Noname manuscript No.
(will be inserted by the editor)

Joint Probabilistic Inference of Causal Structure

Dhanya Sridhar · **Lise Getoor**

Received: date / Accepted: date

Abstract Causal directed acyclic graphical models (DAGs) are powerful reasoning tools in the study and estimation of cause and effect in scientific and socio-behavioral phenomena. In many domains where the cause and effect structure is unknown, a key challenge in studying causality with DAGs is learning the structure of causal graphs directly from observational data. Traditional approaches to *causal structure discovery* are categorized as constraint-based or score-based approaches. Score-based methods perform greedy search over the space of models whereas constraint-based methods iteratively prune and orient edges using structural and statistical constraints. However, both types of approaches rely on heuristics that introduce false positives and negatives. In our work, we cast causal structure discovery as an inference problem and propose a joint probabilistic approach for optimizing over model structures. We use a recently introduced and highly efficient probabilistic programming framework known as *Probabilistic Soft Logic* (PSL) to encode constraint-based structure search. With this novel probabilistic approach to structure discovery, we leverage multiple independence tests and avoid early pruning and variable ordering. We compare our method to the notable PC algorithm on a well-studied synthetic dataset and show improvements in accuracy of predicting causal edges.

Keywords Causal Structure Discovery · Constraint-based Structure Discovery · Probabilistic Programming · Joint Inference

D. Sridhar · L. Getoor
Department of Computer Science
University of California Santa Cruz, CA
E-mail: {dsridhar, getoor}@soe.ucsc.edu

1 Introduction

Causal models encode cause and effect relationships between variables in a system, and are of interest to economists, scientists, and statisticians alike. Perhaps the most notable in this class of models are *causal directed acyclic graphical models* (causal DAGs), which provide an intuitive and compact notation for expressing cause and effect dependencies between variables. Unlike standard, associative DAGs, edges in a causal DAG indicate direct cause relationships between the parent and child. Even with its conditional probability parameters fixed, a causal DAG still represents a family of distributions that admit manipulations to variables, known as interventions. Interventions are studied using a mathematical framework called *do-calculus* and estimate causal effects between variables. Interventions in causal DAGs are important computational mechanisms for reasoning about the effect that changes to a particular variable could have on downstream variables. In some domains, human experts encode causal DAGs and computational methods perform causal inferences with interventions. However, in many domains where expert knowledge is limited or insufficient for developing a causal model, the goal of computational methods is to discover causal DAGs from observed data.

In the context of causal DAGs, *causal structure discovery* refers to learning equivalence classes of causal DAGs directly from observational data. Since passively observed data does not suffice to identify a single causal DAG, structure discovery methods output a partially oriented DAG representing an equivalence class or an exemplar DAG of the class. Approaches to causal structure discovery from observed data can be characterized as score-based, constraint-based, or hybrids of the two. Typically, score-based methods take a Bayesian approach to greedily search and score the posterior probability of the posited structure given the data subject to regularization on the model complexity [2, 3, 4,

5, 13, 9]. Scoring functions include Bayesian scores such as Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) and information theoretic scores such as Minimum Description Length (MDL) [9]. However, these search-and-score methods are computationally expensive due to the exponential space of possible model.

In contrast to score-based approaches, constraint-based methods start with a complete undirected graph and iteratively prune edges to output a partially oriented DAG that satisfies structural and acyclicity constraints implied by conditional independence tests and d-separation criteria [24]. The PC [1] algorithm is perhaps the most notable of constraint-based structure discovery approaches [23]. PC and its subsequent extensions are sound and complete under particular assumptions, and computationally efficient, they remain sensitive to false positives and negatives from independence tests, and inherent orderings in the algorithm. Rich and complex data are especially susceptible to noisy independence testing. Because these approaches are iterative, errors propagate throughout phases of the algorithm. Although constraint-based methods admit domain knowledge constraints through forced edges or non-edges and fixed orderings over variables, use of more flexible domain knowledge constraints remains an open problem.

