

Logic and Random Discrete Structures

Erich Grädel^{*1}, Phokion G. Kolaitis^{*2}, Marc Noy^{*3}, and Matthias Naaf^{†4}

1 RWTH Aachen, DE. graedel@logic.rwth-aachen.de

2 University of California Santa Cruz and IBM Research, US. kolaitis@ucsc.edu

3 UPC Barcelona Tech, ES. marc.noy@upc.edu

4 RWTH Aachen, DE. naaf@logic.rwth-aachen.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 22061 “Logic and Random Discrete Structures”. The main topic of this seminar has been the analysis of large random discrete structures, such as trees, graphs, or permutations, from the perspective of mathematical logic. It has brought together both experts and junior researchers from a number of different areas where logic and random structures play a role, with the goal to establish new connections between such areas and to encourage interactions between foundational research and different application areas, including probabilistic databases.

Seminar February 6–11, 2022 – <http://www.dagstuhl.de/22061>

2012 ACM Subject Classification Theory of computation → Logic; Theory of computation → Computational complexity and cryptography; Mathematics of computing → Discrete mathematics

Keywords and phrases combinatorics, complexity theory, logic, random structures, probabilistic databases

Digital Object Identifier 10.4230/DagRep.12.2.1

1 Executive Summary

Erich Grädel

Phokion G. Kolaitis

Marc Noy

License  Creative Commons BY 4.0 International license
© Erich Grädel, Phokion G. Kolaitis, and Marc Noy

Topic and Goals of the Seminar

The analysis of large random discrete structures, such as trees, graphs, or permutations, is a focus of research in contemporary discrete mathematics. Logic provides a useful and powerful formalism for expressing and classifying discrete structures; moreover, it is intimately linked to the study of algorithms, computational complexity, and structural graph theory. Over the past several decades, researchers have studied random discrete structures from a logical perspective. The first significant result in this direction was the zero-one law for first-order logic under the uniform measure; this seminal result, was followed by the discovery of ‘logical limit laws’ or ‘convergence laws’ for several different models of random discrete structures and for various logics of significance in computer science. In more recent years, a renewed impetus has emerged for research activity on random discrete structures from a logical

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Logic and Random Discrete Structures, *Dagstuhl Reports*, Vol. 12, Issue 2, pp. 1–16

Editors: Erich Grädel, Phokion G. Kolaitis, Marc Noy, and Matthias Naaf



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

perspective. This is in part due to the availability of new methods and techniques, including asymptotic enumeration, discrete harmonic analysis, an extension of Gowers norms, and limit structures. Exciting new results on random geometric graphs, graphs on surfaces, classes of sparse graphs, graph limits, and flag algebras have been established. On the computer science side, there has been a systematic exploration of probabilistic databases, which has brought together databases, logic, and random structures. The main aim of this seminar has been to bring together some of the foremost experts from these different fields, as well as junior researchers who may become motivated to work deeper in the frontier of logic and random structures. In addition to making tangible progress on some of the currently outstanding open problems in this area, we wanted to establish new connections between (classical) random discrete structures, flag algebras, and sparse graph limits, both in terms of identifying new research questions and embarking on new collaborations, as well as fruitful interaction between foundational research and different application areas, including probabilistic databases.

Organisation and Activities

Despite the restrictions and problems caused by corona pandemic, the seminar had originally been intended as a non-hybrid event with all participants on site. At the end, however, this turned out to be infeasible; as a result, two of the invited survey talks and a number of the contributed talks had to be given remotely via Zoom.

The organisers created a schedule consisting of four invited one-hour survey talks, and more focussed regular contributions proposed by the participants. The survey talks were given by

- Albert Atserias on certifying the chromatic number of a random graph;
- Fiona Skerman on the inference of underlying community structures in partially observed graphs;
- Dan Suciú on probabilistic databases;
- Patrice Ossona de Mendez on limits of graphs.

The talks of Fiona Skerman and Patrice Ossona de Mendez were given over Zoom.

In addition, there were 18 contributed talks, 11 of which were given on site, and 7 remotely via Zoom.

Overall, the organisers regard the seminar to have been a very successful scientific event. There was a general view shared by all participants that the community working on logic and random structures is in excellent shape, with interesting new developments and exciting results in many different directions. The participants clearly expressed the wish to have a future meeting of this community, be in in Dagstuhl or elsewhere, within the next two to three years.

