rank of $\phi$

$$\delta_k = \forall x_1, \ldots, x_k \forall y_1, \ldots, y_k \exists z \land_{i,j} x_i \neq y_j \land E(x_i, z) \land \neg E(y_j, z)$$

(Extension Axiom)
0-1 Law

For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Let $k =$ quantifier rank of $\phi$

$$\delta_k = \forall x_1, \ldots, x_k \forall y_1, \ldots, y_k \exists z \wedge_{i,j} x_i \neq y_j \wedge E(x_i, z) \wedge \neg E(y_j, z)$$

(Extension Axiom)

Fact 1: If $G \models \delta_k \wedge H \models \delta_k$ then

Duplicator survives $k$ rounds on $G, H$
For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Let $k = \text{quantifier rank of } \phi$

\[
\delta_k = \forall x_1, \ldots, x_k \forall y_1, \ldots, y_k \exists z \ (i,j) \ x_i \neq y_j \land E(x_i, z) \land \neg E(y_j, z)
\]

(Extension Axiom)

Fact 1: If $G \models \delta_k \land H \models \delta_k$ then Duplicator survives $k$ rounds on $G, H$

Fact 2: $\mu_\infty(\delta_k) = 1$

($\delta_k$ is almost surely true)
For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Let $k = \text{quantifier rank of } \phi$

$$\delta_k = \forall x_1, ..., x_k \forall y_1, ..., y_k \exists z \ \land_{i,j} x_i \neq y_j \land E(x_i, z) \land \neg E(y_j, z)$$

(Extension Axiom)

Fact 1: If $G \models \delta_k \land H \models \delta_k$ then
Duplicator survives $k$ rounds on $G, H$

Fact 2: $\mu_\infty(\delta_k) = 1$
( $\delta_k$ is almost surely true)

a) There is $G \models \delta_k \land \phi$ $\Rightarrow$ (by Fact 1) $\forall H$: If $H \models \delta_k$ then $H \models \phi$

Thus, $\mu_\infty(\delta_k) \leq \mu_\infty(\phi)$

$\Rightarrow$ (by Fact 2) $\mu_\infty(\delta_k) = 1$, hence $\mu_\infty(\phi) = 1$
0-1 Law

For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1.

Let $k = \text{quantifier rank of } \phi$

$$\delta_k = \forall x_1, ..., x_k \forall y_1, ..., y_k \exists z \ (i,j) x_i \neq y_j \land E(x_i, z) \land \neg E(y_j, z)$$

(Extension Axiom)

Fact 1: If $G \models \delta_k \land H \models \delta_k$ then
Duplicator survives $k$ rounds on $G, H$

Fact 2: $\mu_\infty(\delta_k) = 1$
\hspace{1cm} ($\delta_k$ is almost surely true)

2 cases

\begin{align*}
a) \quad & \text{There is } G \models \delta_k \land \phi \quad \Rightarrow \text{ (by Fact 1) } \forall H: \text{ If } H \models \delta_k \text{ then } H \models \phi \\
& \Rightarrow \text{ (by Fact 2) } \mu_\infty(\delta_k) = 1, \text{ hence } \mu_\infty(\phi) = 1
\end{align*}

b) \quad & \text{There is no } G \models \delta_k \land \phi \quad \Rightarrow \text{ (by Fact 2) there is } G \models \delta_k, \\
& \Rightarrow G \models \delta_k \land \neg \phi \quad \Rightarrow \text{ (by case a) } \mu_\infty(\neg \phi) = 1
For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1, and this depends on whether $\text{RADO} \models \phi$.

RADO =
For every $FO$ sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1, and this depends on whether $\text{RADO} \models \phi$. 

Each pair of nodes $i, j$ is connected with probability $1/2$. 

$\text{RADO} =$
0-1 Law

For every $FO$ sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1, and this depends on whether $RADO \models \phi$

RADO =

each pair of nodes $i, j$ is connected if $i$-th bit of $j$ is 1

each pair of nodes $i, j$ is connected with probability $1/2$
For every FO sentence $\phi$, $\mu_\infty(\phi)$ is either 0 or 1, and this depends on whether $\text{RADO} \models \phi$.

\text{RADO} = \text{the unique graph that satisfies } \delta_k \text{ for all } k.

