SHACL: A Description Logic in Disguise

LPNMR

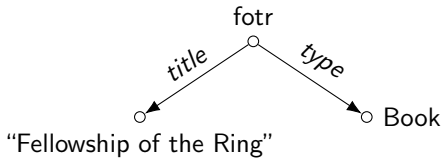
Bart Bogaerts, *Maxime Jakubowski* & Jan Van den Bussche

Vrije Universiteit Brussel & Universiteit Hasselt

September 2022

SHACL

- **Sha**pes **C**onstraint **L**anguage
- Constraint language for RDF graphs
- Conformance checking

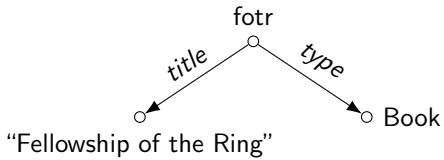


SHACL

- **Sha**pes **C**onstraint **L**anguage
- Constraint language for RDF graphs
- Conformance checking

```
:BookShape
  a sh:PropertyShape;
  sh:path :title;
  sh:minCount 1.

:BookShape sh:targetClass :Book
```

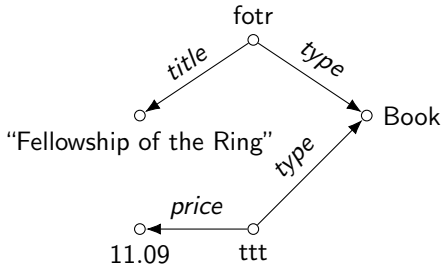


SHACL

- **Sha**pes **C**onstraint **L**anguage
- Constraint language for RDF graphs
- Conformance checking

```
:BookShape
  a sh:PropertyShape;
  sh:path :title;
  sh:minCount 1.

:BookShape sh:targetClass :Book
```



What is an appropriate formalization of SHACL?

SHACL shapes

The language $\mathcal{L}$

$$\phi ::= \top \mid \{c\} \mid \phi \wedge \phi \mid \phi \vee \phi \mid \neg\phi \mid \forall E.\phi \mid \geq_n E.\phi$$

$$E ::= p \mid p^- \mid E \cup E \mid E/E \mid E^*$$

SHACL shapes

The language $\mathcal{L}$

$$\phi ::= \top \mid \{c\} \mid \phi \wedge \phi \mid \phi \vee \phi \mid \neg\phi \mid \forall E.\phi \mid \geq_n E.\phi$$
$$E ::= p \mid p^- \mid E \cup E \mid E/E \mid E^*$$

An interpretation I:

- domain Δ^I
- interprets *node names*
- interprets *property names*

ϕ	$I, a \vDash \phi$ if:
$\{c\}$	$a = [\![c]\!]^I$
$\geq_n E.\psi$	$\sharp\{b \in [\![E]\!]^I(a) \mid I, b \vDash \psi\} \geq n$
$\forall E.\psi$	every $b \in [\![E]\!]^I(a)$ must $I, b \vDash \psi$

SHACL shapes

The language $\mathcal{L}(eq, disj, closed, ?)$

$$\phi ::= \top \mid \{c\} \mid \phi \wedge \phi \mid \phi \vee \phi \mid \neg\phi \mid \forall E.\phi \mid \geq_n E.\phi \mid eq(p, E) \mid disj(p, E) \mid closed(Q)$$
$$E ::= p \mid p^- \mid E \cup E \mid E/E \mid E^* \mid E?$$

Distinctive features:

- Equality
- Disjointness
- Closure
- Zero-or-one path

ϕ	$I, a \vDash \phi$ if:
$\{c\}$	$a = [\![c]\!]^I$
$\geq_n E.\psi$	$\sharp\{b \in [\![E]\!]^I(a) \mid I, b \vDash \psi\} \geq n$
$eq(E, p)$	the sets $[\![E]\!]^I(a)$ and $[\![p]\!]^I(a)$ are equal
$disj(E, p)$	the sets $[\![E]\!]^I(a)$ and $[\![p]\!]^I(a)$ are disjoint
$closed(R)$	$[\![p]\!]^I(a)$ is empty for each $p \in \Sigma - R$

SHACL shapes

The language $\mathcal{L}(eq, disj, closed, ?)$

$$\phi ::= \top \mid \{c\} \mid \phi \wedge \phi \mid \phi \vee \phi \mid \neg\phi \mid \forall E.\phi \mid \geq_n E.\phi \mid eq(p, E) \mid disj(p, E) \mid closed(Q)$$
$$E ::= p \mid p^- \mid E \cup E \mid E/E \mid E^* \mid E?$$

