Banff International Research Station

for Mathematical Innovation and Discovery

15w5080: New Perspectives for Relational Learning
Arriving Sunday, April 19 and departing Friday, April 24

MEALS

*Breakfast (Buffet): 7:00 – 9:30 am, Sally Borden Building, Monday – Friday
*Lunch (Buffet): 11:30 am – 1:30 pm, Sally Borden Building, Monday – Friday
*Dinner (Buffet): 5:30 – 7:30 pm, Sally Borden Building, Sunday – Thursday
Coffee Breaks: As per daily schedule, in the foyer of the TransCanada Pipeline Pavilion (TCPL)
***Please remember to scan your meal card at the host/hostess station in the dining room for each meal.**

MEETING ROOMS

All lectures will be held in the lecture theater in the TransCanada Pipelines Pavilion (TCPL). An LCD projector, a laptop, a document camera, and blackboards are available for presentations.

CONTACT INFORMATION

Daniel Lowd (University of Oregon), lowd@cs.uoregon.edu
Sriraam Natarajan (Indiana University Bloomington), natarasr@indiana.edu
David Poole, (University of British Columbia), poole@cs.ubc.ca
Oliver Schulte (Simon Fraser University), oschulte@cs.sfu.ca. Main Contact.

DRAFT SCHEDULE OVERVIEW

This is a draft schedule; we welcome suggestions for changes. Our goal is to stimulate a maximum of discussion and collaboration.

Monday	Tuesday	Wednesday	Thursday	Friday
Welcome, Introduction. Plenary: What we can do/where we are at. • The state of the art. • High-Level, introductory.	Plenary: What we can't do/where we are going. • Open research challenges. • High-Level, introductory.	Half Day Plenary. Strategic Discussion. • Increasing Impact. • Publications from the workshop.	Specific Topics. • Parallel sessions in smaller groups.	Half Day. Specific Topics. • Parallel sessions in smaller groups. • Conclusion.

DRAFT SCHEDULE DETAILED

Sunday

16:00 Check-in begins (Front Desk – Professional Development Centre - open 24 hours)
17:30-19:30 Buffet Dinner
20:00 Informal gathering in 2nd floor lounge, Corbett Hall
Beverages and small assortment of snacks are available on a cash honor system.

Monday

7:00-8:45 Breakfast
8:45-9:00 Introduction and Welcome by BIRS Station Manager, TCPL
9:00-9:20 Introduction and Welcome by Workshop Organizers.
9:20-11:10 Session 1: What we can do already. Background and State of the Art.
9:20-9:40 Relational Representations. Daniel Lowd.
9:40-10:00 A unified view of generative models for networks: models, methods, opportunities, and challenges. Abigail Jacobs.
*10:00-10:*30 Coffee Break, TCPL.
*10:*30-10:50 Community structure analysis in social networks with relational latent feature models. Manfred Jaeger.
*10:*50-11:10 Latent Variable Modeling and Statistical Relational Learning: The Missing Links. James Foulds.
11:10-11:45 Session 2 (Panel Discussion – Panelists: Speakers of Session 1)
11:45-13:00 Lunch
13:00-14:00 Guided Tour of The Banff Centre; meet in the 2nd floor lounge, Corbett Hall
14:00-14:10 Group Photo; meet in foyer of TCPL (photograph will be taken outdoors so a jacket might be required).
14:10-16:00 *Session 3: What we can do already. Background and State of the Art.*
14:10-15:00 Lifted Inference in Probabilistic Databases -- An Overview. Dan Suciu.
15:00-15:30 Coffee Break, TCPL.
15:30-16:00 Unifying MAX SAT, Local Consistency Relaxations, and Soft Logic with Hinge-Loss Markov Random Fields. Stephen Bach.
16:00-16:30 Session 4 (Panel Discussion – Panelists: Speakers of Session 3)
16:30-17:30 Open Discussion: Current Strengths of Relational Learning, Goals for the Workshop Break-out Sessions
17:30-19:30 Dinner