Each pair of nodes $i, j$ is connected with probability $1/2$. Each pair of nodes $i, j$ is connected if the $i$-th bit of $j$ is 1.
Theorem. The problem of deciding whether an FO sentence is \textit{almost surely true} ($\mu_\infty = 1$) is PSPACE-complete. [Grandjean ’83]
Theorem. The problem of deciding whether an FO sentence is *almost surely true* \( (\mu_\infty = 1) \) is PSPACE-complete. [Grandjean ’83]
**Theorem.** The problem of deciding whether an FO sentence is *almost surely true* ($\mu_\infty = 1$) is PSPACE-complete. [Grandjean ’83]
Theorem. The problem of deciding whether an FO sentence is *almost surely true* ($\mu_\infty = 1$) is PSPACE-complete.

[Grandjean ’83]

Query evaluation on large databases:
Don’t bother evaluating an FO query, it’s either true or false with high probability!
Does the 0-1 Law apply to real-life databases?

Not quite: database constraints easily spoil Extension Axiom.
0-1 Law

Does the 0-1 Law apply to real-life databases?

Not quite: database constraints easily spoil Extension Axiom.

Consider:

- functional constraint $\forall x, x', y, y' \left( (E(x,y) \land E(x',y') \Rightarrow y = y') \land (E(x,y) \land E(x',y) \Rightarrow x = x') \right)$ (E is a permutation)

- FO query $\phi = \neg \exists x \ E(x,x)$
Does the 0-1 Law apply to real-life databases?
Not quite: database constraints easily spoil Extension Axiom.

Consider:

- functional constraint \( \forall x, x', y, y' \left( E(x,y) \land E(x',y) \Rightarrow y = y' \right) \land \left( E(x,y) \land E(x',y') \Rightarrow x = x' \right) \) (E is a permutation)

- FO query \( \phi = \neg \exists x \ E(x,x) \)

Probability that a permutation E satisfies \( \phi = \frac{!n}{n!} \to e^{-1} = 0.3679... \)
0-1 Law

Does the 0-1 Law apply to real-life databases?

Not quite: database *constraints* easily spoil Extension Axiom.

Consider:

- functional constraint $\forall x, x', y, y' \left( (E(x,y) \land E(x,y') \Rightarrow y = y') \land \right.$
  
  $\left. (E(x,y) \land E(x',y) \Rightarrow x = x') \right)$ (E is a permutation)

- FO query $\phi = \neg \exists x \ E(x,x)$

Probability that a permutation E satisfies $\phi = \frac{!n}{n!} \rightarrow e^{-1} = 0.3679...$

The 0-1 Law is a tool for proving expressiveness results, not a statement on the real-life probability of queries being non-empty.
Idea: First order logic can only express “local” properties
Idea: First order logic can only express “local” properties

Local = properties of nodes which are close to one another
Definition. The **Gaifman graph** of a structure $S = (V, R_1, \ldots, R_m)$ is the **undirected** graph $G(S) = (V, E)$ where $E = \{ (u, v) \mid \exists (\ldots, u, \ldots, v, \ldots) \in R_i \text{ for some } i \}$.
Definition. The **Gaifman graph** of a structure $S = (V, R_1, \ldots, R_m)$ is the **undirected** graph $G(S) = (V, E)$ where $E = \{ (u, v) \mid \exists (\ldots, u, \ldots, v, \ldots) \in R_i \text{ for some } i \}$

<table>
<thead>
<tr>
<th>Agent</th>
<th>Name</th>
<th>Drives</th>
</tr>
</thead>
<tbody>
<tr>
<td>007</td>
<td>James Bond</td>
<td>Aston Martin</td>
</tr>
<tr>
<td>200</td>
<td>Mr Smith</td>
<td>Cadillac</td>
</tr>
<tr>
<td>201</td>
<td>Mrs Smith</td>
<td>Mercedes</td>
</tr>
<tr>
<td>3</td>
<td>Jason Bourne</td>
<td>BMW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Car</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aston Martin</td>
<td>UK</td>
</tr>
<tr>
<td>Cadillac</td>
<td>USA</td>
</tr>
<tr>
<td>Mercedes</td>
<td>Germany</td>
</tr>
<tr>
<td>BMW</td>
<td>Germany</td>
</tr>
</tbody>
</table>
Hanf locality