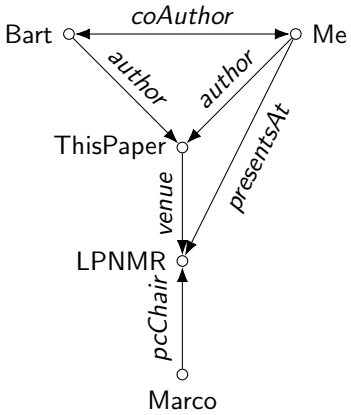
Distinctive features:

- Equality
- Disjointness
- Closure
- Zero-or-one path

ϕ	$I, a \vDash \phi$ if:
$\{c\}$	$a = [\![c]\!]^I$
$\geq_n E.\psi$	$\sharp\{b \in [\![E]\!]^I(a) \mid I, b \vDash \psi\} \geq n$
$eq(E, p)$	the sets $[\![E]\!]^I(a)$ and $[\![p]\!]^I(a)$ are equal
$disj(E, p)$	the sets $[\![E]\!]^I(a)$ and $[\![p]\!]^I(a)$ are disjoint
$closed(R)$	$[\![p]\!]^I(a)$ is empty for each $p \in \Sigma - R$

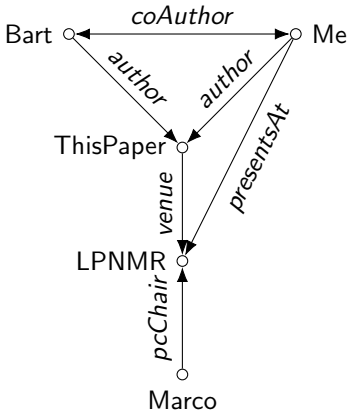
A *shape schema* is a set of shape inclusions, e.g., $\geq_1 type.\{Book\} \subseteq \geq_1 title.\top$

SHACL examples



$LPNMRAuthor \leftarrow \exists author/venue.\{LPNMR\}$

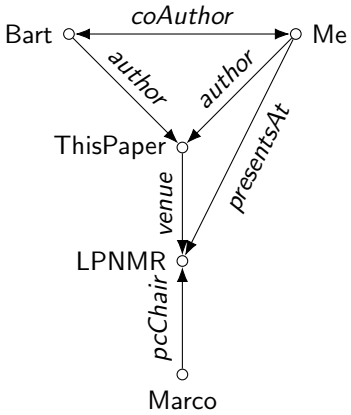
SHACL examples



$LPNMRAuthor \leftarrow \exists author/venue.\{LPNMR\}$

$\exists presentsAt.\{LPNMR\} \subseteq LPNMRAuthor$

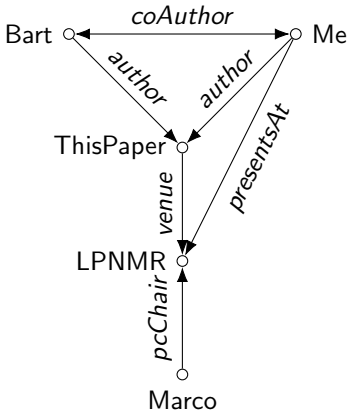
SHACL examples



$LPNMRAuthor \leftarrow \exists author/venue.\{LPNMR\}$
$NotLPNMRAuthor \leftarrow \neg LPNMRAuthor$

$\exists presentsAt.\{LPNMR\} \subseteq LPNMRAuthor$

SHACL examples



$LPNMRAuthor \leftarrow \exists author/venue.\{LPNMR\}$
$NotLPNMRAuthor \leftarrow \neg LPNMRAuthor$

$\exists presentsAt.\{LPNMR\} \subseteq LPNMRAuthor$
$\{Marco\} \subseteq NotLPNMRAuthor$

Observation: SHACL is a Description Logic!

Description Logics

Ontology / Knowledge Base

- Terminology (TBox): what are the concepts and their relations?
- Assertions (ABox): what is the known information?

