

Fundamental Problems in Statistical Relational AI

Sagar Malhotra

TU Wien, Austria

Presenters: Sagar Malhotra

Email: sagar.malhotra@tuwien.ac.at

Proposed Duration: Half-Day Tutorial

1 Introduction

Relational data is characterized by the rich structure it encodes in the dependencies between the individual entities of a given domain. Statistical Relational AI (StarAI) combines first-order logic and probability to learn and reason over relational domains by creating parametric probability distributions over relational structures [1,2]. StarAI models can succinctly represent the complex dependencies in relational data and admit learning and inference under uncertainty. Such expressivity allows for StarAI models to be used in various applications that simultaneously require knowledge representation, learning, reasoning and uncertainty quantification. These models have been extensively used in domains like social network analysis [3], synthesizing biological networks [4] and for various problems on knowledge graphs [5,6]. Furthermore, many widely used Neuro-Symbolic approaches for integrating neural networks with symbolic methods have been obtained as neural extensions of StarAI models [7,8,9,10]. A key feature of StarAI models is the inherent interpretability and explainability. These features have lead to significant interest in applying StarAI models to safety critical areas like Healthcare [11,12] and for analyzing complex social [13,14] and biological systems [15,16].

The need for using StarAI models at scale is consistently rising with the ever-increasing quantities of relational data. However, StarAI models are significantly limited when it comes to the **tractability** of learning and inference. This limitation emerges from the intractability of Weighted First Order Model Counting (WFOMC) [17], as both learning and inference in many StarAI models can be reduced to instances of WFOMC. Furthermore, fundamental properties expected of sound statistical models, like **consistency** of parameter estimation, do not hold for StarAI models.

In this tutorial, we will focus on both tractability and consistency of StarAI models from a foundational perspective. The tutorial will be divided into two focus-topic. First focus-topic will be on tractable WFOMC, where we will discuss the recent developments in the fragments of first order logic that admit tractable WFOMC. In the second focus-topic, we will discuss consistency of parameter estimation and generalization behavior of StarAI models across different domain sizes. The learning goal of the tutorial would be to convey state-of-the-art knowledge of foundations of StarAI models, the existing open-problems, and to motivate further applications of recently introduced theoretical results.

2 Target Audience, Prerequisites and Learning Goals

Recent years have seen increasing interest in Relational AI within the KR community. This interest is reflected in consistent StarAI related events in previous editions of KR, including David Poole’s 2020 keynote speech on StarAI, and the tutorial of Braun, Gehrke and Wilhelm at KR 2023. In this tutorial, we will bring to attention some foundational problems in StarAI. Hence, our tutorial will be of interest to the existing StarAI community at KR. Furthermore, our tutorial is also interesting to the community of complexity analysis of reasoning. As it will contain a thorough introduction to key ideas behind tractable WFOMC. We believe that this tutorial can be attended by anyone with basic knowledge of probability and first order logic.

The attendees will learn about the relation between problems like: model counting and probabilistic inference; the complexity of model counting problems; and about fragments of first order logic that admit tractable WFOMC. The tutorial will guide the audience through some of the key combinatorial ideas behind tractable WFOMC. The tutorial will also discuss the current state-of-the-art first-order model counters, their applications and their inefficiencies. The audience will be made aware of both theoretical and practical open-problems, whose resolution would lead to efficient probabilistic inference and learning in StarAI models.

The attendees will also be introduced to the problem of consistency of inference, and the recently developed theories of StarAI models that admit consistent parameter estimation. Such models, also known as *projective models*, open up a vast array of new avenues in StarAI, and the tutorial will provide the audience with the most recent developments. The audience will gain deep understanding of present theoretical models of projectivity and their recent applications. Finally, we will also discuss some recent results on domain-size generalization of non-projective models, and discuss how simple techniques like parameter re-scaling and regularization lead to provably better generalization across domain sizes.

3 Tutorial Outline

The tutorial will be divided into three parts:

Introduction [30 Minutes] We will first discuss some background on relational data and probability distributions on relational structures. We will formalize a general notion of probabilistic inference on relational data, and show how it maps to WFOMC for StarAI models like Markov Logic Networks and Problog. We will provide some initial examples, where tractability and inconsistency of inference can be major hurdles in real life applications.