Definition. The Gaifman graph of a structure $S = (V, R_1, ..., R_m)$ is the undirected graph $G(S) = (V, E)$ where $E = \{ (u, v) | \exists (..., u, ..., v, ...) \in R_i \text{ for some } i \}$

<table>
<thead>
<tr>
<th>Agent</th>
<th>Name</th>
<th>Drives</th>
</tr>
</thead>
<tbody>
<tr>
<td>007</td>
<td>James Bond</td>
<td>Aston Martin</td>
</tr>
<tr>
<td>200</td>
<td>Mr Smith</td>
<td>Cadillac</td>
</tr>
<tr>
<td>201</td>
<td>Mrs Smith</td>
<td>Mercedes</td>
</tr>
<tr>
<td>3</td>
<td>Jason Bourne</td>
<td>BMW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Car</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aston Martin</td>
<td>UK</td>
</tr>
<tr>
<td>Cadillac</td>
<td>USA</td>
</tr>
<tr>
<td>Mercedes</td>
<td>Germany</td>
</tr>
<tr>
<td>BMW</td>
<td>Germany</td>
</tr>
</tbody>
</table>
Definition. The Gaifman graph of a structure $S = (V, R_1, \ldots, R_m)$ is the undirected graph $G(S) = (V, E)$ where $E = \{ (u, v) \mid \exists (\ldots, u, \ldots, v, \ldots) \in R_i \text{ for some } i \}$.
Hanf locality

- \( \text{dist}(u, v) = \text{distance} \) between \( u \) and \( v \) in the Gaifman graph
- \( S[u, r] = \text{ball} \) around \( u \) of radius \( r \) = sub-structure induced by \( \{v \mid \text{dist}(u, v) \leq r\} \)

<table>
<thead>
<tr>
<th>Agent</th>
<th>Name</th>
<th>Drives</th>
<th>Car</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>007</td>
<td>James Bond</td>
<td>Aston Martin</td>
<td>Aston Martin</td>
<td>UK</td>
</tr>
<tr>
<td>200</td>
<td>Mr Smith</td>
<td>Cadillac</td>
<td>Cadillac</td>
<td>USA</td>
</tr>
<tr>
<td>201</td>
<td>Mrs Smith</td>
<td>Mercedes</td>
<td>Mercedes</td>
<td>Germany</td>
</tr>
<tr>
<td>3</td>
<td>Jason Bourne</td>
<td>BMW</td>
<td>BMW</td>
<td>Germany</td>
</tr>
</tbody>
</table>

\[
\text{dist}(u, v) = \text{distance} \quad \text{between} \quad u \quad \text{and} \quad v \quad \text{in the Gaifman graph}
\]
\[
S[u, r] = \text{ball} \quad \text{around} \quad u \quad \text{of radius} \quad r \quad = \text{sub-structure induced by} \quad \{v \mid \text{dist}(u, v) \leq r\}
\]
Hanf locality

- \( \text{dist}(u, v) = \text{distance} \) between \( u \) and \( v \) in the Gaifman graph
- \( S[u, r] = \text{ball around} \ u \ \text{of radius} \ r = \text{sub-structure induced by} \ \{ v \mid \text{dist}(u, v) \leq r \} \)

<table>
<thead>
<tr>
<th>Agent</th>
<th>Name</th>
<th>Drives</th>
</tr>
</thead>
<tbody>
<tr>
<td>007</td>
<td>James Bond</td>
<td>Aston Martin</td>
</tr>
<tr>
<td>200</td>
<td>Mr Smith</td>
<td>Cadillac</td>
</tr>
<tr>
<td>201</td>
<td>Mrs Smith</td>
<td>Mercedes</td>
</tr>
<tr>
<td>3</td>
<td>Jason Bourne</td>
<td>BMW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Car</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aston Martin</td>
<td>UK</td>
</tr>
<tr>
<td>Cadillac</td>
<td>USA</td>
</tr>
<tr>
<td>Mercedes</td>
<td>Germany</td>
</tr>
<tr>
<td>BMW</td>
<td>Germany</td>
</tr>
</tbody>
</table>
### Hanf locality