Tuesday

7:00-9:00 Breakfast
9:00-10:50 9:00-10:50 Session 5: What we can't do. Open Research Challenges.
9:00-9:30 Probabilistic logic programming for robotics. Luc de Raedt.
9:30-10:00 Probabilistic modeling of cyclic relations and its application to plan recognition. Taisuke Sato.
*10:00-10:*30 Coffee Break, TCPL.
*10:*30-10:50 Relational Stochastic Programming: Frontiers for Planning and Learning. Scott Sanner.
10:50-11:45 Session 6 (Panel Discussion – Panelists: Speakers of Session 5).
11:45-13:00 Lunch
13:00-15:10 *Session 7:* What we can't do. Open Research Challenges.
13:00-13:30. Approximate Symmetries in Lifted Inference. Guy Van den Broeck.
13*:30-14:00* A simple method for multi-relational outlier detection. Oliver Schulte.

14:00-14:30 Coffee Break, TCPL.
14:30-14:50 (Panel Discussion – Panelists: Speakers of Session 7).
14:50-16:15 Open Discussion:
The Science of Relational Learning. Chair: Manfred Jaeger.
16:15-17:30 Break-out Sessions
17:30-19:30 Dinner

Wednesday

7:00-9:00 Breakfast
9:00-9:30 Session 8 (Publication based on Workshop)
BIRS Report
Book, Memos, Technical Reports, Manifestos,...
9:30-10:00 Success Stories: Relational Learning in Practice
10:00-10:30 Coffee Break, TCPL.
10:30-12:15 Session 9 (Open Discussion: Increasing Impact)
More Success Stories
Software Tools
Applications
- Bioinformatics
- Sports Analytics
- Computer Vision
- Health Data
- ...

Dataset repository
Education
Start new conference
Facilitate Cross-Community Interactions
- Relational Deep Learning
- Relational Planning/Game Playing/Multi-Agent Systems
- Probabilistic Databases

....
12:15-13:30 Lunch
Free Afternoon [*Tour of Banff Centre, Mountain Hike]*
17:30-19:30 Dinner

Thursday

7:00-9:00 Breakfast
9:00-10:00 Session 9 (Specific Topics) parallel sessions, 2-3 talks each.

Time	Topic	
	Inference and Learning	Open. Perhaps Ph.D. research spotlights?
9:00-9:30	Lifted Generative Learning of Markov Logic Networks. Jesse Davis.	Open
9:30-10:00	Paired-Dual Learning for Circumventing the Inference Bottleneck. Bert Huang.	Open.

10:00-10:30 Coffee Break, TCPL – available from 10:00 am onwards but must finish by 11:00 am
10:30-11:30 Session 10 (Specific Topics)

Time	Topic	
	Inference and Learning	Open
10:30-11:00	Efficient Online Collective Inference for Graphical Models. Jay Pujara.	Open.

11:00-11:30	Learning Guarantees for Relational Models in the “One Example” Setting. Ben London.	Open.

11:30-12:00 Panel Discussion
12:00-13:10 Lunch
13:10-14:20 Session 12 (Plenary Session), Demos and Successor Stories.
14:20-14:50 Coffee Break, TCPL – available from 2:00 pm onwards but must finish by 3:30 pm
14:50-16:20 Open Session. Representation Challenge: comparing strengths and weaknesses of relational representations. Chair: Scott Sanner.
16:20-17:30 Break-out Sessions
17:30-19:30 Dinner

Friday

7:00-9:00 Breakfast
9:00-10:15 Demo Session.
Informal Discussions recommended, as many participants must catch early flights
10:15-10:45 Coffee Break, TCPL – available from 10:00 am onwards but must finish by 11:00 am
10:45-11:30 Session 14: Closing and Concluding Remarks
11:30-12:00 Checkout
12:00-1:30 Lunch

** 5-day workshop participants are welcome to use BIRS facilities (BIRS Coffee Lounge, TCPL and Reading Room) until 3 pm on Friday, although participants are still required to checkout of the guest rooms by 12 noon. **

Abstracts to follow (if desired) in alphabetical order by last name of speaker.