Constraint-based causal structure discovery has also been studied from the lens of *MAX-SAT* and linear programming [16, 15], two well-studied ways to solve constraint satisfaction problems. In this work, we recognize that causal structure discovery can also be viewed as inference problem. We extend this viewpoint by formulating the problem of joint probabilistic inference of causal structure and introducing an approach for inferring causal edges and adjacencies between variables from observational data. Our approach uses a recently introduced probabilistic programming framework, Probabilistic Soft Logic (PSL). PSL is a templating language for a particular class of continuous Markov random field models that enjoy exact and efficient inference. Motivated by the PC algorithm, we model structural and d-separation constraints with the PSL framework to jointly infer causal structure, without early pruning or reliance on variable orderings.

Our main technical contributions include:

- Extending the causal structure discovery problem within the PSL framework as a joint inference problem
- Formulating a constraint-based probabilistic modeling approach motivated by the PC algorithm.
- Evaluating our approach on publicly available, well-studied synthetic dataset from the Causality Workbench [14].

Section 2 details the problem of causal structure discovery and provides background on the constraint-based PC algorithm. In Section 3, we formulate the problem of joint causal structure inference and describe our probabilistic framework for encoding such a model. Section 4 presents the dataset and results from our evaluation experiment. In Section 5, we outline related work in causal structure discovery and propose multiple avenues for future study in Section 6.

[1] PC stands for Peter-Clark, the authors of the algorithm

2 Causal Structure Discovery Problem

Formally, a directed acyclic graphical model (DAG) $\mathscr{G} = (\mathbf{V}, \mathbf{E})$ over a set of variables $\mathbf{V} = \{X_1 \dots X_n\}$ defines a joint probability distribution of the form

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i | \mathit{Parents}(X_i))$$

where $\mathit{Parents}(X_i)$ refer to the direct ancestors of a variable X_i in $\mathscr{G}$ and $P(X_i|\mathit{Parents}(X_i)$ parametrize the model. Thus, $\mathscr{G}$ encodes conditional independences between variables in the underlying distribution. The criteria for *d-separation* determine all conditional independences of the form $I(X,Y;S)$ where S is the conditioning set. $P(X_1, \dots, X_n)$ for $\mathscr{G}$ satisfies the *global Markov condition* if and only if for any three disjoint subsets of variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ from $\mathbf{V}$, if $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$, then $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$ in P. Intuitively, the property guarantees a one-to-one mapping between the independences in $\mathscr{G}$ and P, allowing us to reason over $\mathscr{G}$ directly instead of P.

As mentioned, causal DAGs have additional semantics to encode cause-and-effect relations between variables. Formally, we say that X is a *cause* of Y if the probability distribution of Y changes as values of X change. Furthermore, X is a *direct cause* of Y if manipulations of X change the probability distribution of Y regardless of manipulations to any other variables in $\mathbf{V} \backslash \{X, Y\}$. A causal DAG $\mathscr{G}_c = (\mathbf{V}, \mathbf{E})$ contains an edge from X_i to X_j only if X_i is a direct cause of X_j. In the causal structure discovery problem, we assume that there exists a true, underlying $\mathscr{G}_c$ that describes $P(X_1 \dots X_n)$ for variables $\mathbf{X} = \{X_1 \dots X_n\}$. Given a set of m observations $\{\mathbf{X}_1 \dots \mathbf{X}_m\}$ of the variables, the goal of the structure discovery algorithm is to identify causal edges $\mathbf{E}_c$ and undirected edges $\mathbf{E}_u$ when adjacency but not orientation can be determined.

2.1 Constraint-based Discovery with PC

This section focuses on constraint-based structure discovery and provides background on the foundational PC algorithm. PC outputs an equivalence class of DAGs that contain $\mathscr{G}_c$, known as a *CP-DAG*. A CP-DAG consists of both directed edges where a causal edge is uncovered and undirected edges when only associational, not causal, dependence can be established.