- \( \text{dist}(u, v) = \text{distance} \) between \( u \) and \( v \) in the Gaifman graph
- \( S[u, r] = \text{ball} \) around \( u \) of radius \( r \) = sub-structure induced by \( \{ v | \text{dist}(u, v) \leq r \} \)

<table>
<thead>
<tr>
<th>Agent</th>
<th>Name</th>
<th>Drives</th>
</tr>
</thead>
<tbody>
<tr>
<td>007</td>
<td>James Bond</td>
<td>Aston Martin</td>
</tr>
<tr>
<td>200</td>
<td>Mr Smith</td>
<td>Cadillac</td>
</tr>
<tr>
<td>201</td>
<td>Mrs Smith</td>
<td>Mercedes</td>
</tr>
<tr>
<td>3</td>
<td>Jason Bourne</td>
<td>BMW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Car</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aston Martin</td>
<td>UK</td>
</tr>
<tr>
<td>Cadillac</td>
<td>USA</td>
</tr>
<tr>
<td>Mercedes</td>
<td>Germany</td>
</tr>
<tr>
<td>BMW</td>
<td>Germany</td>
</tr>
</tbody>
</table>
**Hanf locality**

- \( \text{dist}(u, v) = \text{distance} \) between \( u \) and \( v \) in the Gaifman graph
- \( S[u, r] = \text{ball} \) around \( u \) of radius \( r \) = sub-structure induced by \( \{ v | \text{dist}(u, v) \leq r \} \)

<table>
<thead>
<tr>
<th>Agent</th>
<th>Name</th>
<th>Drives</th>
</tr>
</thead>
<tbody>
<tr>
<td>007</td>
<td>James Bond</td>
<td>Aston Martin</td>
</tr>
<tr>
<td>200</td>
<td>Mr Smith</td>
<td>Cadillac</td>
</tr>
<tr>
<td>201</td>
<td>Mrs Smith</td>
<td>Mercedes</td>
</tr>
<tr>
<td>3</td>
<td>Jason Bourne</td>
<td>BMW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Car</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aston Martin</td>
<td>UK</td>
</tr>
<tr>
<td>Cadillac</td>
<td>USA</td>
</tr>
<tr>
<td>u Mercedes</td>
<td>Germany</td>
</tr>
<tr>
<td>BMW</td>
<td>Germany</td>
</tr>
</tbody>
</table>

[Diagram of agent and car relationships]
**Hanf locality**

- \( \text{dist}(u, v) = \text{distance} \) between \( u \) and \( v \) in the Gaifman graph
- \( S[u, r] = \text{ball} \) around \( u \) of radius \( r = \) sub-structure induced by \( \{ v | \text{dist}(u, v) \leq r \} \)

<table>
<thead>
<tr>
<th>Agent</th>
<th>Name</th>
<th>Drives</th>
</tr>
</thead>
<tbody>
<tr>
<td>007</td>
<td>James Bond</td>
<td>Aston Martin</td>
</tr>
<tr>
<td>200</td>
<td>Mr Smith</td>
<td>Cadillac</td>
</tr>
<tr>
<td>201</td>
<td>Mrs Smith</td>
<td>Mercedes</td>
</tr>
<tr>
<td>3</td>
<td>Jason Bourne</td>
<td>BMW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Car</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aston Martin</td>
<td>UK</td>
</tr>
<tr>
<td>Cadillac</td>
<td>USA</td>
</tr>
<tr>
<td>Mercedes</td>
<td>Germany</td>
</tr>
<tr>
<td>BMW</td>
<td>Germany</td>
</tr>
</tbody>
</table>
Definition. Two structures $S_1$ and $S_2$ are **Hanf $(r, t)$-equivalent**

iff for each structure $B$, the two numbers

$$\# u \text{ s.t. } S_1 [u, r] \cong B \quad \# v \text{ s.t. } S_2 [v, r] \cong B$$

are *either the same* or *both* $\geq t$. 
Hanf locality

Definition. Two structures $S_1$ and $S_2$ are **Hanf ($r$, $t$)-equivalent**

iff for each structure $B$, the two numbers

$$\#u \text{ s.t. } S_1[u, r] \cong B \quad \#v \text{ s.t. } S_2[v, r] \cong B$$

are either the same or both $\geq t$.