Description Logics

Ontology / Knowledge Base

- Terminology (TBox): what are the concepts and their relations?
- Assertions (ABox): what is the known information?
- Example:

TBox:

$$Author \sqsubseteq Human \sqcap \exists hasWritten\,.\,Publication$$

ABox:

$$Author\ :\ tolkien$$

$$hasWritten\ :\ (tolkien, fotr)$$

SHACL as a Description Logic

- TBox is a finite set of shape inclusions, given by the shape schema
  - Definitions: :BookShape $\equiv \exists$:title . $\top$
  - Targeting: $\exists$:writtenBy . $\top \sqsubseteq$:BookShape
- There is no ABox

Different reasoning tasks:

- Normally, higher-order reasoning tasks: consistency, entailment, ...
- SHACL does *model checking*

SHACL as a Description Logic

- TBox is a finite set of shape inclusions, given by the shape schema
  - Definitions: :BookShape $\equiv \exists$:title . $\top$
  - Targeting: $\exists$:writtenBy . $\top \sqsubseteq$:BookShape
- There is no ABox

Different reasoning tasks:

- Normally, higher-order reasoning tasks: consistency, entailment, ...
- SHACL does *model checking*

... but what then is the interpretation?

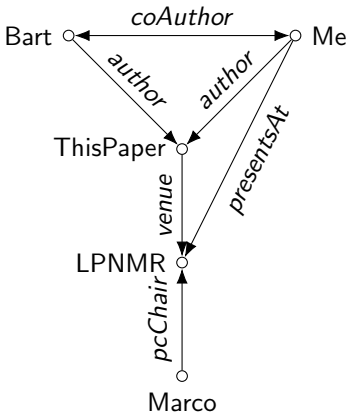
What's in an RDF graph?

- A graph is a finite set of *facts*
- A fact is of the form $p(a, b)$ with p a property name and a, b nodes of G.

We associate to any given graph an interpretation I:

- The domain is the universe of *all nodes*
- Every constant is interpreted as itself
- The interpretation of a property name is fixed by the facts

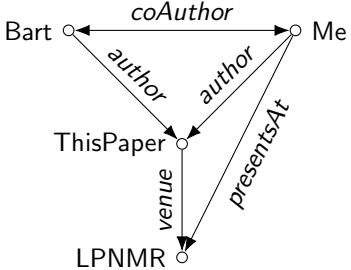
Illustration



$NotLPNMRAuthor \leftarrow \neg\exists author/venue.\{LPNMR\}$

$\{Marco\} \subseteq NotLPNMRAuthor$

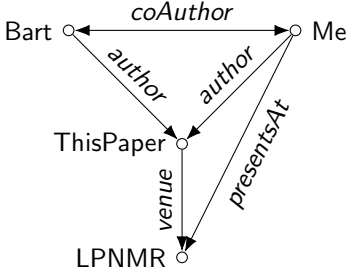
Illustration



$NotLPNMRAuthor \leftarrow \neg\exists author/venue.\{LPNMR\}$

$\{Marco\} \subseteq NotLPNMRAuthor$

Illustration

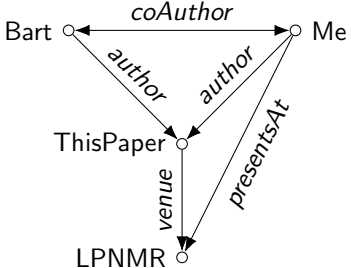


$NotLPNMRAuthor \leftarrow \neg\exists author/venue.\{LPNMR\}$

$\{Marco\} \subseteq NotLPNMRAuthor$

- *NotLPNMRAuthor* evaluates to $N - \{Bart, Me\}$

Illustration

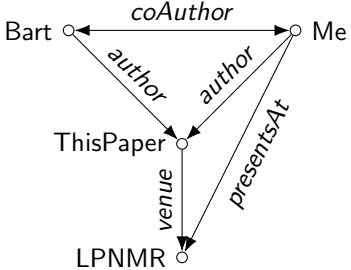


$NotLPNMRAuthor \leftarrow \neg\exists author/venue.\{LPNMR\}$

$\{Marco\} \subseteq NotLPNMRAuthor$

- $NotLPNMRAuthor$ evaluates to $N - \{Bart, Me\}$
  ... because the domain is the universe of all nodes (N)