2.1.1 The PC Algorithm

Algorithm 1 below describes the key idea of PC at a high-level.

Algorithm 1 PC Algorithm

1: **procedure** PC(Observations $X_1 \dots X_n$, Conditional Independence Testing Procedure) **Require:** An ordering over variables X
2: Find the skeleton graph $\mathscr{C}$ and separation sets according to the iterative procedure described below.
3: Orient unshielded triples in $\mathscr{C}$ based on the separation sets
4: In $\mathscr{C}$, orient as many of the remaining undirected edges as possible using rules 1-3 described below.

Finding the skeleton graph The first step of the PC algorithm relies on conditional independence tests and the principle of d-separation in a DAG to find all undirected edges between nodes, producing a skeleton graph. The later steps of the algorithm orient as many edges as possible in this skeleton graph. The skeleton graph subroutine begins with a complete graph and uses conditional independence to iteratively remove edges. That is, if $x_i \perp\!\!\!\perp x_j | \mathbf{S}$, the edge $x_i - x_j$ is removed. The algorithm typically queries a procedure that tests for independence using the sample data.

The algorithm iterates over the size l of conditioning set $\mathbf{S}$, starting with $l = 0$ when marginal independences between variables are tested. To prevent the combinatorial number of possible independence tests, the algorithm assumes an ordering over the variables and performs tests based on the ordering and each variable's adjacency list. If an edge is removed, it will never be considered again, thereby admitting more efficient search over the space.

As edges are iteratively removed based on l and the ordering, we define a separation set of i and j as the conditioning set $\mathbf{S_{ij}}$ such that $x_i \perp\!\!\!\perp x_j | \mathbf{S_{ij}}$. The algorithm stops when there are no more conditioning sets of size l or when l reaches a pre-specified number. The resulting graph is denoted $\mathscr{C}$. The separation sets are used in the next stage of orienting unshielded triples.

Orienting triples and other orientation rules The remainder of the algorithm repeatedly applies the following rules until as many edges are oriented as possible:

- Determine v-structures by considering all consecutive edges $x_i - x_j - x_k$ in $\mathscr{C}$ and orienting as a v-structure if $x_j \notin \mathbf{S_{ik}}$.
- Orient $x_j - x_k$ into $x_j \to x_k$ whenever there is a directed edge $x_i \to x_j$ such that x_i and x_k are not adjacent.
- Orient $x_i - x_j$ into $x_i \to x_j$ whenever there is a chain $x_i \to x_k \to x_j$.
- Orient $x_i - x_j$ into $x_i \to x_j$ whenever there are two chains $x_i - x_k \to x_j$ and $x_i - x_l \to x_j$ such that x_k and x_l are not adjacent.

The resulting output after repeated application is a partially oriented graph known as the CP-DAG. At a high-level, PC consists of a graph search phase followed by a constraint satisfaction phase to avoid cycles and additional v-structures.

3 Joint Probabilistic Causal Structure Inference

This section first introduces the problem of inferring causal structure and provides an overview on the Probabilistic Soft Logic (PSL) framework. We describe our approach to modeling causal structure discovery with PSL for joint inference of causal and adjacency edges.

3.1 The Probabilistic Causal Structure Discovery Problem

In the previous section, we review the PC algorithm. PC suffers from two major drawbacks: (1) noise from conditional independence testing and (2) incorrect edge removals are never recovered and errors propagate. In this section, we introduce the novel problem of joint probabilistic inference of causal structure to better trade-off against noisy information and more robustly uncover the underlying causal graph. Unlike previous methods that either greedily search over structures or iteratively prune edges to find a structure, we encode the search over possible causal DAGs as optimization within PSL.

Again given n observations of values for X, the challenge of probabilistic inference of causal DAGs is to model the distribution over possible CP-DAGs. That is, we want to jointly infer all adjacencies and oriented edges. In this setting, finding the most probable CP-DAG reduces to *maximum a posteriori* (MAP) inference in this distribution.

