&

	Jan 15	Jan 16	Jan 17	Jan 18	Jan 19
	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
7:30-8:45	Breakfast	Breakfast	Breakfast	Breakfast	Breakfast
9h00-9h30	**Intro (one min slide + write name on board)**	**Session: Model counting, Shapley, Probabilistic DB** (38 + 7 mins talks) -- Dan S. -- Antoine	**Session: Enumeration and direct access** -- Cristian – Benny – Nikos	**Session:Factorized DB, Probabilistic DB + OPEN PROBLEMS** -- Christoph B. -- Pierre S. **– OPEN PROBLEMS**	**Session: Discussions & future plan**
9h30-10h00	**Session: Semiring and Datalog** – Reinhard				
10h00-10h30					free collaboration
10h30-11h00	Coffee	Coffee	Coffee	Coffee	Coffee
11h00-11h30	**Session: Explanations overview** -- Sudeepa – Bernardo/ Pablo	**Session: Probabilistic DB and direct access** – Mikael -- Stefan	**Session: Circuits** -- Pierre B. -- Florent	**Session: Causal inference for explanations** -- Babak -- Amir	free collaboration
11h30-12h00					
12h15-14h00	Lunch	Lunch	Lunch	Lunch	Lunch
14h00-15h15	**Session: Shapley Values** – (Logistics) -- Pablo -- Ester	**Session: Monoids and Semirings** -- Liat -- Albert – Reinhard (contd.)	~~**Excursion**~~ (bad weather) free collaboration	**Session: Short ad-hoc talks** – Bertram – Christoph S. – Boris	**Departure**
15h15-16h15	Coffee & Cake	Coffee & Cake		Coffee & Cake	
16h15-17h30	**Session: Short talks** (18+2 mins each) -- Floris -- Wolfgang -- Leo	free collaboration		free collaboration	
18:00	Dinner	Dinner	Dinner	Dinner	

TALK DETAILS

1. **Reinhard Pichler**
   TU Wien, Austria (https://dbai.tuwien.ac.at/staff/pichler/)

Title: Datalog over (Pre-)Semirings

Abstract:
Datalog is a successful query language that extends relational calculus by recursion, has an elegant declarative semantics as well as a simple operational semantics, and admits several powerful optimizations such as semi-naive evaluation and magic set rewriting. However, datalog also has its limitations since it only supports monotone queries over sets. This means, for instance, that aggregates (which are crucial in many data analytics tasks but are not monotone under set inclusion) are not supported in pure datalog.

In a seminal paper by Green, Karvounarakis, and Tannen at PODS 2007, K-relations were introduced as a generalization of standard relations. In a K-relation, tuples are mapped to some semiring K. We can then consider standard relations as K-relations over the Boolean semiring, bags of tuples as K-relations over the natural numbers, sparse tensors as K-relations over the reals, etc. Also provenance information at various levels of detail can be captured by an appropriate choice of the semiring K.

In this talk, I will present our recent work on the query language datalogo, which is based on the concept of K-relations and generalizes datalog to (pre-)semirings. In particular, I will show how it can capture various computations involving aggregates as well as provenance information. Moreover, I will briefly mention convergence properties of datalogo and some optimization techniques.

Relevant papers:

- Todd J. Green, Gregory Karvounarakis, Val Tannen: Provenance semirings. PODS 2007: 31-40: https://dl.acm.org/doi/10.1145/1265530.1265535
- Yisu Remy Wang, Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu: Optimizing Recursive Queries with Progam Synthesis. SIGMOD Conference 2022: 79-93: https://dl.acm.org/doi/10.1145/3514221.3517827
- Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu, Yisu Remy Wang: Convergence of Datalog over (Pre-) Semirings. PODS 2022: 105-117: https://dl.acm.org/doi/10.1145/3517804.3524140
  full version (to appear in J.ACM): https://arxiv.org/abs/2105.14435
- Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu, Yisu Remy Wang: Datalog in Wonderland. SIGMOD Rec. 51(2): 6-17 (2022): https://dl.acm.org/doi/10.1145/3552490.3552492

2. **Sudeepa Roy**
   Duke University, USA (https://users.cs.duke.edu/~sudeepa/)

Title: Explanations for Aggregate Query Answers - An Overview

Abstract: I will give an overview of different types of explanations for aggregate query answers answering user questions like why a value is high/low or higher/lower than another value. I will discuss explanations by intervention, counterbalance, augmented provenance, causal explanations, and actionable explanations. Explanations by Shapley Value will be covered in other talks.