Banff International Research Station

for Mathematical Innovation and Discovery

Name of 2013 5-day Workshop
Date of 2013 5-day Workshop

ABSTRACTS

(in alphabetic order by speaker surname)

Speaker: Stephen Bach (University of Maryland College Park)
Title: Unifying MAX SAT, Local Consistency Relaxations, and Soft Logic with Hinge-Loss MRFs
Abstract:
Relational machine learning offers the ability to model rich, structured problems but quickly runs into problems of scalability. Many approaches have been proposed to scale up inference and learning algorithms, often using convex programming to approximate NP-hard problems. In this talk I will present our latest results on unifying different approaches to convex MAP inference, a crucial problem that arises during both prediction and learning tasks. I will focus on three seemingly distinct approaches, (1) randomized algorithms for MAX SAT, (2) local consistency relaxation for Markov random fields, and (3) reasoning about continuous information with fuzzy logic. I will show how all three actually optimize the same convex inference objective, which is extremely useful. It means that a single formalism and set of algorithms can combine the advantages of all three into a tool that can reason scalably and accurately about both discrete and continuous information. We have generalized this unified inference objective into a new kind of graphical model, hinge-loss Markov random fields (HL-MRFs), and created a probabilistic programming language called probabilistic soft logic (PSL) to make them easy to use. I will then discuss how HL-MRFs and PSL can be applied to a wide range of relational learning problems, leading to models that are accurate, easy to define, and highly scalable.

Speaker: Jesse Davis (Katholieke Universiteit Leuven)
Title: Lifted Generative Learning of Markov Logic Networks
Abstract:
Markov logic networks (MLNs) are a well-known statistical relational learning formalism that combine Markov networks with first-order logic. MLNs attach weights to formulas in first-order logic. Learning MLNs from data is a challenging task as it requires searching through the huge space of possible theories. Additionally, evaluating a theory's likelihood requires learning the weight of all formulas in the theory. This in turn requires performing probabilistic inference, which, in general, is intractable in MLNs. Lifted inference speeds up probabilistic inference by exploiting symmetries in a model. We explore how to use lifted inference when learning MLNs. Specifically, we investigate generative learning where the goal is to maximize the observed data's likelihood. First, we provide a generic algorithm for learning maximum likelihood weights that works with any exact lifted inference approach. In contrast, most existing approaches optimize approximate measures such as the pseudo-likelihood. Second, we provide a concrete parameter learning algorithm based on first-order knowledge compilation. Third, we propose a structure learning algorithm that learns liftable MLNs, which is the first MLN structure learning algorithm that exactly optimizes the

likelihood of the data. Finally, we perform an empirical evaluation on three real-world datasets. Our parameter learning algorithm results in more accurate models than several competing approximate approaches. It learns more accurate models in terms of the test set log-likelihood as well as prediction tasks. Furthermore, our tractable learner outperforms intractable models on prediction tasks suggesting that liftable models are a powerful hypothesis space, which may be sufficient for many standard learning problems.

Speaker: James Foulds (University of California Santa Cruz)
Title: Latent Variable Modeling and Statistical Relational Learning: The Missing Links
Abstract:
Probabilistic and statistical machine learning has become a broad community, and it is perhaps unsurprising that there are a diversity of opinions on how to go about probabilistic modeling. The statistical relational learning community has designed powerful general-purpose models to encode complex dependency structures between random variables, such as Markov logic networks, relational logistic regression, Bayesian logic programs, and probabilistic soft logic. In another direction, the latent variable modeling camp has developed useful specialized models for identifying interpretable latent structure, such as topic models and latent community membership models. In this context, we propose latent topic networks, a modeling framework which aims to be a ``missing link'' between the latent variable and statistical relational learning communities: latent topic networks provide a general-purpose framework for reasoning over interpretable ``missing'' or hidden variables with complex dependency ``links'' between them. This work was performed in collaboration with Shachi Kumar and Lise Getoor, from the University of California, Santa Cruz.