Example. $S_1, S_2$ are Hanf ($1$, $1$)-equivalent iff they have the same balls of radius 1
Hanf locality

Definition. Two structures $S_1$ and $S_2$ are **Hanf ($r,t$)-equivalent** iff for each structure $B$, the two numbers

$$
\# u \text{ s.t. } S_1 [u, r] \cong B \quad \# v \text{ s.t. } S_2 [v, r] \cong B
$$

are either the same or both $\geq t$.

Example. $S_1, S_2$ are Hanf $(1, 1)$-equivalent iff they have the *same balls* of radius 1.
Hanf locality

Definition. Two structures $S_1$ and $S_2$ are **Hanf** $(r, t)$-equivalent

iff for each structure $B$, the two numbers

$$\#u \text{ s.t. } S_1[u, r] \cong B \quad \#v \text{ s.t. } S_2[v, r] \cong B$$

are either the same or both $\geq t$.

Example. $K_n, K_{n+1}$ are **not** Hanf $(1, 1)$-equivalent
Hanf locality

Definition. Two structures $S_1$ and $S_2$ are **Hanf** $(r, t)$ - equivalent iff for each structure $B$, the two numbers

$$\# u \text{ s.t. } S_1[u, r] \cong B \quad \# v \text{ s.t. } S_2[v, r] \cong B$$

are either the same or both $\geq t$.

Example. $K_n, K_{n+1}$ are **not** Hanf $(1, 1)$ - equivalent
**Theorem.** $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$.

[Hanf '60]
Theorem. $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$.

[Hanf '60]

Exercise: prove that acyclicity is not FO-definable (even on finite structures)
**Theorem.** $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$.

[Hanf '60]

**Exercise:** prove that *acyclicity* is not FO-definable (even on finite structures)
Hanf locality

**Theorem.** $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$.

[Hanf '60]

**Exercise:** prove that *acyclicity* is not FO-definable (even on finite structures)
Exercise: prove that *acyclicity* is not FO-definable (even on finite structures)
Hanf locality

**Theorem.** $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$.

[Hanf '60]

**Exercise:** prove that *acyclicity* is not FO-definable (even on finite structures)
**Theorem.** $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf$(r, t)$-equivalent, with $r = 3^n$ and $t = n$. [Hanf '60]

**Exercise:** prove that testing whether a binary tree is complete is not FO-definable.
**Theorem.** $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$.

[Hanf '60]

**Exercise:** prove that testing whether a binary tree is *complete* is not FO-definable
**Theorem.** $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$. 

[Hanf '60]
Theorem. $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$. [Hanf '60]
Theorem. \( S_1, S_2 \) are \( n \)-equivalent (they satisfy the same sentences with quantifier rank \( n \)) whenever \( S_1, S_2 \) are Hanf \((r, t)\)-equivalent, with \( r = 3^n \) and \( t = n \).

[Hanf '60]

Why so BIG?

Remember \( d_k(x, y) = \) “there is a path of length \( 2^k \) from \( x \) to \( y \)”
Theorem. $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$. [Hanf '60]

Remember $d_k(x, y) =$ “there is a path of length $2^k$ from $x$ to $y$”

$$d_0(x, y) = E(x, y), \text{ and}$$

$$d_k(x, y) = \exists z \left( d_{k-1}(x, z) \land d_{k-1}(z, y) \right)$$

$$qr(d_k) = k$$
**Theorem.** $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf $(r, t)$-equivalent, with $r = 3^n$ and $t = n$.

[Hanf '60]

Why so BIG?

Remember $d_k(x, y) = “$there is a path of length $2^k$ from $x$ to $y”$\)

\[
d_0(x, y) = E(x, y), \text{ and} \\
d_k(x, y) = \exists z \ (d_{k-1}(x, z) \land d_{k-1}(z, y))
\]

\[q_{r}(d_k) = k\]

2·2$^n$ + 1

2·2$^n$
Theorem. $S_1, S_2$ are $n$-equivalent (they satisfy the same sentences with quantifier rank $n$) whenever $S_1, S_2$ are Hanf($r, t$)-equivalent, with $r = 3^n$ and $t = n$.

[Hanf '60]

Why so BIG?