Relevant papers:
SIGMOD'14: [Link]
SIGMOD'19: [Link]
SIGMOD'21: [Link]
SIGMOD'22: [Link]
(also SIGMOD'24 - causal explanations).

3. **Bernardo Subercaseaux (Talk given by Pablo Barcelo)**
   Title: Model Interpretability through the Lens of Computational Complexity

   Abstract: This talk revisits a framework for judging and comparing the interpretability of classes of Machine Learning models. Said framework allows us to formalize and prove a nuanced version of claims like "decision trees are more interpretable than neural networks". Interestingly, such a formalization pointed out the first result establishing the hardness of interpreting decision trees, and provided tools to analyze how hyper-parameters such as the number of layers in a network can impact its interpretability. Our framework relied on a few assumptions that will be discussed explicitly in the talk, such as the role of well-defined interpretability queries or the adequacy of computational complexity for capturing the practical complexity of real-life instances.

   Relevant papers:
   - Model Interpretability through the Lens of Computational Complexity. Pablo Barceló, Mikaël Monet, Jorge Pérez, Bernardo Subercaseaux. (https://arxiv.org/abs/2010.12265)
   - On Computing Probabilistic Explanations for Decision Trees. Marcelo Arenas, Pablo Barceló, Miguel Romero, Bernardo Subercaseaux. (https://arxiv.org/abs/2207.12213)

4. **Pablo Barcelo**
   Title: SHAP-Scores and Its Computation over ML Models

Abstract: SHAP scores are expressions designed to capture the contribution of a feature to the output of a machine learning model. They are grounded in the well-studied game-theoretical notion of Shapley values. In this discussion, I will elucidate the meaning of these SHAP scores expressions and explain how they are obtained from first principles. Subsequently, I will delve into the examination of the problem of computing SHAP scores over machine learning models. I will provide insights into when and why this problem becomes computationally intractable. Additionally, I will identify a large and practically relevant class of models for which the problem can be solved in polynomial time. Finally, I will show that even for slight extensions of this class, the computation of SHAP scores is not only intractable but also does not admit a Fully Polynomial Randomized Approximation Scheme (FPRAS).

Relevant papers:

Marcelo Arenas, Pablo Barceló, Leopoldo E. Bertossi, Mikaël Monet:
On the Complexity of SHAP-Score-Based Explanations: Tractability via Knowledge Compilation and Non-Approximability Results. J. Mach. Learn. Res. 24: 63:1-63:58 (2023)

5. **Ester Livshits**

University of Edinburgh, UK
https://esterliv.com/

Title: The Shapley Value in Database Management

Abstract: We consider two situations where we wish to quantify the responsibility of individual database tuples to the outcome. The first is query answering, where we wish to provide an explanation as to why we obtained a specific answer. The second is database inconsistency, where the goal is to identify the most problematic tuples. Some tuples may contribute more than others to the outcome, which can be a bit in the case of a Boolean query, a tuple or a number for conjunctive and aggregate queries, respectively, or a number indicating how inconsistent the database is (i.e., an inconsistency measure). To quantify the contribution of tuples, we use the well-known Shapley value that was introduced in cooperative game theory in the 1950s and has found applications in a plethora of domains. We investigate the applicability of the Shapley value in the two settings, as well as the computational aspects of its calculation in terms of complexity, algorithms, and approximation.