Speaker: Bert Huang (Virginia Tech.)

Title: Paired-Dual Learning for Circumventing the Inference Bottleneck

Abstract:
Inference is often the primary computational expense during training of probabilistic models, and inference in structured or relational models is not cheap. Advances on tractable model families and algorithms for fast inference have significantly improved the scalability of learning in complex settings. Nevertheless, many learning approaches for settings with latent variables, such as expectation-maximization, still require many rounds of inference to optimize their objective functions. In this talk, I will describe an approach called paired-dual learning (PDL) that enables training of models with partially labeled data that can avoid the need for multiple inferences. PDL uses a saddle-point objective function that pairs two dual inference objectives. Gradient-based optimization of this paired-dual objective can be done using early, unconverged solutions of the respective inferences. The paired-dual objective is itself a dual of a variational likelihood, so its solutions are the same as traditional approaches. In experiments, PDL finds accurate models much faster than these traditional learning algorithms. In the last part of the talk, I will discuss where PDL fits in among other lines of research on learning algorithms that circumvent the bottleneck of inference.

Speaker: Ben London (University of Maryland College Park)
Title: Learning Guarantees for Relational Models in the “One Example” Setting
Abstract:
Statistical relational learning (SRL) techniques have been very effective in the “one example” inductive setting, in which the training data consists of a single fully labeled graph, and the related transductive setting, in which the goal is to complete the labeling of a fixed, partially labeled graph. Despite the empirical success of SRL, there are few formal theoretical guarantees on the quality of the learned predictor. The lack of guarantees is partly due to the difficulties of analyzing interdependent random variables. For collective models (i.e., those that perform joint inference), another challenge is caused by the dependence induced by the

predictor, which affects the stability (i.e., variance) of the predictions. In this talk, I will present some recent advances towards a better understanding of generalization error in SRL, focusing on new PAC learning guarantees for the one example learning paradigm. The bounds highlight the importance of templating (i.e., parameter-tying), which is implicitly used by many SRL models. I will also highlight some open theoretical problems to motivate new research in the SRL community.

Speaker: Abigail Zoe Jacobs (University of Colorado Boulder)
Title: A unified view of generative models for networks: models, methods, opportunities, and challenges
Abstract:
Research on probabilistic models of networks now spans a wide variety of fields, including physics, sociology, biology, statistics, and machine learning. These efforts have produced a diverse ecology of models and methods. Despite this diversity, many of these models share a common underlying structure: pairwise interactions (edges) are generated with probability conditional on latent vertex attributes. Differences across these models generally stem from different philosophical choices about how to learn from data or different empirically-motivated goals. The highly interdisciplinary nature of work on these generative models, however, has inhibited the development of a unified view of their similarities and differences. For instance, novel theoretical models and model-fitting techniques developed in machine learning are largely unknown within the social and biological sciences, which have instead emphasized model interpretability. Here, we describe a unified view of generative models for networks that draws together many of these disparate threads and highlights the fundamental similarities and differences that span these fields. Navigating this space of models and modeling approaches illuminates new areas of exploration and points of connection with nearby domains. Finally, we describe a number of opportunities and challenges for future work that are revealed by this view.

Speaker: Manfred Jaeger (Aalborg University)
Title: Community structure analysis in social networks with relational latent feature models
Abstract:
The design of hybrid probabilistic relational models that combine discrete and numeric variables faces serious challenges in terms of computational tractability. In this talk I present a limited but already very useful extension of relational Bayesian networks with numeric input relations. These numeric relations are only used to condition the distribution of discrete variables on numeric data, and they are algorithmically handled just like model parameters. The usefulness of the resulting framework is demonstrated by showing how latent feature models for community structure in social networks can be implemented, and used to discover complex structural information in multi-relational social networks.