The organisers are grateful to the Scientific Directorate and to the staff of the Center for their support of this seminar.

2 Table of Contents

Executive Summary

| | |
|--|---|
| <i>Erich Grädel, Phokion G. Kolaitis, and Marc Noy</i> | 1 |
|--|---|

Overview of Talks

| | |
|---|----|
| On certifying the chromatic number of a random graph <i>Albert Atserias</i> | 5 |
| The almost-sure theories of classes defined by forbidden homomorphisms <i>Manuel Bodirsky</i> | 5 |
| The modal logic of almost sure frame validities in the finite <i>Valentin Goranko</i> | 6 |
| Limit laws for existential monadic second order logic <i>Maksim Zhukovskii</i> | 6 |
| Improved bounds for acyclic coloring parameters <i>Lefteris M. Kirousis</i> | 7 |
| Quasirandom combinatorial structures <i>Daniel Král'</i> | 7 |
| Partially observing graphs – when can we infer underlying community structure? <i>Fiona Skerman</i> | 8 |
| Heavy-tailed distributions for random SAT <i>Andrei A. Bulatov</i> | 8 |
| Convergence law for random permutations <i>Valentin Féray</i> | 9 |
| Logic and property testing on graphs of bounded degree <i>Isolde Adler</i> | 9 |
| Limiting probabilities of first order properties of random sparse graphs <i>Marc Noy</i> | 10 |
| Logical limit laws for Mallows random permutations <i>Tobias Müller</i> | 10 |
| On first order model checking on Erdős-Rényi graphs <i>Peter Rossmanith</i> | 10 |
| Probabilistic databases – overview <i>Dan Suciu</i> | 11 |
| Explaining query answers through probabilistic databases <i>Benny Kimelfeld</i> | 11 |
| Probabilistic query evaluation with bag semantics <i>Peter Lindner</i> | 12 |
| Limits of graphs <i>Patrice Ossona de Mendez</i> | 12 |
| Zero-one laws and almost sure valuations of first-order logic in semiring semantics <i>Matthias Naaf</i> | 13 |

| | |
|--|----|
| Moser-Tardos algorithm with small number of random bits <i>Oleg Pikhurko</i> | 13 |
| The repeated insertion model (RIM): probabilistic inference and applications <i>Batya Kenig</i> | 14 |
| Repairs, measures, and complexity for constraints violations in databases <i>Sudeepa Roy</i> | 14 |
| Ordered graphs of bounded twin-width <i>Szymon Torunczyk</i> | 15 |
| Participants | 16 |
| Remote Participants | 16 |

3 Overview of Talks

The talks are listed in the order of presentation during the seminar.

3.1 On certifying the chromatic number of a random graph

Albert Atserias (UPC Barcelona Tech, ES)

License  Creative Commons BY 4.0 International license
 © Albert Atserias

A standard first-moment calculation shows that, for every fixed integer $q > 1$, the chromatic number $\chi(G)$ of an Erdős-Rényi random graph G of any sufficiently large constant density $d > c(q)$ of edges is, asymptotically almost surely, larger than q . We study the question whether there exist efficiently checkable certificates for the statement “ $\chi(G) > q$ ” that apply to such a random graph with high probability. First we overview the known negative results that show that bounded local consistency methods do not suffice. Then we ask whether semi-definite programming methods suffice and provide some new observations indicating that they do. This is in sharp contrast to the case of random k -CNF formulas where it is known that not even semi-definite programming methods suffice to certify their almost sure unsatisfiability at any constant density of clauses to variables.

3.2 The almost-sure theories of classes defined by forbidden homomorphisms

Manuel Bodirsky (TU Dresden, DE)

License  Creative Commons BY 4.0 International license
 © Manuel Bodirsky

Joint work of Manuel Bodirsky, Colin Jahel

Main reference Manuel Bodirsky, Colin Jahel: “Asymptotic Theories of Classes Defined by Forbidden Homomorphisms”, CoRR, Vol. abs.2204.01404, 2022.

URL <https://arxiv.org/abs/2204.01404>

This talk is about the almost-sure theories for classes of finite structures that are specified by homomorphically forbidding a finite set \mathcal{F} of finite structures. If \mathcal{F} consists of undirected graphs, a full description of these theories can be derived from the Kolaitis-Proemel-Rothschild theorem, which treats the special case where $\mathcal{F} = K_n$. The corresponding question for finite sets \mathcal{F} of directed graphs is wide open. We present a description of the almost-sure theories of classes described by homomorphically forbidding finite sets \mathcal{F} of oriented trees.