Relevant papers:

- Ester Livshits, Leopoldo E. Bertossi, Benny Kimelfeld, and Moshe Sebag. “The Shapley Value of Tuples in Query Answering”. Logical Methods in Computer Science (2021). https://lmcs.episciences.org/8437

- Ester Livshits and Benny Kimelfeld. “The Shapley Value of Inconsistency Measures for Functional Dependencies”. Logical Methods in Computer Science (2022). https://lmcs.episciences.org/9705

6. **Floris Geerts**
University of Antwerp
https://fgeerts.github.io/

~~Title: Graph Explainability Explained~~

~~Abstract: We delve into diverse approaches within graph learning explainability, hereby aiming to establish meaningful connections with explainability in database research.~~

~~Relevant paper(s):~~
- ~~*A Survey on Explainability of Graph Neural Networks.* Jaykumar Kakkad, Jaspal Jannu, Kartik Sharma, Charu Aggarwal, Sourav Medya https://arxiv.org/abs/2306.01958~~

Title: Graph Explainability and Shapley

Graph explainability is a critical aspect in understanding and interpreting complex relationships within graph-structured data. The need for transparent and interpretable models has led to the exploration of various methodologies, with a focus on providing insights into the contribution of individual nodes or edges in a graph. Shapley values, inspired by cooperative game theory, offer a principled approach to attribute values to each node, reflecting their marginal contributions to different coalitions. Myerson value further refines this concept by considering the externalities of a coalition, providing a more comprehensive understanding of node importance. In the context of graph explainability, Hamiache and Navarro score introduces a novel perspective by evaluating the relevance of nodes based on the information flow and connectivity patterns, offering a nuanced interpretation of their impact on the overall graph structure. Together, these approaches contribute to the development of explainable graph models, enabling stakeholders to gain deeper insights into the dynamics and significance of individual elements within complex graph data.

Relevant papers:
- GStarX: Explaining Graph Neural Networks with Structure-Aware Cooperative Games, Shichang Zhang Yozen Liu Neil Shah Yizhou Sun https://arxiv.org/pdf/2201.12380.pdf

7. **Wolfgang Gatterbauer**

Northeastern University, Boston:
https://gatterbauer.name/

Title: Factorization of Provenance as a Reverse Data Management problem

Abstract: The talk outlines a connection between the problem of finding minimal size provenance factorizations and well-known reverse data management problems such as resilience. This connection is possible through insights on the relationship between provenance factorizations and minimal query plans (as originally defined for probabilistic databases). We show that by this connection a Max-Cut Min-Flow algorithm can be used to solve all known PTIME cases for this problem.

Relevant papers:

- Neha Makhija, Wolfgang Gatterbauer: Towards a Dichotomy for Minimally Factorizing the Provenance of Self-Join Free Conjunctive Queries. https://arxiv.org/pdf/2105.14307
- Neha Makhija, Wolfgang Gatterbauer: A Unified Approach for Resilience and Causal Responsibility with Integer Linear Programming (ILP) and LP Relaxations. SIGMOD'24. https://arxiv.org/pdf/2212.08898, https://github.com/northeastern-datalab/resilience-responsibility-ilp, https://northeastern-datalab.github.io/unified-reverse-data-management/

8. **Leo Bertossi**
SKEMA Business School Canada (Montreal)
https://people.scs.carleton.ca/~bertossi/

Title: Tractability and Optimization of Shap-Score Computation for Explainable AI

Abstract: We will present recent research on the Shap Scores in Explainable Machine Learning, concentrating on the tractability of Shap on open-box classifiers defined by interesting classes of Boolean circuits, and its application to binary neural networks via knowledge compilation techniques.

Relevant papers:

- Leopoldo Bertossi and Jorge E. Leon. "Efficient Computation of Shap Explanation Scores for Neural Network Classifiers via Knowledge Compilation". Proc. of JELIA'23, Springer LNCS 14281, 2023, pp. 49-64. https://arxiv.org/pdf/2303.06516.pdf

- Marcelo Arenas, Pablo Barcelo, Leopoldo Bertossi, Mikael Monet. "On the Complexity of SHAP-Score-Based Explanations: Tractability via Knowledge Compilation and Non-Approximability Results". Journal of Machine Learning Research, 2023, 24(63):1-58. https://jmlr.org/papers/volume24/21-0389/21-0389.pdf

9. **Dan Suciu**
   University of Washington (https://homes.cs.washington.edu/~suciu/)

   Title: From Shapley Value to Model Counting and Back

   Abstract: We study the problem of quantifying the contribution of each Boolean variable to the satisfying assignments of a Boolean function, based on the Shapley value. This problem was introduced by Livshits et al. in order to quantify the contribution of an input tuple to the output of a query. We prove polynomial-time equivalence between computing Shapley values and model counting, for classes of Boolean functions that are closed under substitutions of variables with disjunctions of fresh variables. This result settles an open problem raised by Deutch et at, which sought to connect the Shapley value computation to probabilistic query evaluation.