Tractability [60 Minutes] This session will focus on conveying the key recent results in the field of WFOMC. We will first introduce the notion of types and tables in first order logic. We will then provide a thorough introduction to WFOMC in the universally quantified fragment of first order logic with two variables, providing a closed form formula for WFOMC of any universally quantified two-variable formula [17]. We will

then extend this result to admit existential quantifiers using the principle of inclusion-exclusion [18,19]. Finally, we will extend these results to admit cardinality constraints and counting quantifiers [20]. The tutorial will convey key algorithmic and combinatorial ideas [19] that will bring the audience a deep understanding of the state-of-the-art WFOMC. We will introduce key open problems in this area, ranging from scalability of existing model counters [21] to fine-grain analysis of known (potentially sub-optimal) upper-bounds on complexity of WFOMC. Finally, we will give a brief overview of recent results that allow for efficient model counting in first order logic fragments, with graph-theoretic constraints [22,23]. Such constraints, such as axiomatising a relation to represent a tree, forest, a connected graph etc., significantly extend the tractable fragment of WFOMC.

Consistency [60 Minutes] We will begin by introducing recent results [24] that have shown that a large variety of widely used probabilistic models, encompassing almost all of existing StarAI models [25], do not admit basic *consistency* requirements expected of sound statistical models. The lack of such consistency conditions means that StarAI models do not admit consistency of parameter estimation, i.e., as you see more and more data, it is not true that your parameters converge to the true value of the model parameters. Such inherent inconsistencies in StarAI models make them hard to be used across varying domain sizes [26]. We will discuss recent theoretical developments in identifying fragments of StarAI models, that admit consistency of parameter estimation [27,28]. These fragments also admit tractable inference in the strongest theoretical sense, as inference complexity is independent of the domain size. We will discuss the recently introduced rich theoretical framework for projective models [29], and the recent developments in their practical implementation and applications [30]. We will especially focus on the vast array of potential applications of these models in problems like reasoning over large scale data and random relational structure generation [30]. Finally, we will look into recent developments in understanding generalization properties of StarAI models that do not admit consistent parameter estimation [26,31], and how simple regularization and re-scaling approaches can lead to provably improved domain-size generalization in practice.

We will close the session with a discussion on open problems and potential new applications in this domain.

4 Presenter Bio

Sagar Malhotra is a PostDoc at TU Wien, hosted by Prof. Thomas Gärtner. He obtained his PhD in 2023 from the university of Trento on “Tractability and Consistency of Probabilistic Inference in Relational Domains” under the supervision of Luciano Serafini. During his PhD he worked on novel fragments of first order logic that admit tractable WFOMC [19,23]. He also worked on consistency of inference of Markov Logic Networks [27]. Recently, he has been interested in quantifying domain-size generalization of StarAI models which do not admit consistent parameter estimation [31].

References

1. Lise Getoor and Ben Taskar. *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2007.
2. Luc De Raedt, Kristian Kersting, Sriraam Natarajan, and David Poole. *Statistical Relational Artificial Intelligence: Logic, Probability, and Computation*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2016.
3. Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. Problog: A probabilistic prolog and its application in link discovery. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI'07*, page 2468–2473, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.
4. Céline Brouard, Christel Vrain, Julie Dubois, David Castel, Marie-Anne Debily, and Florence d’Alché Buc. Learning a markov logic network for supervised gene regulatory network inference. *BMC Bioinformatics*, 14:273, September 2013.
5. Luigi Bellomarini, Eleonora Laurenza, Emanuel Sallinger, and Evgeny Sherkhonov. Swift markov logic for probabilistic reasoning on knowledge graphs, 2022.
6. Maximilian Nickel, Kevin P. Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104:11–33, 2015.
7. Vaishak Belle. Statistical relational learning and neuro-symbolic ai: what does first-order logic offer?, 2023.
8. Robin Manhaeve, Sebastijan Dumancic, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. Deepproblog: neural probabilistic logic programming. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18*, page 3753–3763, Red Hook, NY, USA, 2018. Curran Associates Inc.
9. Manfred Jaeger. Learning and Reasoning with Graph Data: Neural and Statistical-Relational Approaches. In Camille Bourgaux, Ana Ozaki, and Rafael Peñaloza, editors, *International Research School in Artificial Intelligence in Bergen (AIB 2022)*, volume 99 of *Open Access Series in Informatics (OASIs)*, pages 5:1–5:42, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
10. Giuseppe Marra and Ondřej Kuželka. Neural markov logic networks, 2020.
11. Sriraam Natarajan, Kristian Kersting, Tushar Khot, and Jude Shavlik. *Boosted Statistical Relational Learners: From Benchmarks to Data-Driven Medicine*. Springer Publishing Company, Incorporated, 2015.
12. Sriraam Natarajan, Vishal Bangera, Tushar Khot, Jose Picado, Anurag Wazalwar, Vitor Santos Costa, David Page, and Michael Caldwell. Markov logic networks for adverse drug event extraction from text. *Knowledge and information systems*, 51:435–457, 2017.
13. Haiquan Chen, Wei-Shinn Ku, Haixun Wang, Liang Tang, and Min-Te Sun. Scaling up markov logic probabilistic inference for social graphs. *IEEE Transactions on Knowledge and Data Engineering*, 29(2):433–445, 2017.
14. Weizhe Zhang, Xiaoqiang Li, Hui He, Xing Wang, et al. Identifying network public opinion leaders based on markov logic networks. *The scientific world journal*, 2014, 2014.
15. Sebastian Riedel, Hong-Woo Chun, Toshihisa Takagi, and Jun’ichi Tsujii. A markov logic approach to bio-molecular event extraction. In *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*, pages 41–49, 2009.
16. Nikita A Sakhanenko and David J Galas. Markov logic networks in the analysis of genetic data. *Journal of Computational Biology*, 17(11):1491–1508, 2010.
17. Paul Beame, Guy Van den Broeck, Eric Gribkoff, and Dan Suciu. Symmetric weighted first-order model counting. In Tova Milo and Diego Calvanese, editors, *Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, pages 313–328. ACM, 2015.