3.2 Probabilistic Soft Logic and Hinge-loss Markov Random Fields

Our approach focuses on a special class of Markov random field (MRF) known as Hinge-loss MRF (HL-MRF). This section reviews the foundations of HL-MRFs and gives background on Probabilistic Soft Logic (PSL), the templating language that describes these models.

Like other probabilistic modeling frameworks, notably Markov logic networks, PSL uses a logic-like language for defining the potential functions for a special form of conditional random field [20]. HL-MRFs are log-linear exponential family models that admit efficient, scalable and exact maximum a posteriori (MAP) inference [1]. These models

Rule 1: $\text{INDEPENDENT}(A,B,SepSet) \rightarrow \neg\text{ADJ}(A,B)$

Rule 2: $\text{ADJ}(A,B) \wedge \text{ADJ}(B,C) \wedge \neg\text{INSEPSET}(B,A,C) \rightarrow \text{CAUSES}(A,B)$

Rule 3: $\text{ADJ}(A,B) \wedge \text{ADJ}(B,C) \wedge \neg\text{INSEPSET}(B,A,C) \rightarrow \text{CAUSES}(C,B)$

Rule 4: $\text{ADJ}(B,C) \wedge \text{CAUSES}(A,B) \wedge \neg\text{ADJ}(C,B) \rightarrow \text{CAUSES}(B,C)$

Rule 5: $\text{ADJ}(A,B) \wedge \text{CAUSES}(A,C) \wedge \text{CAUSES}(C,B) \rightarrow \text{CAUSES}(A,B)$

Rule 6: $\text{ADJ}(A,B) \wedge \text{ADJ}(A,C) \wedge \text{CAUSES}(C,B) \wedge \text{ADJ}(A,D) \wedge \text{CAUSES}(D,B) \wedge \neg\text{ADJ}(C,D) \rightarrow \text{CAUSES}(A,B)$

Fig. 1 Encoding constraint-based PC algorithm causal structure search criteria as logical rules in PSL.

are defined over continuous random variables, which provide a natural interpretation for real-valued similarities. MAP inference in HL-MRFs is a convex optimization problem over these variables. Formally, a hinge-loss MRF defines a joint probability density function of the form:

$$P(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z}\exp\Big(-\sum_{r=1}^{M}\lambda_r\phi_r(\mathbf{Y},\mathbf{X})\Big)$$

where the entries of target variables $\mathbf{Y}$ and observed variables $\mathbf{X}$ are in $[0,1]$, λ is a vector of weight parameters, $\mathscr{Z}$ is a normalization constant, and

$$\phi_r(\mathbf{Y},\mathbf{X}) = (\max\{l_r(\mathbf{Y},\mathbf{X}),0\})^{\rho_r}$$

is a *hinge-loss potential* specified by a linear function l_r and optional exponent $\rho_r \in \{1,2\}$. Relaxations of first-order logic rules are one way to derive the linear functions $l_r(\cdot)$ in the hinge-loss potentials ϕ_r. Given a collection of logical implications based on domain knowledge described in PSL and a set of observations from data, the rules are instantiated, or grounded out, with known entities in the dataset. Each instantiation of the rules maps to a hinge-loss potential function $\phi_r(\mathbf{Y},\mathbf{X})$ as shown above, and the potential functions define an HL-MRF model.

To illustrate modeling in PSL, we consider a prototypical similarity based rule that encourages transitive closure for link prediction between entities a,b,c:

$$\text{SIMILAR}(a,b) \wedge \text{LINK}(b,c) \rightarrow \text{LINK}(a,c)$$

where instantiations of the predicate LINK represent continuous target variables for a link prediction task and instantiations of SIMILAR are continuous observed variables. The convex relaxation of this logical implication derived using the well-known Lukasiewicz logic for continuous truth values is equivalent to the hinge-loss function

$$\max(\text{SIMILAR}(a,b) + \text{LINK}(b,c) - \text{LINK}(a,c) - 1, 0)$$

and can be understood as its *distance to satisfaction*. The distance to satisfaction of this ground rule is a linear function of the variables and thus, exactly corresponds to