3.3 The modal logic of almost sure frame validities in the finite

Valentin Goranko (University of Stockholm, SE)

License © Creative Commons BY 4.0 International license
© Valentin Goranko

Main reference Valentin Goranko: “The Modal Logic of Almost Sure Frame Validities in the Finite”, in Proc. of the 13th Conference on Advances in Modal Logic, AiML 2020, Helsinki, Finland, August 24-28, 2020, pp. 249–268, College Publications, 2020.

URL <http://www.aiml.net/volumes/volume13/Goranko.pdf>

A modal formula is almost surely frame-valid in the finite if the probability that it is valid in a randomly chosen finite frame with n states is asymptotically 1 as n grows unboundedly. In this talk I discuss the normal modal logic ML^{as} of all modal formulae that are almost surely frame-valid in the finite. Because of the failure of the zero-one law for frame validity in modal logic, the logic ML^{as} extends properly the modal logic of the countable random frame ML^f , which was completely axiomatized in a 2003 paper by Goranko and Kapron. Thus, the logic ML^{as} is still relatively little known and understood.

In this talk I first introduce the logic ML^{as} , present what is known about it, including the so far known axioms coming from ML^f and a certain model-theoretic characterisation of its additional validities beyond those in ML^f . I then raise some open problems and conjectures regarding the missing additional axioms over ML^f and the explicit description of the complete axiomatisation of ML^{as} , which may turn out to hinge on difficult combinatorial-probabilistic arguments and calculations.

3.4 Limit laws for existential monadic second order logic

Maksim Zhukovskii (MIPT – Dolgoprudny, RU)

License © Creative Commons BY 4.0 International license
© Maksim Zhukovskii

The classical result of Glebskii, Kogan, Liogon’kii, Talanov (1969) and of Fagin (1976) states that the dense (the probability of appearance of an edge p is constant) binomial random graph $G(n, p)$ obeys first order 0-1 law. On the other hand, Kaufmann and Shelah (1985) proved that there exists a monadic second order (MSO) sentence such that its truth probability on $G(n, p)$ does not converge. In 2001, Le Bars proved the same non-convergence result for the existential fragment of MSO. In the talk, I am going to review limit laws for existential monadic second order (EMSO) properties of binomial random graphs both in dense and sparse regimes. In particular, I will state our recent result that disproves the conjecture of Le Bars: there exists an EMSO sentence with 2 first order variables such that its truth probability does not converge.

3.5 Improved bounds for acyclic coloring parameters

Lefteris M. Kirousis (University of Athens, GR)

License © Creative Commons BY 4.0 International license
© Lefteris M. Kirousis

Joint work of Lefteris M. Kirousis, John Livieratos

The *acyclic chromatic number* of a graph is the least number of colors needed to properly color its vertices so that none of its cycles has only two colors. The *acyclic chromatic index* is the analogous graph parameter for edge colorings. We first show that the acyclic chromatic index is at most $2\Delta - 1$, where Δ is the maximum degree of the graph. We then show that for all $\epsilon > 0$ and for Δ large enough (depending on ϵ), the acyclic chromatic number of the graph is at most $\lceil (2^{-1/3} + \epsilon)\Delta^{4/3} \rceil + \Delta + 1$. Both results improve long chains of previous successive advances. Both are algorithmic, in the sense that the colorings are generated by randomized algorithms. However, in contrast with extant approaches, where the randomized algorithms assume the availability of enough colors to guarantee properness deterministically, and use additional colors for randomization in dealing with the bichromatic cycles, our algorithms may initially generate colorings that are not necessarily proper; they only aim at avoiding cycles where all pairs of edges, or vertices, that are one edge, or vertex, apart in a traversal of the cycle are homochromatic (of the same color). When this goal is reached, they check for properness and if necessary they repeat until properness is attained.