Remember $d_k(x, y) = \text{“there is a path of length } 2^k \text{ from } x \text{ to } y\text{”}$

$$d_0(x, y) = E(x, y), \text{ and}$$

$$d_k(x, y) = \exists z ( d_{k-1}(x, z) \land d_{k-1}(z, y) )$$

$qr(d_k) = k$

Not $(n+2)$-equivalent yet they have the same $2^n-1$ balls.
Gaifman locality

What about queries?

Eg: Is reachability expressible in FO?

What about equivalence on the same structure?

When are two points indistinguishable?
Gaifman locality

\[ S[(a_1, ..., a_n), r] = \text{induced substructure of } S \]

of elements at distance \( \leq r \) of some \( a_i \) in the Gaifman graph.
Gaifman locality

\[ S[(a_1, ..., a_n), r] = \text{induced substructure of } S \]

of elements at distance \( \leq r \) of some \( a_i \) in the Gaifman graph.
Gaifman locality

\[ S \left[ (a_1, ..., a_n), r \right] = \text{induced substructure of } S \]

of elements at distance \( \leq r \) of some \( a_i \) in the Gaifman graph.
For any $\phi \in \text{FO}$ of quantifier rank $k$ and structure $S$,

$$S[(a_1, \ldots, a_n), r] \cong S[(b_1, \ldots, b_n), r] \text{ for } r = 3^{k+1}$$

implies

$$(a_1, \ldots, a_n) \in \phi(S) \iff (b_1, \ldots, b_n) \in \phi(S)$$

**Idea:** If the neighbourhoods of two tuples are the same, the formula cannot distinguish them.
Gaifman locality vs Hanf locality

Difference between Hanf- and Gaifman-locality:

Hanf-locality relates two different structures,

Gaifman-locality talks about definability in one structure
Gaifman locality vs Hanf locality

Difference between Hanf- and Gaifman-locality:

Hanf-locality relates **two different structures**, while Gaifman-locality talks about definability in **one structure**.

Inside $S$,
\[ 3^{k+1} \text{-balls of } (a_1, \ldots, a_n) = 3^{k+1} \text{-balls of } (b_1, \ldots, b_n) \]
\[ \Downarrow \]
\[ (a_1, \ldots, a_n) \text{ indistinguishable from } (b_1, \ldots, b_n) \text{ through formulas of } qr \leq k \]
Gaifman locality vs Hanf locality

Difference between Hanf- and Gaifman-locality:

Hanf-locality relates **two different structures**,

- $S_1$ and $S_2$ have the same # of balls of radius $3^k$, **up to threshold $k**
- $\Downarrow$
- They verify the same sentences of $qr \leq k$

Gaifman-locality talks about definability in **one structure**

- Inside $S$,
  
  $3^{k+1}$-balls of $(a_1,\ldots,a_n) = 3^{k+1}$-balls of $(b_1,\ldots,b_n)$
  
  $\Downarrow$

  $(a_1,\ldots,a_n)$ indistinguishable from $(b_1,\ldots,b_n)$ through **formulas** of $qr \leq k$
A query \( Q(x_1, \ldots, x_n) \) is not FO-definable if:

for every \( k \) there is a structure \( S_k \) and \( (a_1, \ldots, a_n), (b_1, \ldots, b_n) \) such that

1. \( S_k [(a_1, \ldots, a_n), 3^{k+1}] \equiv S_k [(b_1, \ldots, b_n), 3^{k+1}] \)
2. \( (a_1, \ldots, a_n) \in Q(S_k), \ (b_1, \ldots, b_n) \notin Q(S_k) \)

Proof: If \( Q \) were expressible with a formula of quantifier rank \( k \), then \( (a_1, \ldots, a_n) \in Q(S_k) \) iff \( (b_1, \ldots, b_n) \in Q(S_k) \). Absurd!
Reachability is not FO definable.

For every \( k \), we build \( S_k \):
Reachability is not FO definable.

For every $k$, we build $S_k$:
Reachability is not FO definable.

For every $k$, we build $S_k$: 
Gaifman locality

Reachability is not FO definable.

For every $k$, we build $S_k$: 
Gaifman locality

Reachability is not FO definable.