   Relevant papers:

   Ester Livshits, Leopoldo E. Bertossi, Benny Kimelfeld, Moshe Sebag:
   The Shapley Value of Tuples in Query Answering. Log. Methods Comput. Sci. 17(3) (2021)

   Daniel Deutch, Nave Frost, Benny Kimelfeld, Mikaël Monet:
   Computing the Shapley Value of Facts in Query Answering. SIGMOD Conference 2022: 1570-1583

   Ahmet Kara, Dan Olteanu, Dan Suciu:
   From Shapley Value to Model Counting and Back. CoRR abs/2306.14211 (2023)
   (to appear in PODS'2024)

10. **Antoine Amarilli, Télécom Paris, https://a3nm.net/**
    Title: Lower Bounds on Probabilistic Query Evaluation

    Abstract: This talk focuses on the task of computing the probability that a fixed query holds on an input probabilistic database. The problem can also be specialized to several contexts, e.g., computing the probability that an input graph with probabilistic edges contains a specific pattern, or in the unweighted case counting how many subgraphs of the input have a certain property.

We will review recent hardness results on this problem. We will cover two kinds of results: lower bounds on the computational complexity of the problem, and lower bounds on the size of the query provenance when represented in structured circuit classes.

Relevant papers:

Antoine Amarilli, Timothy van Bremen, Kuldeep S. Meel.
Conjunctive Queries on Probabilistic Graphs: The Limits of Approximability.
ICDT 2024.

Antoine Amarilli.
Uniform Reliability for Unbounded Homomorphism-Closed Graph Queries.
ICDT 2023. [slides, video]

Antoine Amarilli, Benny Kimelfeld.
Uniform Reliability of Self-Join-Free Conjunctive Queries.
LMCS, 2022. [conference version]

Antoine Amarilli, Mikaël Monet.
Weighted Counting of Matchings in Unbounded-Treewidth Graph Families.
MFCS 2022. [slides by Mikaël Monet, code]

11. **Mikael Monet**
Inria Lille, France (https://mikael-monet.net/en/home.html)

Title: The Intensional-Extensional Problem in Probabilistic Databases

Abstract: Dalvi and Suciu established a dichotomy for probabilistic query evaluation (PQE) over tuple-independent databases, for unions of conjunctive queries (UCQs): for each UCQ, the problem is either solvable in PTIME, or is #P-hard. The UCQs for which the problem is in PTIME are called *safe*. Dalvi and Suciu's algorithm on such a safe query relies essentially on the following three probabilistic rules: Independence, Negation, and Inclusion-Exclusion. In parallel, another method to obtain PTIME algorithms for PQE is through *knowledge compilation*: one first compiles the provenance of a query Q on a TID D as a Boolean circuit or diagram from the field of knowledge compilation (e.g., OBDDs, FBDDs, d-DNNFs, etc), and then uses this circuit to compute the probability. At a high-level, this type of algorithm makes use of the following three probabilistic rules: Independence, Negation, and *Disjoint union* (instead of inclusion-exclusion). This naturally leads to the following question, called the intensional-extensional problem: letting Q be a safe UCQ, can the tractability of PQE(Q) be captured with the knowledge compilation approach?

In this talk I will talk about this problem, present a technique that allowed to handle a specific class of UCQs, and discuss our ongoing work on the problem. In particular, I will present a neat combinatorial conjecture, that we named the "non-cancelling intersections" conjecture, that talks only about sets and the so-called Möbius function (i.e., no databases, no queries, no complexity). This talk is based on ongoing work with Antoine Amarilli, Louis Jachiet, and Dan Suciu.