Speaker: Jay Pujara (University of Maryland College Park)

Title: Efficient Online Collective Inference for Graphical Models

Abstract:
Many applications of statistical relational learning operate online, where the inferences of a probabilistic model must be updated in response to new evidence. When these models contain millions of interdependent variables, jointly updating the most likely (i.e., MAP) configuration of the variables each time new evidence is encountered can be infeasible, even if inference is tractable. We address this problem setting by formulating the problem of efficient online collective inference, where a graphical model is updated efficiently by revising the assignments to a subset of the variables while holding others fixed. To formalize the consequences of partially updating the MAP state as new evidence becomes available, we introduce the concept of inference regret. We derive inference regret

bounds for a class of graphical models with strongly-convex free energies. These theoretical insights, combined with a thorough analysis of the optimization solver, motivate new approximate methods for efficiently updating the variable assignments under a budget constraint. We show how these methods are important in the real-world problem of knowledge graph identification, where we infer a knowledge base with millions of facts using millions of ontological dependencies between a continually evolving set of extractions from a noisy Web corpus. Our experimental results demonstrate that budgeted online inference can reduce inference time dramatically while maintaining accuracy comparable to full inference.

Collaborators: Ben London, Hui Miao, William Cohen, Lise Getoor

Speaker: Luc De Raedt (KU Leuven)
Title: Probabilistic logic programming for robotics
Abstract:
Robotics is a rich domain that requires both high-level reasoning and reasoning about uncertainty. Therefore, probabilistic logics and statistical relational learning has a lot of potential for contributing to this domain.
While on the one hand, there is a rapidly increasing interest in using such SRL or PL methods in robotics, there are several important remaining challenges for probabilistic programming and statistical relational learning that must be tackled before SRL or PL can be routinely used. These challenges include: dealing with dynamics, with continuous and discrete distributions and with changing domain sizes. In this talk I shall report on these challenges and on the progress that has been made with the framework of Distributional Clauses to tackle these.

Speaker: Scott Sanner (Australian National University)
Title: Relational Stochastic Programming: Frontiers for Planning and Learning
Abstract:
I discuss a (relatively) new language called RDDL, which is a relational stochastic programming formalism developed to compactly model real-world planning and scheduling problems, including those required to enable applications in Smart Cities. Along with RDDL come many interesting and unsolved challenges related to model translation, learning, and optimization (planning) that I will discuss.

Speaker: Taisuke Sato (Tokyo Institute of Technology)
Title: Probabilistic modeling of cyclic relations and its application to plan recognition
Abstract:
Probabilistic modeling of cyclic relations such as Markov chains requires the computation of infinite sum of probabilities. In this talk we present recent advances in the logic-based modeling language PRISM which include a new built-in predicate to compute such infinite sum of probaiblities specified by a set of cyclic propositional formulas. We also apply our technique to plan recognition form incomplete observations.

Speaker: Oliver Schulte (Simon Fraser University)
Title: A simple method for unsupervised multi-relational outlier detection. (With tools that you probably have around the lab.)
Abstract:
I describe a simple method for leveraging statistical-relational parameter learning to compute an outlier score. Suppose we have learned a model structure for a domain (e.g., a Markov Logic Network), with parameter values estimated from the entire population. For an individual that is a potential outlier, we can construct an individual data view that summarizes the information for that individual only. Applying parameter learning to the

individual data view produces another set of individual parameter estimates. The distance between the population parameter values and the individual parameter values can then be used as an outlier score. For example, in soccer the distance between the parameters estimated from the matches of a specific player, vs. the parameters estimated from all matches, can be used as an outlier score for that player. We apply this approach to synthetic data, players and teams from the UK Premier league, and movies in the IMDb. I will briefly describe how this approach relates to (1) propositionalization for outlier detection, and (2) using distribution divergences for comparing individuals in multi-relational data.

Speaker: Dan Suciu (University of Washington)
Title: Lifted Inference in Probabilistic Databases -- An Overview
Abstract:
Probabilistic Databases (PDBs) represents uncertainty in a relational form, much like statistical relational learning (SRL). Both are connected through the common use of relational language, and both have explored lifted inference as an approach to scale up probabilistic inference. Talk will describe the commonalities and differences between PDBs and SRLs, then will present the key results on lifted inference in PDBs, including a list of open problems.