Illustration

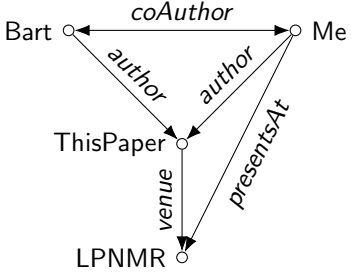


$NotLPNMRAuthor \leftarrow \neg\exists author/venue.\{LPNMR\}$

$\{Marco\} \subseteq NotLPNMRAuthor$

- $NotLPNMRAuthor$ evaluates to $N - \{Bart, Me\}$
  ... because the domain is the universe of all nodes (N)
- $\{Marco\}$ evaluates to $\{Marco\}$

Illustration

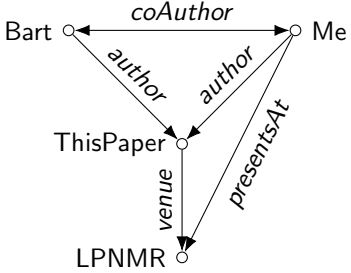


$NotLPNMRAuthor \leftarrow \neg\exists author/venue.\{LPNMR\}$

$\{Marco\} \subseteq NotLPNMRAuthor$

- $NotLPNMRAuthor$ evaluates to $N - \{Bart, Me\}$
  ... because the domain is the universe of all nodes (N)
- $\{Marco\}$ evaluates to $\{Marco\}$
  ... because all constants are interpreted as themselves

Illustration



$NotLPNMRAuthor \leftarrow \neg\exists author/venue.\{LPNMR\}$

$\{Marco\} \subseteq NotLPNMRAuthor$

- $NotLPNMRAuthor$ evaluates to $N - \{Bart, Me\}$
  ... because the domain is the universe of all nodes (N)
- $\{Marco\}$ evaluates to $\{Marco\}$
  ... because all constants are interpreted as themselves

$\Longrightarrow$ This is also the behavior of **real** SHACL!

Question: what about expressiveness?

Expressiveness

There is no shape that can express these concepts:

4-clique

"The node belongs to a 4-clique"

- SHACL is subsumed by *3-variable infinitary logic*
- Known to be not expressible

Majority

"A conference has more attendees than it has papers"

- SHACL is subsumed by *first-order logic*
- Known to be not expressible

SHACL shapes

The language $\mathcal{L}(eq, disj, closed, ?)$

$$\phi ::= \top \mid \{c\} \mid \phi \wedge \phi \mid \phi \vee \phi \mid \neg\phi \mid \forall E.\phi \mid \geq_n E.\phi \mid eq(p, E) \mid disj(p, E) \mid closed(Q)$$
$$E ::= p \mid p^- \mid E \cup E \mid E/E \mid E^* \mid E?$$

E are regular path queries with inverse

Distinctive features:

- Equality
- Disjointness
- Closure
- Zero-or-one path

ϕ	$I, a \vDash \phi$ if:
$\{c\}$	$a = [\![c]\!]^I$
$\geq_n E.\psi$	$\sharp\{b \in [\![E]\!]^I(a) \mid I, b \vDash \psi\} \geq n$
$eq(E, p)$	the sets $[\![E]\!]^I(a)$ and $[\![p]\!]^I(a)$ are equal
$disj(E, p)$	the sets $[\![E]\!]^I(a)$ and $[\![p]\!]^I(a)$ are disjoint
$closed(R)$	$[\![p]\!]^I(a)$ is empty for each $p \in \Sigma - R$

Relative Expressiveness

For each distinguished feature $X \in \{eq, disj, closed, ?\}$ we define a class of graphs Q_X such that:

- Q_X is definable by a simple shape constraint using only the feature X
- Q_X is **not** definable without X

For example, let $X = eq$:

- Q_{eq} is the class of symmetric graphs
- Q_{eq} is expressible with the constraint $\exists r.\top \subseteq eq(r, r^-)$
- Q_{eq} is not expressible without eq

Question: semantics of recursion?

Fixpoint semantics

Method:

- We need an operator on interpretations
- Define recursive semantics using fixpoints of this operator

By applying Approximation Fixpoint Theory we obtain known semantics for SHACL like stable models and well-founded models

- No need to reinvent semantics
- We directly obtain strong formal foundation for the study of recursive SHACL

Conclusion

- It is of value to put emphasis on the formalization of SHACL
  - What does the RDF graph represent?
  - What are the exact semantics of the language?
- In the semantic web, both OWL and SHACL are used for modeling tasks
  - OWL has its logical foundations in Description Logic
  - ... according to us, SHACL does too
- With a proper formalization, we can better study SHACL