Relevant papers:

- Abhay Kumar Jha, Dan Suciu: Knowledge Compilation Meets Database Theory: Compiling Queries to Decision Diagrams. Theory Comput. Syst. 52(3): 403-440 (2013)
- Solving a Special Case of the Intensional vs Extensional Conjecture in Probabilistic Databases. PODS 2020: 149-163

12. Stefan Mengel

CNRS, CRIL
http://www.cril.univ-artois.fr/~mengel/

Title: Impact of Self-Joins on Enumeration and Direct Access on Join Queries

Abstract: It has been known essentially since the introduction of conjunctive queries that self-joins have an impact on the evaluation of join queries. While in settings like answering Boolean queries and counting their complexity implications are completely understood, the situation is far less clear for other query answering tasks. In this talk, I will present some recent progress for enumeration (Carmeli and Segoufin 2023) and direct access (Bringmann, Carmeli, and Mengel 2023) showing that, even though these settings are often conceptually very close, self-joins behave very differently for them.

Relevant papers:

Karl Bringmann, Nofar Carmeli, Stefan Mengel:
Tight Fine-Grained Bounds for Direct Access on Join Queries.
https://arxiv.org/abs/2201.02401

Nofar Carmeli, Luc Segoufin:
Conjunctive Queries With Self-Joins, Towards a Fine-Grained Enumeration Complexity Analysis. PODS 2023: 277-289
https://arxiv.org/abs/2206.04988

13. Liat Peterfreund

The Hebrew University of Jerusalem, https://sites.google.com/view/liatpeterfreund/

Title:Revisiting Semiring Provenance for Datalog

Abstract: While the definition of semiring provenance is uncontroversial for unions of conjunctive queries, the picture is less clear for Datalog. Indeed, the original definition might include infinite computations, and is not consistent with other proposals for Datalog semantics over annotated data. In this work, we propose and investigate several provenance semantics, based on different approaches for defining classical Datalog semantics. We study the relationship between these semantics, and introduce properties that allow us to analyze and compare them.

Relevant papers: https://proceedings.kr.org/2022/10/ with Camille Bourgaux, Pierre Bourhis and Michaël Thomazo.

14. **Albert Atserias**

Universitat Politècnica de Catalunya, Barcelona, Spain

Title: Consistency of Relations over Monoids

Abstract: The interplay between local consistency and global consistency has been the object of study in several different areas, including probability theory, relational databases, and quantum information. For relational databases, Beeri, Fagin, Maier, and Yannakakis showed that a database schema is acyclic if and only if it has the local-to-global consistency property for relations, which means that every collection of pairwise consistent relations over the schema is globally consistent. More recently, the same result has been shown under bag semantics. In this paper, we carry out a systematic study of local vs. global consistency for relations over positive commutative monoids, which is a common generalization of ordinary relations and bags. Let K be an arbitrary positive commutative monoid. We begin by showing that acyclicity of the schema is a necessary condition for the local-to-global consistency property for K-relations to hold. Unlike the case of ordinary relations and bags, however, we show that acyclicity is not always sufficient. After this, we characterize the positive commutative monoids for which acyclicity is both necessary and sufficient for the local-to-global consistency property to hold; this characterization involves a combinatorial property of monoids, which we call the transportation property. We then identify several different classes of monoids that possess the transportation property. As our final contribution, we introduce a modified notion of local consistency of K-relations, which we call pairwise consistency up to the free cover. We prove that, for all positive commutative monoids K, even those without the transportation property, acyclicity is both necessary and sufficient for every family of K-relations that is pairwise consistent up to the free cover to be globally consistent.

This is joint work with Phokion G. Kolaitis, UC Santa Cruz and IBM Research

Relevant papers:

Albert Atserias and Phokion G. Kolaitis. Consistency of Relations over Monoids. https://arxiv.org/abs/2312.02023

15. Pierre Bourhis

CNRS, CRISTAL, Lille

Title: Circuits for Query evaluation over tree

Abstract: Querying trees via Tree automata presents a lot of interest because several important questions can be executed with a guaranteed efficient time. Over the last decades, different approaches have been presented to solve major query answering questions such as enumeration, probabilistic evaluation... In this survey, we review a particular approach which can be adapted to all these questions: a knowledge compilation approach. We present the different results that can be resolved by this approach and also its limits.