$$\phi_r(\text{LINK}(b,c), \text{LINK}(a,c), \text{SIMILAR}(a,b))$$

the feature function that scores configurations of assignments to the three variables. Intuitively, distance to satisfaction represents the degree to which the rule is violated by assignments to the random variables conditioned on the observations. Thus, MAP inference minimizes the weighted, convex distances to satisfaction to find an consistent joint assignment for all the target variables:

$$\arg\min_{\mathbf{y}\in[0,1]^n} \sum_{r=1}^{m} w_r \max\{l_r(\mathbf{y},\mathbf{x}),0\}$$

Higher rule weights induce higher penalties for violating the rule increasing its relative importance to other rules. Weights are learned from data through maximum likelihood estimation using training data and the structured perceptron algorithm. Exact MAP inference is performed on the learned model to find the most likely assignments for variables using the consensus based ADMM algorithm. PSL supports latent variable modeling with additional EM-based learning algorithms. For a full description of PSL, see Bach et al. [1] Thus, PSL rules encode the domain knowledge that leads to a consistent assignment to all target variables. HL-MRFs have achieved state-of-the-art performance in many domains including knowledge graph identification [18], student engagement understanding in MOOCs [19], drug-target interaction prediction [11, 10], social spam detection [12], and recommendation [17]. The open source PSL software can be downloaded from the website (`http://psl.umiacs.umd.edu/`).

3.3 The PC-PSL Model

Since MAP inference in PSL can be viewed as a soft-constraint optimization, we encode the skeleton and orientation constraints used by PC as a PSL model. We describe in detail below the predicates and rules used in our PSL model of joint causal structure inference. Figure 1 shows all the logical formulas defined in PSL to represent the variables and constraints used in the PC algorithm.

3.3.1 Predicates for Causal Inference

We define the key predicates INDEPENDENT, CAUSES, and ADJ. We include an auxiliary predicate INSEPSET for use in orienting colliders based on d-separation. From conditional independence tests performed between pairs of variables, with varying separation sets, we fully observe p-values and treat them as evidence for INDEPENDENT. Thus, INDEPENDENT is an observed predicate whose groundings represent the evidence $\mathbf{X}$ in the conditional HL-MRF distribution.

Orientation of v-structures, or colliders, requires reasoning about separation sets for conditional independence. We enable checking separation sets with groundings of the INSEPSET predicate. The model jointly infers values of both $\text{CAUSES}(A,B)$ and $\text{ADJ}(A,B)$ for all possible pairs of nodes A,B. Thus, groundings of CAUSES and ADJ are the unobserved, or target, variables $\mathbf{Y}$ in $P(\mathbf{Y}|\mathbf{X})$.

3.3.2 PC-PSL Rules

Figure 1 includes all the rules used in PC represented as logical implications involving the predicates described above. Rule 1 corresponds to the skeleton phase of the PC algorithm, where undirected edges are pruned based on strong evidence of independence between variables. PC prunes iteratively, exploiting the reduction in adjacencies to more efficiently perform conditional independence tests between adjacent variables given adjacent conditioning sets. However, the iterative pruning introduces false positives and negatives. In contrast, the PC-PSL model optimizes and infers adjacencies, allowing for more than one independence test between variables based on many possible separation sets.

Rules 2-3 correspond to the orientation of colliders from chains, or unshielded triples. This rule encodes the well-known d-separation criterion that a variable not in the conditioning set of independent variables must be a collider. The unshielded triples are represented by conjunctions of ADJ variables and checks for membership in separation set are represented with INSEPSET.

The remainder of the rules, 4-6, correspond to orientation rules 1-3 used by PC and infer causal edges represented by predicate CAUSES. The rules encode constraints to avoid cycles and any additional v-structures. Specifically, rule 4 prevents 3-cycles within partially oriented triples, or chains of three variables. Rule 5 prevents additional v-structures from forming among these partially oriented triples. Finally, rule 6 prevents cycles from forming along two separate chains of three variables.

