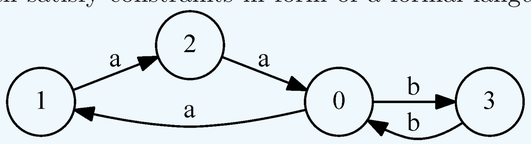
Context-Free Path Querying with All-Path Semantics by Matrix Multiplication

Rustam Azimov
JetBrains Research, Saint Petersburg University, Russia
Rustam.Azimov@jetbrains.com

All-Path CFPQ

Find all paths which satisfy constraints in form of a formal language



$L = \{a^n b^n \mid n > 0\}$
Query = grammar for L: $S \to a\ b \mid a\ S\ b$

In some cases we want to find all data dependencies, for example in static code analysis when searching for vulnerabilities

Results

- We provide the matrix-based algorithm for CFPQ with all-path query semantics
- We implement the provided algorithm using the GraphBLAS API
- The proposed matrix-based solution for all-path query semantics compared to the Kronecker product-based solution consumes more memory, but allows one to extract paths significantly faster

Future Research

- We plan to obtain GPU-based and distributed implementations
- Try to update the query results dynamically when data changes
- We plan to provide the multiple-source modifications for all linear algebra-based CFPQ algorithms
- Find new applications that required CFPQ

Matrix-Based Algorithm [2]

- We store the additional information about the paths found as the sets of the intermediate vertices
- We introduce the following matrix multiplication operation
- $T^A \odot T^B = T^C$ where $T^C_{i,j} = \bigcup_{k=1}^{n} (T^A_{i,k} \otimes T^B_{k,j})$ and $T^A_{i,k} \otimes T^B_{k,j} = \begin{cases} \{k\}, & \text{if } T^A_{i,k} \neq \emptyset \wedge T^B_{k,j} \neq \emptyset \\ \emptyset, & \text{otherwise} \end{cases}$

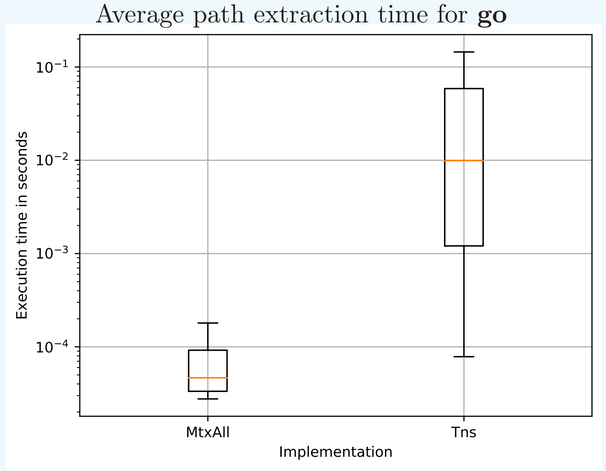
Path Extraction

- After constructing a set of matrices with sets of intermediate vertices, we can extract all required paths $i \pi j$ for every vertex pair i, j if such paths exist
- It is assumed that the sets of paths are computed lazily, to ensure the termination in case of an infinite number of paths

CFPQ Evaluation with Relational, Single-Path and All-Path Query Semantics

Graph	#V	#E	MtxRel		MtxSingle		MtxAll		Tns	
			Time	Mem	Time	Mem	Time	Mem	Time	Mem
eclass_514en	239 111	523 727	0.06	181	0.16	216	0.22	126	0.27	193
go	272 770	534 311	0.94	246	0.93	217	1.13	990	1.27	243
geospecies	450 609	2 311 461	7.48	7645	15.54	22941	32.06	44235	26.32	19537
taxonomy	5 728 398	14 922 125	0.72	1175	1.15	2250	3.84	1507	3.56	1776

- Time in seconds and memory is measured in megabytes
- Graph: real-world ontologies (RDFs), query: same-generation query
- Example of a grammar: $S \to scor\ S\ sco \mid tr\ S\ t \mid scor\ sco \mid tr\ t$



- **MtxAll** constructs index up to 2-3 times slower and consumes more memory than **MtxSingle**
- If we must extract paths many times for a once constructed index then **MtxAll** is preferable than **Tns**

Contact Us

- Semyon Grigorev: s.v.grigoriev@spbu.ru
- Rustam Azimov:Rustam.Azimov@jetbrains.com
- Ilya Epelbaum: iliyepelbaun@gmail.com

- Dataset: `https://github.com/JetBrains-Research/CFPQ_Data`
- Implementations: `https://github.com/JetBrains-Research/CFPQ_PyAlgo`

References

[1] Rustam Azimov and Semyon Grigorev. Context-free path querying by matrix multiplication. In *Proceedings of the 1st ACM SIGMOD Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA)*, GRADES-NDA '18, pages 5:1–5:10, 2018.

[2] Rustam Azimov, Ilya Epelbaum, and Semyon Grigorev. Context-free path querying with all-path semantics by matrix multiplication. In *Proceedings of the 4th Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA)*, GRADES-NDA '21, 2021.

[3] Egor Orachev, Ilya Epelbaum, Rustam Azimov, and Semyon Grigorev. Context-free path querying by kronecker product. In *European Conference on Advances in Databases and Information Systems*, pages 49–59. Springer, 2020.