FLOW-SHACL Dataflow-Driven SHACL Validation and Inference System

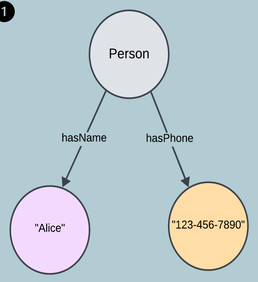
Sowndarya Krishnan Navaneetha Kannan, Dr. Gabe Fierro

DEPARTMENT OF COMPUTER SCIENCE
COLORADO SCHOOL OF MINES
EARTH • ENERGY • ENVIRONMENT

Background

Knowledge graph: Structured data connecting entities with relationships.

1



2

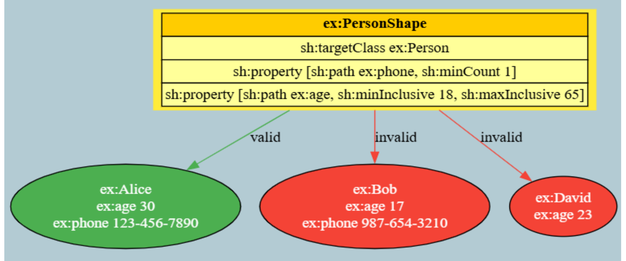
Resource description framework – standard way to define Kgs

RDF graph: ex:Person ex:hasName "Alice" ; ex:hasPhone "123-456-7890" .

3

Shapes Constraint Language- Integrity Constraints on the RDF Kgs

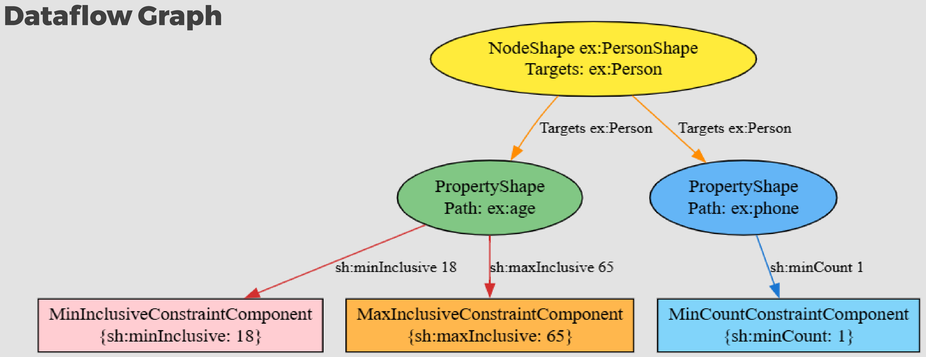
Example Execution



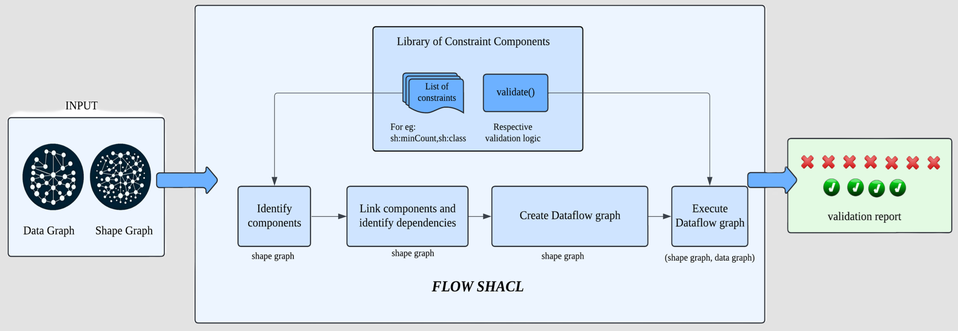
Abstract

Flow-SHACL addresses the performance limitations of existing SHACL validators, particularly when validating large-scale RDF data. SHACL (Shapes Constraint Language) is a W3C standard used to validate RDF data against predefined "shapes." Traditional SHACL validators struggle with performance as knowledge graphs scale to millions or billions of nodes. Flow-SHACL overcomes this challenge by utilizing a dataflow-based approach, which constructs an optimized data structure that represents SHACL operations as a dataflow graph. This novel approach not only enhances validation performance but also supports inference capabilities, providing deeper insights and enabling scalable solutions for handling large knowledge graphs efficiently.