Table 1 Average accuracy and F1 score and standard deviation of *causal* edge discovery comparing PC and PC-PSL on LUCAS dataset.

Method	Causal Edge Accuracy	Causal Edge F1-score
PC	0.91 ± 0.06	0.53 ± 0.26
PC-PSL	0.94 ± 0.02	0.58 ± 0.19

4 Evaluation

In this section, we describe the synthetic dataset used for experimental evaluation. We outline the evaluation setup and task, and present results from our preliminary studies.

4.1 Dataset

We use the Lung Cancer Simple Set (LUCAS) synthetic dataset made publicly available through the Causality Workbench [14]. The true causal DAG consists of 12 binary variables: (1) Smoking, (2) Yellow Fingers, (3) Anxiety, (4) Peer Pressure, (5) Genetics, (6) Attention Disorder, (7), Born an Even Day, (8) Car Accident, (9) Fatigue, (10) Allergy, (11) Coughing and (12) Lung Cancer.

The true causal graph consists of 12 causal edges between variables. Figure 2 shows the underlying ground truth causal DAG for this system. The dataset contains 2000 observations of all variables.

4.2 Preliminary Results

We perform G^2 statistical independence tests between all pairs of variables. The G^2 test is closely related to the χ^2 test of independence and also Kullback-Leibler divergence. Then, for all pairs of variables, we enumerate over all combinations of conditioning sets up to maximum set size of 3 and perform G^2 tests of conditional independence. Each X, Y, and separation set S constitutes a grounding of INDEPENDENT, with the independence test's p-value encoding the soft truth score used by PSL. For each $Z \in S$ between X and Y, we construct groundings for INSEPSET. The inference task is to predict truth values for CAUSES and ADJ between all pairs of X and Y.

We compare the PC-PSL model against the PC algorithm implemented in the `pcalg` library for the Python language. We run PC also with G^2 independence tests, maximum separation set size of 3, and with pruning threshold $\alpha = 0.01$. We evaluate PC and PC-PSL for the accuracy and F-measure of predicted causal edges with 3-fold cross validation. We split the possible causal edges into folds and use training folds in turn to select thresholds for rounding PC-PSL [0,1] output values to $\{0,1\}$. We perform grid search and select the best performing thresholds on the training

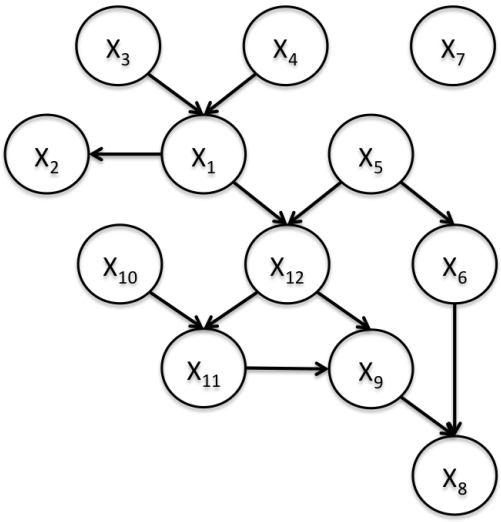

Fig. 2 Synthetic LUCAS ground truth causal DAG among 12 binary variables described below.

folds and evaluate on the test fold. We compare our results with the performance of PC on each test fold.

Table 1 shows the average accuracy and F1-score and standard deviations for causal edge identification comparing PC and PC-PSL. The PC-PSL sees gains over the PC algorithm in identifying true causal edges. Moreover, inference for PC-PSL completes in seconds compared to minutes required by the PC algorithm, even with additional independence test evidence.