3.6 Quasirandom combinatorial structures

Daniel Král' (Masaryk University – Brno, CZ)

License © Creative Commons BY 4.0 International license
© Daniel Král'

Joint work of Timothy Chan, Jacob Cooper, Robert Hancock, Adam Kabelá, Daniel Král', Ander Lamaison, Taísa Martins, Samuel Mohr, Jonathan Noel, Roberto Parente, Yanitsa Pehova, Oleg Pikhurko, Maryam Sharifzadeh, Fiona Skerman, Jan Volec

A combinatorial structure is said to be quasirandom if it resembles a random structure in a certain robust sense. The notion of quasirandom graphs, developed in the work of Rödl, Thomason, Chung, Graham and Wilson in 1980s, is particularly robust as several different properties of truly random graphs, e.g., subgraph density, uniform edge distribution and spectral properties, are satisfied by a large graph if and only if one of them is.

We will discuss how the classical results on quasirandom graphs can be viewed through the lenses of the theory of combinatorial limits and apply the tools offered by this theory to study quasirandomness of tournaments, permutations and Latin squares. We show that the same phenomenon as in the case of graphs, when quasirandomness is captured by the densities of finitely many substructures (in the case of graphs, an edge and any even cycle), appears in relation to these structures, too. We also discuss what minimal sets of substructures have this property and give characterizations of such sets in several of the considered settings.

References

- 1 Timothy Chan, Daniel Král', Jonathan Noel, Yanitsa Pehova, Maryam Sharifzadeh, and Jan Volec. Characterization of quasirandom permutations by a pattern sum. *Random Structures & Algorithms*, 57(4):920–939, 2020. doi:10.1002/rsa.20956.
- 2 Jacob Cooper, Daniel Král', Ander Lamaison, and Samuel Mohr. Quasirandom latin squares. *Random Structures & Algorithms*, 2020, doi:https://doi.org/10.1002/rsa.21060.

- 3 Robert Hancock, Adam Kabela, Daniel Král', Taísa Martins, Roberto Parente, Fiona Skerman, and Jan Volec. No additional tournaments are quasirandom-forcing. [arXiv:1912.04243](https://arxiv.org/abs/1912.04243), 2019.
- 4 Daniel Král' and Oleg Pikhurko. Quasirandom permutations are characterized by 4-point densities. *Geometric and Functional Analysis*, 23(2):570–579, 2013. doi:10.1007/s00039-013-0216-9.

3.7 Partially observing graphs – when can we infer underlying community structure?

Fiona Skerman (Uppsala University, SE)

License © Creative Commons BY 4.0 International license
© Fiona Skerman

Joint work of Colin McDiarmid, Fiona Skerman.

Main reference Colin McDiarmid, Fiona Skerman: “Modularity and edge sampling”, CoRR, Vol. abs/2112.13190, 2021.

URL <https://arxiv.org/abs/2112.13190>

Suppose edges in an underlying graph G appear independently with some probability in our observed graph G' – or alternately that we can query uniformly random edges. We describe how high a sampling probability we need to infer the modularity of the underlying graph.

Modularity is a function on graphs which is used in algorithms for community detection. For a given graph G , each partition of the vertices has a modularity score, with higher values indicating that the partition better captures community structure in G . The (max) modularity $q^*(G)$ of the graph G is defined to be the maximum over all vertex partitions of the modularity score, and satisfies $0 \leq q^*(G) \leq 1$.

In the seminar I will spend time on intuition for the behaviour of modularity, how it can be approximated, links to other graph parameters and to present some open problems.

3.8 Heavy-tailed distributions for random SAT

Andrei A. Bulatov (Simon Fraser University – Burnaby, CA)

License © Creative Commons BY 4.0 International license
© Andrei A. Bulatov

Joint work of Andrei A. Bulatov, Oleksii Omelchenko

While the focus of research on Random SAT have been on uniformly distributed instances, several models were suggested, in which the distribution from which instances are sampled is far from uniform. Sometimes such models are motivated by analyzing ‘practical’ distributions, sometimes they appear to be mathematically natural and sound. We study one such model, the Configuration Model, which is parametrized by a distribution of degrees of variables. To sample from this model we start with a set of variables, then for each of them we sample a degree and create the prescribed number of copies. Each copy (or clone) is then negated with probability $1/2$, and all the clones are grouped into clauses uniformly at random. We study properties of Random SAT problems generated this way for a wide range of degree distributions.