For every $k$, we build $S_k$:

And $S_k [(a_1, a_2), 3^{k+1}] \cong S_k [(b_1, b_2), 3^{k+1}]$
Gaifman locality

Reachability is not FO definable.

For every $k$, we build $S_k$:

And $S_k [(a_1, a_2), 3^{k+1}] \equiv S_k [(b_1, b_2), 3^{k+1}]$

However,

- $b_2$ is reachable from $b_1$,
- $a_2$ is not reachable from $a_1$. 
Gaifman locality

Reachability is not FO definable.

For every $k$, we build $S_k$:

And $S_k [(a_1, a_2), 3^{k+1}] \equiv S_k [(b_1, b_2), 3^{k+1}]$

However,

- $b_2$ is reachable from $b_1$,
- $a_2$ is not reachable from $a_1$.

Your turn! $Q(x) = “x$ is a vertex separator”
Gaifman Theorem

Basic local sentence:

$$\exists x_1, \ldots, x_n \quad r \quad x_1 \quad r \quad x_2 \quad \cdots \quad r \quad x_n \quad \land \psi_1(x_1) \land \cdots \land \psi_n(x_n)$$
Gaifman Theorem

Basic local sentence:

$$\exists x_1, \ldots, x_n \quad \land \psi_1(x_1) \land \cdots \land \psi_n(x_n)$$

disjoint $r$-balls around $x_1, \ldots, x_n$
Gaifman Theorem

Basic local sentence:

$$\exists x_1, \ldots, x_n$$

$$\exists x_1, \ldots, x_n$$

$$r$$-local formulas

$$\land \psi_1(x_1) \land \cdots \land \psi_n(x_n)$$

disjoint $$r$$-balls around $$x_1, \ldots, x_n$$
Gaifman Theorem

Basic local sentence:

$$\exists x_1, \ldots, x_n \wedge \psi_1(x_1) \wedge \cdots \wedge \psi_n(x_n)$$

disjoint $r$-balls around $x_1, \ldots, x_n$

Inside $\psi_i(x_i)$ we interpret $\exists y. \phi$ as $\exists y. d(x_i, y) \leq r \wedge \phi$
Gaifman Theorem

Basic local sentence:

\[ \exists x_1, \ldots, x_n \\land \psi_1(x_1) \land \cdots \land \psi_n(x_n) \]

disjoint \( r \)-balls around \( x_1, \ldots, x_n \)

**Gaifman Theorem**: Every FO sentence is equivalent to a boolean combination of **basic local sentences**.
Recap

EF games

FO sentences with quantifier rank n

= winning strategies for Spoiler in the n-round EF game
Recap

**EF games**

FO sentences with quantifier rank \( n \)

= winning strategies for Spoiler in the \( n \)-round EF game

**0-1 Law**

FO sentences are almost always true or almost always false
Recap

**EF games**

FO sentences with quantifier rank n

= winning strategies for Spoiler in the n-round EF game

**0-1 Law**

FO sentences are almost always true or almost always false

**Hanf locality**

FO sentences with quantifier rank n

= counting $3^n$ sized balls up to n
Recap

**EF games**

FO sentences with quantifier rank $n$

= winning strategies for Spoiler in the $n$-round EF game

**0-1 Law**

FO sentences are almost always true or almost always false

**Hanf locality**

FO sentences with quantifier rank $n$

= counting $3^n$ sized balls up to $n$

**Gaifman locality**

Queries of quantifier rank $n$ output tuples closed under $3^{n+1}$ balls.
Recap

**EF games**

FO sentences with quantifier rank n

= winning strategies for Spoiler in the n-round EF game

**0-1 Law**

FO sentences are almost always true or almost always false

**Hanf locality**

FO sentences with quantifier rank n

= counting $3^n$ sized balls up to n

**Gaifman locality**

Queries of quantifier rank n output tuples closed under $3^{n+1}$ balls.

**Gaifman Theorem**

An FO sentence can only say

“there are some points at distance $\geq 2r$

whose $r$-balls are isomorphic to certain structures”

or a boolean combination of that.

• Otto, “Finite Model Theory”, Springer, 2005
  (freely available at www.mathematik.tu-darmstadt.de/~otto/LEHRE/FMT0809.ps)

  (available at www.math.helsinki.fi/logic/people/jouko.vaanananen/shortcourse.pdf)