Relevant papers:

- Antoine Amarilli, Pierre Bourhis, Florent Capelli, Mikaël Monet: Ranked Enumeration for MSO on Trees via Knowledge Compilation. CoRR abs/2310.00731 (2023)
- Antoine Amarilli, Pierre Bourhis, Stefan Mengel, Matthias Niewerth: Enumeration on Trees with Tractable Combined Complexity and Efficient Updates. PODS 2019: 89-103
- Antoine Amarilli, Pierre Bourhis, Stefan Mengel: Enumeration on Trees under Relabelings. ICDT 2018: 5:1-5:18
- Antoine Amarilli, Pierre Bourhis, Louis Jachiet, Stefan Mengel: A Circuit-Based Approach to Efficient Enumeration. ICALP 2017: 111:1-111:15
- Antoine Amarilli, Pierre Bourhis, Pierre Senellart: Provenance Circuits for Trees and Treelike Instances. ICALP (2) 2015: 56-68

16. Florent Capelli

Université d'Artois, CRIL, https://florent.capelli.me

Title: From Queries to Circuits

Abstract: In this talk, we will review two algorithms to construct tractable circuits from a conjunctive query and a database whose size can be bounded using fractional hypertree width. The first one is a classical bottom up dynamic programming on a join tree of the conjunctive query, which can be seen as a generalization of Yannakakis approach. The second one is based on a top-down approach akin to exhaustive DPLL, an algorithm originally devised for solving #SAT. We will show that both algorithms construct very

similar circuits on conjunctive queries but that DPLL can be applied to a more general setting without changing much of its structure.

Relevant papers:

- *Direct Access for Conjunctive Queries with Negation*, Florent Capelli, Oliver Irwin, to appear in ICDT24
-

17. Cristian Riveros

Title: MSO enumeration over words and their representations

Abstract: I will present a survey of MSO enumeration problems over words based on the model of annotated automata, a model for encoding MSO queries with output. In the first half, I will present the basic MSO enumeration problem and the representations needed for efficient enumeration. In the second half, I will go through extensions of this MSO enumeration problem, with the required extensions on the representations. Toward the end, I will present some open problems.

Relevant papers:

- Martin Muñoz, Cristian Riveros: Constant-Delay Enumeration for SLP-Compressed Documents. ICDT 2023.
- Martin Muñoz, Cristian Riveros: Streaming Enumeration on Nested Documents. ICDT 2022: 19:1-19:18
- Antoine Amarilli, Pierre Bourhis, Louis Jachiet, Stefan Mengel: A Circuit-Based Approach to Efficient Enumeration. ICALP 2017.

18. Benny Kimelfeld

Title: Answering Database Queries Using Direct-Access Structures

Abstract: The talk will describe recent results on the fine-grained complexity of database queries that involve joins, grouping, aggregation, and ordering. For some common aggregate functions (e.g., min, max, count, sum), such a query can be phrased as an ordinary conjunctive query over a database annotated with a suitable commutative semiring. I will discuss the ability to evaluate such queries by constructing, in quasilinear time in the database size (i.e., roughly the time it takes to read the database), a data structure that provides logarithmic-time direct access to the answers, ordered by a desired lexicographic order. This task is nontrivial since the number of answers might be larger than quasilinear in the database size, so, the data structure needs to provide a representation that is compact, easy to construct, and fast to access. The results provide classifications of queries, orderings, and semirings by the feasibility of such complexity guarantees.

Relevant papers:

- Nofar Carmeli, Nikolaos Tziavelis, Wolfgang Gatterbauer, Benny Kimelfeld, Mirek Riedewald: Tractable Orders for Direct Access to Ranked Answers of Conjunctive Queries. PODS 2021: 325-341
- Idan Eldar, Nofar Carmeli, Benny Kimelfeld: Direct Access for Answers to Conjunctive Queries with Aggregation. CoRR abs/2303.05327 (2023). To appear in ICDT 2024.