References

- 1 Oleksii Omelchenko, Andrei A. Bulatov. Satisfiability and algorithms for non-uniform random k -SAT. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5):3886-3894, 2021. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/16507>.
- 2 Oleksii Omelchenko, Andrei A. Bulatov. Satisfiability threshold for power law random 2-SAT in configuration model. *Theoretical Computer Science*, 888:70-94, 2021. doi: 10.1016/j.tcs.2021.07.028.
- 3 Oleksii Omelchenko, Andrei A. Bulatov. Concentration inequalities for sums of random variables, each having power bounded tails. [arXiv:1903.02529](https://arxiv.org/abs/1903.02529), 2019.

3.9 Convergence law for random permutations

Valentin Féray (CNRS – Vandoeuvre-lès-Nancy, FR)

License  Creative Commons BY 4.0 International license
© Valentin Féray

Joint work of Michael H. Albert, Mathilde Bouvel, Valentin Féray, Marc Noy

There are two natural ways to see permutations as models of some logical theory: either as bijections from a set A to itself or as a pair of linear orders on a set A . In this talk we discuss the question of finding convergence laws for models of random permutations. In particular, we prove a convergence law for uniform 231-avoiding permutations, seen as pairs of orders.

3.10 Logic and property testing on graphs of bounded degree

Isolde Adler (University of Leeds, GB)

License  Creative Commons BY 4.0 International license
© Isolde Adler

Joint work of Isolde Adler, Polly Fahey, Frederik Harwath, Noleen Köhler, Pan Peng

Property testing (for a property P) asks for a given graph, whether it has property P , or is “structurally far” from having that property. A “testing algorithm” is a probabilistic algorithm that answers this question with high probability correctly, by only looking at small parts of the input. Testing algorithms are thought of as “extremely efficient”, making them relevant in the context of large data sets.

In this talk I will present recent positive and negative results about testability of properties definable in first-order logic and monadic second-order logic on classes of bounded-degree graphs.

3.11 Limiting probabilities of first order properties of random sparse graphs

Marc Noy (UPC Barcelona Tech, ES)

License  Creative Commons BY 4.0 International license
© Marc Noy

Joint work of Alberto Larrauri, Tobias Müller, Marc Noy

Main reference Lázaro Alberto Larrauri, Tobias Müller, Marc Noy: “Limiting probabilities of first order properties of random sparse graphs and hypergraphs”, *Random Struct. Algorithms*, Vol. 60(3), pp. 506–526, 2022.

URL <https://doi.org/10.1002/rsa.21041>

Let G_n be the binomial random graph $G(n, p = c/n)$ in the sparse regime, which as is well-known undergoes a phase transition at $c = 1$. Lynch [1] showed that for every first order sentence ϕ , the limiting probability that G_n satisfies ϕ as $n \rightarrow \infty$ exists, and moreover it is an analytic function of c . In this paper we consider the closure $\overline{L_c}$ in the interval $[0, 1]$ of the set L_c of all limiting probabilities of first order sentences in G_n . We show that there exists a critical value $c_0 \approx 0.93$ such that $\overline{L_c} = [0, 1]$ when $c \geq c_0$, whereas $\overline{L_c}$ misses at least one subinterval when $c < c_0$.

References

- 1 James F. Lynch. Probabilities of sentences about very sparse random graphs. *Random Structures & Algorithms*, 3(1):33–53, 1992. doi:10.1002/rsa.3240030105.

3.12 Logical limit laws for Mallows random permutations

Tobias Müller (University of Groningen, NL)

License  Creative Commons BY 4.0 International license
© Tobias Müller

Joint work of Tobias Müller, Fiona Skerman, Teun Verstraaten

The Mallows distribution samples a permutation on $1, \dots, n$ at random, where each permutation has probability proportional to $q^{\text{inv}(\pi)}$, where $q > 0$ is a parameter and $\text{inv}(\cdot)$ denotes the number of inversions of π . So in particular, setting $q = 1$ we retrieve the uniform distribution.

In the talk, I discussed some preliminary results on logical limit laws for Mallows random permutations, using two different logical languages of permutations, called “theory of one bijection” (TOOB) and “theory of two orders” (TOTO) by Albert, Bouvel and Feray.

3.13 On first order model checking on Erdős-Rényi graphs

Peter Rossmanith (RWTH Aachen, DE)

License  Creative Commons BY 4.0 International license
© Peter Rossmanith

Clique can be solved in expected FPT time on uniformly distributed graphs of size n while this is not clear for Dominating Set. We show that it is indeed unlikely that Dominating Set can be solved efficiently on random graphs: If yes, then every first-order expressible graph property can be solved in expected FPT time, too. Furthermore, this remains true when we consider random graphs with an arbitrary constant edge probability. There is a very simple problem on random matrices that is equally hard to solve on average: Given a square boolean matrix, are there k rows whose logical AND is the zero vector?