18. Guy Van den Broeck, Wannes Meert, and Adnan Darwiche. Skolemization for weighted first-order model counting. In Chitta Baral, Giuseppe De Giacomo, and Thomas Eiter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press, 2014.
19. Sagar Malhotra and Luciano Serafini. Weighted model counting in FO2 with cardinality constraints and counting quantifiers: A closed form formula. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 5817–5824. AAAI Press, 2022.
20. Ondrej Kuzelka. Weighted first-order model counting in the two-variable fragment with counting quantifiers. *J. Artif. Intell. Res.*, 70:1281–1307, 2021.
21. Timothy van Bremen and Ondřej Kuželka. Faster lifting for two-variable logic using cell graphs. In Cassio de Campos and Marloes H. Maathuis, editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 1393–1402. PMLR, 27–30 Jul 2021.
22. Timothy van Bremen and Ondřej Kuželka. Lifted Inference with Tree Axioms. In *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning*, pages 599–608, 11 2021.
23. Sagar Malhotra, Davide Bizzaro, and Luciano Serafini. Lifted inference beyond first-order logic. *ArXiv*, abs/2308.11738, 2023.
24. Cosma Rohilla Shalizi and Alessandro Rinaldo. Consistency under sampling of exponential random graph models. *Annals of statistics*, 41 2:508–535, 2013.
25. Manfred Jaeger and Oliver Schulte. Inference, learning, and population size: Projectivity for SRL models. *CoRR*, abs/1807.00564, 2018.
26. Happy Mittal, Ayush Bhardwaj, Vibhav Gogate, and Parag Singla. Domain-size aware markov logic networks. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*, volume 89 of *Proceedings of Machine Learning Research*, pages 3216–3224. PMLR, 2019.
27. Sagar Malhotra and Luciano Serafini. On projectivity in markov logic networks. In Massih-Reza Amini, Stéphane Canu, Asja Fischer, Tias Guns, Petra Kralj Novak, and Grigorios Tsoumakas, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 223–238, Cham, 2023. Springer Nature Switzerland.
28. Felix Q. Weitkämper. An asymptotic analysis of probabilistic logic programming, with implications for expressing projective families of distributions. *Theory Pract. Log. Program.*, 21(6):802–817, 2021.
29. Manfred Jaeger and Oliver Schulte. A complete characterization of projectivity for statistical relational models. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4283–4290. International Joint Conferences on Artificial Intelligence Organization, 7 2020. Main track.
30. Manfred Jaeger, Antonio Longa, Steve Azzolin, Oliver Schulte, and Andrea Passerini. A simple latent variable model for graph learning and inference. In *The Second Learning on Graphs Conference*, 2023.
31. Florian Chen, Felix Weitkämper, and Sagar Malhotra. Understanding domain-size generalization in markov logic networks, 2024.