19. Nikos Tziavelis

Title: Answering Quantile Join Queries by Representing Inequality Predicates Efficiently

Abstract: This talk will discuss the complexity of answering Quantile Join Queries, which ask for the answer at a specified relative position (e.g., 50% for the median) under some ordering over the answers to an ordinary Join Query (JQ). Compared to the task of direct access, this task is easier since only one access is required. The goal is to avoid materializing the set of all join answers, and to achieve quasilinear time in the size of the database, regardless of the total number of answers. The tractability of such a query does not only depend on the join structure, but also on the desired order. We show an algorithm that covers all known tractable cases by iteratively using a “trimming” subroutine which removes query answers that are higher or lower (according to the ranking function) than a certain answer determined as the “pivot”. Trimming essentially adds inequality predicates to our initial query and an efficient representation of these inequalities implies efficient Quantile Join Query answering for a large class of ranking functions.

Relevant papers:

- Nikolaos Tziavelis, Nofar Carmeli, Wolfgang Gatterbauer, Benny Kimelfeld, and Mirek Riedewald. 2023. Efficient Computation of Quantiles over Joins. PODS 2023: 303–315. https://doi.org/10.1145/3584372.3588670

20. Christoph Berkholz

Title: A dichotomy for succinct representations of homomorphisms

Abstract: The talk is based on the ICALP'23 paper cited below. It will be about factorized databases for multi-way join queries, or, in other words, succinct representations of all homomorphisms between two structures A and B. The main result is a characterisation of (bounded-arity) structures A where this is efficiently doable. In the talk I will mainly focus on lower bounds for factorized representations.

Relevant papers:
Christoph Berkholz, Harry Vinall-Smeeth:

A Dichotomy for Succinct Representations of Homomorphisms. ICALP 2023: 113:1-113:19

21. **Pierre Senellart**

Title: Expected Shapley-Like Scores of Boolean Functions: Complexity and Applications to Probabilistic Databases

Abstract: Shapley values, originating in game theory and increasingly prominent in explainable AI, have been proposed to assess the contribution of facts in query answering over databases, along with other similar power indices such as Banzhaf values. In this work we adapt these Shapley-like scores to probabilistic settings, the objective being to compute their expected value. We show that the computations of expected Shapley values and of the expected values of Boolean functions are interreducible in polynomial time, thus obtaining the same tractability landscape. We investigate the specific tractable case where Boolean functions are represented as deterministic decomposable circuits, designing a polynomial-time algorithm for this setting. We present applications to probabilistic databases through database provenance, and an effective implementation of this algorithm within the ProvSQL system, which experimentally validates its feasibility over a standard benchmark.

Relevant papers:

- Pratik Karmakar, MIkaël Monet, Pierre Senellart, and Stéphane Bressan, 2024. Expected Shapley-Like Scores of Boolean Functions: Complexity and Applications to Probabilistic Databases, https://arxiv.org/abs/2401.06493
- Daniel Deutch, Nave Frost, Benny Kimelfeld, and Mikaël Monet. 2022. Computing the Shapley value of facts in query answering. In SIGMOD Conference. ACM, 1570–1583.
- Pierre Senellart, Louis Jachiet, Silviu Maniu, and Yann Ramusat. 2018. ProvSQL: Provenance and Probability Management in PostgreSQL. Proc. VLDB Endow. 11, 12 (2018), 2034–2037.

22. **Boris Glavic**

University of Illinois, Chicago,

Title: Lessons Learned from Building Systems for Provenance and Explanations

Abstract: In this talk I will introduce to the audience lessons learned from my work and other group's work on building systems for capturing and managing provenance and explanations. For instance, an underappreciated concept in developing such systems is that provenance creates a separate information flow in the system that does not conform to the standard way of how data flows through the operators of a query.