3.14 Probabilistic databases – overview

Dan Suciu (University of Washington – Seattle, US)

License © Creative Commons BY 4.0 International license
© Dan Suciu

Main reference Dan Suciu: “Probabilistic Databases for All”, in Proc. of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2020, Portland, OR, USA, June 14-19, 2020, pp. 19–31, ACM, 2020.

URL <https://doi.org/10.1145/3375395.3389129>

In probabilistic databases the data is uncertain and is modeled by a probability distribution. The central problem in probabilistic databases is query evaluation, which requires performing not only traditional data processing such as joins, projections, unions, but also probabilistic inference in order to compute the probability of each item in the answer. At their core, probabilistic databases are a proposal to integrate logic with probability theory. This talk gives an overview of the probabilistic data models and the main results on query evaluation on probabilistic databases.

3.15 Explaining query answers through probabilistic databases

Benny Kimelfeld (Technion – Haifa, IL)

License © Creative Commons BY 4.0 International license
© Benny Kimelfeld

Joint work of Antoine Amarilli, Leopoldo Bertossi, Daniel Deutch, Nave Frost, Benny Kimelfeld, Ester Livshits, Mikaël Monet

Main reference Daniel Deutch, Nave Frost, Benny Kimelfeld, Mikaël Monet: “Computing the Shapley Value of Facts in Query Answering”, CoRR, Vol. abs/2112.08874, 2021.

URL <https://arxiv.org/abs/2112.08874>

One of the challenges in explanations for data-analysis tools is the quantification of the responsibility of individual data items to the overall result. Nowadays a common approach is to deploy concepts from the theory of cooperative games. A primary example is the Shapley value – a conventional and well-studied function for determining the contribution of a player to the coalition. I will describe our recent research on the computation of the Shapley value (and values of a similar nature) of a database tuple as its contribution to the result of a query. It turns out that the complexity of this task has tight connections to the complexity of query evaluation over tuple-independent probabilistic databases. Moreover, this connection highlights the need to understand the complexity of query evaluation in cases where probability assignments are restricted, such as the uniform case where each tuple has the probability 0.5 (and the task is to count the database subsets that satisfy the query). I will present a recent resolution of the uniform case, as well as every case where the probability is fixed in every relation, for an important class of database queries.

3.16 Probabilistic query evaluation with bag semantics

Peter Lindner (EPFL Lausanne, CH)

License © Creative Commons BY 4.0 International license
© Peter Lindner

Joint work of Martin Grohe, Peter Lindner, Christoph Standke

Main reference Martin Grohe, Peter Lindner, Christoph Standke: “Probabilistic Query Evaluation with Bag Semantics”, CoRR, Vol. abs/2201.11524, 2022.

URL <https://arxiv.org/abs/2201.11524>

Probabilistic databases (PDBs) are usually introduced as probability spaces over the subsets of a finite set of possible facts. Probabilistic query evaluation (PQE) is the problem of finding the probability that a Boolean query evaluates to true on a given PDB. The data complexity of PQE is well-understood for the class of unions of conjunctive queries on tuple-independent PDBs: for every query, the problem can be classified to be either solvable in polynomial time, or is #P-hard [1].

In this talk, we discuss a version of PQE, where both the query evaluation, and the probabilistic databases use a bag semantics. That is, the input PDB may contain duplicates, and the query evaluation takes tuple multiplicities into account. Our main focus lies on self-join free conjunctive queries where we obtain a dichotomy similar to the one above. This is achieved by combining known results for the set version of the problem with novel techniques that are required to handle distributions over multiplicities.

References

- 1 Nilesch Dalvi and Dan Suciu. The dichotomy of probabilistic inference for unions of conjunctive queries. *Journal of the ACM*, 59(6):30:1–30:87, 2012. doi:10.1145/2395116.2395119.