Dataflow Graph



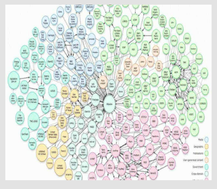
System Architecture



Next Steps

- ❑ Fully complete the SHACL core features.
- ❑ Profile the system to identify time consumption across different parts of the process
- ❑ Benchmark its performance for better comparison and optimization.
- ❑ Expand as an inference engine, helping users discover hidden relationships and insights in their knowledge graphs.

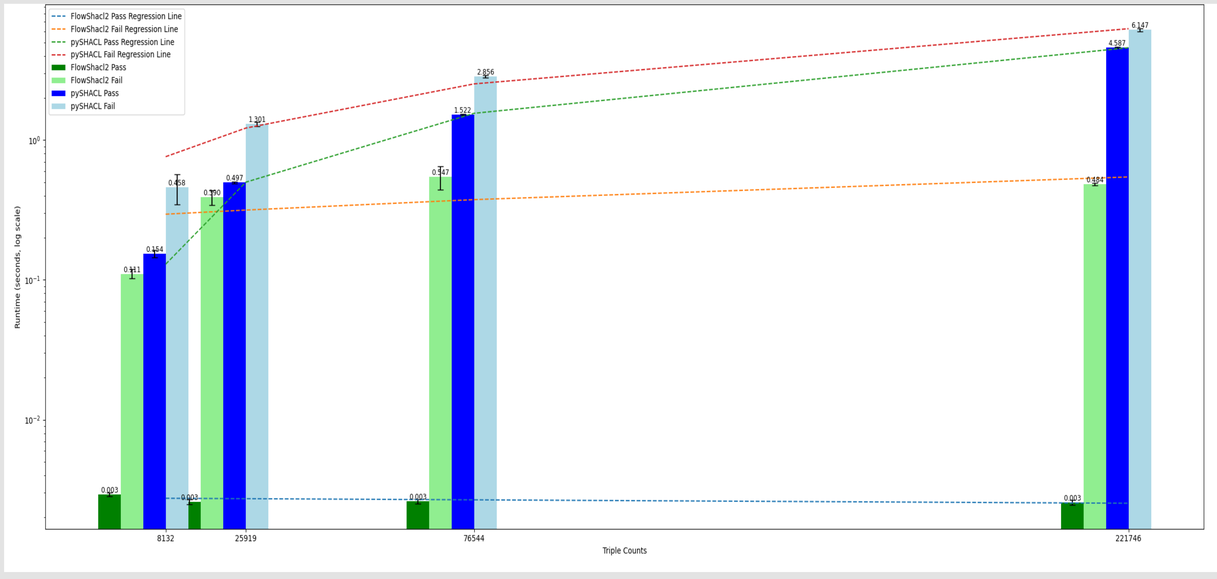
Introduction



Knowledge graphs are widely used in domains like web data (e.g., DBpedia, Wikidata) and industry-specific applications. Despite their utility, they often face quality issues, scalability challenges, and performance bottlenecks in validation. Flow-SHACL addresses these limitations by enhancing SHACL (Shapes Constraint Language) validation, a W3C standard for verifying RDF data. Traditional SHACL validators struggle with large-scale graphs, leading to inefficiencies. Flow-SHACL introduces a dataflow-based approach, optimizing validation by mapping SHACL operations as a dataflow graph. This innovation ensures faster validation for large graphs, seamless handling of massive datasets, and improved reasoning capabilities through efficient inference, offering a robust and scalable solution for modern knowledge graphs.

Preliminary Results

This comparison shows the performance of FlowSHACL versus pySHACL, the widely-used, fully-featured SHACL validator. Please keep in mind that FlowSHACL currently only handles basic SHACL core features and property paths, with `sh:qualifieds` and other advanced features still to be implemented.



Conclusion

- ❑ FLOW-SHACL revolutionizes SHACL validation, enhancing speed and scalability through novel data-flow approach.
- ❑ Sets foundation for future advancements, including development of inference system for deeper insights.
- ❑ Aims to transform how knowledge graphs are utilized and understood in data-driven environments.

Contact

❑ Sowndarya Krishnan Navaneetha Kannan
Email: sowndaryakrishnanna@mines.edu

❑ Dr. Gabe Fierro
Email: gtfierro@mines.edu

MINES | Computer Science