Relevant papers:

- https://vldb.org/pvldb/vol15/p451-niu.pdf
- https://arxiv.org/pdf/1804.07156.pdf

- http://sites.computer.org/debull/A18mar/p51.pdf
- http://www.vldb.org/pvldb/vol13/p912-lee.pdf
- https://dl.acm.org/doi/pdf/10.1145/3555041.3589731
- https://inria.hal.science/hal-01851538/document
- https://dl.acm.org/doi/pdf/10.14778/2824032.2824089

23. **Babak Salimi**
    Title: Training Invariant Machine Learning Models with Incomplete Data

    Abstract: In this talk, I aim to discuss the significant challenge of learning machine learning models that satisfy invariant properties under conditional independence constraints. The importance of this problem will be illustrated through various real-world examples, emphasizing its relevance and urgency. Subsequently, I will analyze existing approaches and their shortcomings, especially in situations where data is compromised by quality issues such as selection bias. To overcome these obstacles, I will introduce a framework inspired by techniques for querying incomplete data in data management. This framework is tailored to effectively handle the specific challenges posed by incomplete datasets. Additionally, I will demonstrate its application in the context of algorithmic fairness.

    Relevant papers: https://arxiv.org/pdf/2212.10839.pdf (VLDB'23)

24. **Amir Gilad (https://amirgilad.github.io/)**
    The Hebrew University, Israel

    Title: Privately generating justifiably fair data

    Abstract: In this talk, I will present our recent work that develops a framework for synthetic data generation that is both differentially-private and fair, where fairness is modeled by an adaptation of the causal definition for justifiable fairness.

    Relevant papers: https://www.vldb.org/pvldb/vol16/p1573-pujol.pdf (VLDB 23')

Ad-hoc short talks:

25. **Bertram Ludäscher (iSchool, dblp)**
    University of Illinois, Urbana-Champaign

    Title: Provenance in Queries, Games, and Argumentation: Time for a Family Reunion!

Abstract: What do queries, games, and argumentation have in common? A lot (and probably more than is commonly acknowledged). Consider the non-stratified, recursive query *Q*(x) :- *E*(x,y), **not** *Q*(y). It has been studied in database theory (to show that stratified Datalog is less expressive than Fixpoint [Kol91]), in LPNMR (to illustrate the 3-valued well-founded semantics [VRS91]), and as an argument processing unit (APU), i.e., a meta-interpreter and logical foundation for abstract argumentation [Dun95].
The *n*-ary version of *Q* can be understood as a normal form for Fixpoint queries using the well-founded semantics [VRS91], even when restricted to total (i.e., 2-valued) programs [FKL97]. In a delightfully elegant characterization, solved games (well-founded models of *Q*) represent their own provenance [KLZ13].
The underlying graph- and game-theoretic notions [Flu00] lend themselves to new ways of looking at provenance for queries [KLZ13] and argumentation frameworks [LBX23].

References:

- [Kol91] P. Kolaitis, The expressive power of stratified logic programs, *Information and Computation* (1991).
- [Dun95] P. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games, *AI* (1995)
- [FKL97] J. Flum, M. Kubierschky, and B. Ludäscher. Total and partial well-founded Datalog coincide. *ICDT*, Delphi. LNCS 1186 (1997)
- [Flu00] J. Flum, Games, Kernels, and Antitone Operations, *Order* 17 (2000).
- [KLS12] Köhler, S., Ludäscher, B., & Smaragdakis, Y. Declarative Datalog Debugging for Mere Mortals. *Datalog in Academia and Industry* (2012)
- [KLZ13] S. Köhler, B. Ludäscher, and D. Zinn. First-Order Provenance Games. *In Search of Elegance in the Theory and Practice of Computation* (Peter Buneman Festschrift). LNCS 8000 (2013) [small updates on arxiv]
- [LBX23] B. Ludäscher, S. Bowers, and Y. Xia. Games, Queries, and Argumentation Frameworks: Towards a Family Reunion. *AI³@AI*IA* (2023)

26. Christoph Standke

RWTH Aachen University, Germany

Title: The Importance of Parameters in Database Queries

Abstract: In this talk, I will introduce a framework for quantifying the importance of the choices of parameter values to the result of a query over a database. In our framework, the importance of a parameter is its SHAP score and we make the case for the rationale of using this score by showing that we arrive at this score in two different, apparently opposing, approaches to quantifying the contribution of a parameter. We then point out that this framework yields an interesting complexity-theoretic landscape.

Based on joint work with Martin Grohe, Benny Kimelfeld and Peter Lindner.

Relevant Papers: https://arxiv.org/abs/2401.04606 (ICDT 2024)