3.17 Limits of graphs

Patrice Ossona de Mendez (EHESS – Paris, FR)

License © Creative Commons BY 4.0 International license
© Patrice Ossona de Mendez

Joint work of Jaroslav Nešetřil, Patrice Ossona de Mendez

How to represent limits of networks? In this survey, several kinds of limits are considered: elementary limits, left limits, local limits, and structural limits, as well as different types of limits objects (distributional or analytical). A particular attention is given to the notion of structural limits, based on the convergence of the satisfaction probabilities of first-order formulas, which generalize classical notions and offers a both a distributional limit objects (as a probability distribution on a Stone space) and an analytic limit object (a modeling, which is a totally Borel structure on a standard probability space).

3.18 Zero-one laws and almost sure valuations of first-order logic in semiring semantics

Matthias Naaf (RWTH Aachen, DE)

License © Creative Commons BY 4.0 International license
© Matthias Naaf

Joint work of Erich Grädel, Hayyan Helal, Matthias Naaf, Richard Wilke

Main reference Erich Grädel, Hayyan Helal, Matthias Naaf, Richard Wilke: “Zero-One Laws and Almost Sure Valuations of First-Order Logic in Semiring Semantics”, CoRR, Vol. abs/2203.03425, 2022.

URL <https://doi.org/10.48550/arXiv.2203.03425>

Semiring semantics evaluates logical statements by values in some commutative semiring K , and random semiring interpretations, induced by a probability distribution on K , generalise random structures. In this talk, we investigate how the classical 0-1 laws of Glebskii et al. and Fagin generalise to semiring semantics.

Using algebraic representations of FO-formulae, we show that for many semirings, including min-max-semirings and the natural semiring, 0-1 laws hold under reasonable assumptions on the probability distribution. We can further partition the FO-sentences into classes $(\Phi_j)_{j \in K}$ for all semiring elements j , such that all sentences in Φ_j almost surely evaluate to j in random semiring interpretations. For finite min-max and lattice semirings, this partition actually collapses to just three classes Φ_0 , Φ_1 and Φ_ε of sentences that, respectively, almost surely evaluate to 0, to the greatest semiring value 1, and to the smallest non-zero value ε . Computing this almost sure valuation is a PSPACE-complete problem.

3.19 Moser-Tardos algorithm with small number of random bits

Oleg Pikhurko (University of Warwick – Coventry, GB)

License © Creative Commons BY 4.0 International license
© Oleg Pikhurko

Joint work of Endre Csóka, Lukasz Grabowski, András Máthé, Oleg Pikhurko, Konstantinos Tyros

Main reference Endre Csóka, Lukasz Grabowski, András Máthé, Oleg Pikhurko, Konstantinos Tyros: “Moser-Tardos Algorithm with small number of random bits”, CoRR, Vol. abs/2203.05888, 2022.

URL <https://doi.org/10.48550/arXiv.2203.05888>

We present a variant of the parallel Moser-Tardos Algorithm [1] and show that, for a class of problems whose dependency graphs have some subexponential growth, the expected number of random bits used by the algorithm is constant. In particular the expected number of used random bits is independent from the total number of variables. This is achieved by using the same random bits to resample variables which are far enough in the dependency graph.

References

- 1 Robin A. Moser and Gábor Tardos. A constructive proof of the general Lovász local lemma. *Journal of the ACM*, 57(2):11:1–11:15, 2010. doi:10.1145/1667053.1667060.

3.20 The repeated insertion model (RIM): probabilistic inference and applications

Batya Kenig (Technion – Haifa, IL)

License © Creative Commons BY 4.0 International license

© Batya Kenig

Joint work of Lovro Ilijasić, Batya Kenig, Benny Kimelfeld, Haoyue Ping, Julia Stoyanovich

Main reference Batya Kenig, Lovro Ilijasic, Haoyue Ping, Benny Kimelfeld, Julia Stoyanovich: “Probabilistic Inference Over Repeated Insertion Models”, in Proc. of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pp. 1897–1904, AAAI Press, 2018.

URL <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16252>

Distributions over rankings are used to model user preferences in elections, commerce, and more. The Repeated Insertion Model (RIM) gives rise to various known probability distributions over rankings, in particular to the popular Mallows model. However, probabilistic inference over RIM is provably intractable in the general case. In this talk, I will describe an algorithm for computing the marginal probability of an arbitrary partially ordered set over RIM. The complexity of the algorithm is captured in terms of a new measure termed the “cover width”. I will briefly discuss an application of RIM for the task of estimating the probability of an outcome in an election over probabilistic votes.

References

- 1 Batya Kenig and Benny Kimelfeld. Approximate inference of outcomes in probabilistic elections. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):2061–2068, 2019. doi:10.1609/aaai.v33i01.33012061.

3.21 Repairs, measures, and complexity for constraints violations in databases

Sudeepa Roy (Duke University – Durham, US)

License © Creative Commons BY 4.0 International license

© Sudeepa Roy

Joint work of Benny Kimelfeld, Ester Livshits, Ihab Ilyas, Rina Kochirgan, Sudeepa Roy, Segev Tsur

Noisy or inconsistent databases that violate one or more integrity constraints expected to hold on the database are abundant in practice. First, I will discuss the complexity of computing an optimal “repair” for an inconsistent database where integrity constraints are Functional Dependencies, both in terms of “subset repairs” (by a minimum number of tuple deletion) and “update repairs” (by a minimum number of value or cell updates), and draw a connection to the complexity of the “most probable database” problem. Then I will discuss theoretical properties of “measures” of inconsistencies in a database where integrity constraints are violated, which can be useful for reliability estimation for datasets and progress indication in data repair.

3.22 Ordered graphs of bounded twin-width

Szymon Torunczyk (*University of Warsaw, PL*)

License  Creative Commons BY 4.0 International license
© Szymon Torunczyk

We establish a list of characterizations of bounded twin-width for hereditary classes of totally ordered graphs: as classes of at most exponential growth studied in enumerative combinatorics, as NIP classes studied in model theory, as classes that do not transduce the class of all graphs studied in finite model theory, and as classes for which model checking first-order logic is fixed-parameter tractable studied in algorithmic graph theory.

This has several consequences. First, it allows us to show that every hereditary class of ordered graphs either has at most exponential growth, or has at least factorial growth. This settles a question first asked by Balogh, Bollobás, and Morris [1] on the growth of hereditary classes of ordered graphs, generalizing the Stanley-Wilf conjecture/Marcus-Tardos theorem. Second, it gives a fixed-parameter approximation algorithm for twin-width on ordered graphs. Third, it yields a full classification of fixed-parameter tractable first-order model checking on hereditary classes of ordered binary structures. Finally, it provides a model-theoretic characterization of classes with bounded twin-width.

References

- 1 József Balogh, Béla Bollobás, and Robert Morris. Hereditary properties of partitions, ordered graphs and ordered hypergraphs. *European Journal of Combinatorics*, 27(8):1263–1281, 2006. doi:10.1016/j.ejc.2006.05.004.

Participants

- Aida Abiad
TU Eindhoven, NL
- Isolde Adler
University of Leeds, GB
- Albert Atserias
UPC Barcelona Tech, ES
- Manuel Bodirsky
TU Dresden, DE
- Andrei A. Bulatov
Simon Fraser University –
Burnaby, CA
- Anuj Dawar
University of Cambridge, GB
- Valentin Féray
CNRS – Vandoeuvre-lès-
Nancy, FR
- Erich Grädel
RWTH Aachen, DE
- Benny Kimelfeld
Technion – Haifa, IL
- Peter Lindner
EPFL Lausanne, CH
- Tobias Müller
University of Groningen, NL
- Matthias Naaf
RWTH Aachen, DE
- Marc Noy
UPC Barcelona Tech, ES
- Oleg Pikhurko
University of Warwick –
Coventry, GB
- Peter Rossmanith
RWTH Aachen, DE
- Dan Suciu
University of Washington –
Seattle, US
- Lidia Tendra
University of Opole, PL
- Teun Verstraaten
University of Groningen, NL



Remote Participants

- Andreas R. Blass
University of Michigan –
Ann Arbor, US
- Valentin Goranko
University of Stockholm, SE
- Batya Kenig
Technion – Haifa, IL
- Lefteris M. Kirousis
University of Athens, GR
- Phokion G. Kolaitis
University of California –
Santa Cruz, US
- Daniel Král’
Masaryk University – Brno, CZ
- Jerzy Marcinkowski
University of Wrocław, PL
- Patrice Ossona de Mendez
EHESS – Paris, FR
- Sudeepa Roy
Duke University – Durham, US
- Fiona Skerman
Uppsala University, SE
- Christoph Standke
RWTH Aachen, DE
- Szymon Torunczyk
University of Warsaw, PL
- Maksim Zhukovskii
MIPT – Dolgoprudny, RU