5 Related Work

Our work builds on the foundational constraint-based PC algorithm introduced by Spirtes and Glymour [23] for structure discovery. As we note in section 3, PC remain sensitive to false positives and negatives from independence tests, and because the algorithm is iterative, errors propagate to later phases of the algorithm. Extensions of PC include the Fast Causal Inference (FCI) algorithm [22] that admits latent variables and confounders, and a variable order-independent variant of PC [7]. In our work, we instead encode the constraints of the PC algorithm in the PSL framework to perform constraint-based optimization and find a joint assignment to all causal edges without iterative pruning or dependencies on ordering.

Perhaps most similar to our work, Hyttinen et al. [15] study discovery of cyclic DAGs by encoding causal edges as variables in a MAX-SAT problem with constraints based on d-separation criteria. However, in our work, we develop a fully probabilistic that can be viewed as a relaxation of MAX-SAT. Recently, ongoing work on inference-based structure discovery uses PSL to encode the constraint-based Logical Causal Inference (LoCI) algorithm [6], mapping the proposed logical formulas to HL-MRFs templated with PSL.

In a separate line of work, *score-based* approaches to structure learning search over the space of possible models to find the model that best fits the data. Chickering [2, 3, 4], Chickering et al. [5], Friedman and Goldszmidt [13], De Campos and Ji [9] take a Bayesian approach to greedily search and score the posterior probability of the posited structure given the data, $P(S|D)$, subject to regularization on the model complexity Scoring functions include Bayesian scores such as Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) and information theoretic scores such as Minimum Description Length (MDL) [9]. In contrast to constraint-based methods, score-based approaches start with an empty or randomly initialized DAG and add, remove or negate edges, scoring each manipulation to the model until a sufficiently good structures is found. Although in this work we focus on applying PSL as a constraint-based method, we see our approach as performing a meta-search, or optimization, over the space of possible model structures. In between score- and constraint-based approaches, recently there has been a rich body of work on hybrid structure learning approaches that use constraint-based methods to initialize a partially oriented DAG for input to search-and-score methods. Dash and Druzdzel [8] apply PC to observational data and use the resulting CP-DAG as input to a greedy search that uses the standard likelihood score for DAGs. Tsamardinos et al. [25] similarly use PC in the first phase to identity an equivalence class of models given by the CP-DAG, and perform hill-climbing in the space of models to find best fitting DAG. In contrast, Schmidt et al. [21] prune the search space with L1-regularization based variable selection by fitting logistic regression models to the variables. We see multiple opportunities to extend our approach to be a hybrid method that further optimizes model structure

over additional score-based criteria beyond our initial set of constraints.

6 Discussion and Future Work

In this work, we formulate the problem of joint inference of causal structure from observational data and introduce a probabilistic approach that uses PSL framework to build on constraint-based structure discovery methodology. From our preliminary evaluation on the LUCAS synthetic dataset made available through the Causality Workbench, our PC-PSL approach enjoys improvements over PC in accuracy and F1 of causal edge prediction. In our future work, we plan to extensively study multiple synthetic and real datasets, including challenge problems posed on the Causality Workbench. We plan to extend our approach significantly by incorporating additional domain knowledge and variable selection techniques, fusing multiple sources of observational and experimental evidence, and more directly optimizing model structure by following hybrid approaches to searching and scoring.

7 Acknowledgements

We thank Sara Magliacane for the insightful and valuable discussions while formulating and developing this work.

References

1. S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. Hinge-loss Markov random fields and probabilistic soft logic. arXiv:1505.04406 [cs.LG], 2015.
2. David Maxwell Chickering. Learning bayesian networks is np-complete. In *Learning from data*, pages 121–130. Springer, 1996.
3. David Maxwell Chickering. Learning equivalence classes of bayesian-network structures. *The Journal of Machine Learning Research*, 2:445–498, 2002.
4. David Maxwell Chickering. Optimal structure identification with greedy search. *The Journal of Machine Learning Research*, 3:507–554, 2003.
5. David Maxwell Chickering, David Heckerman, and Christopher Meek. A bayesian approach to learning bayesian networks with local structure. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pages 80–89. Morgan Kaufmann Publishers Inc., 1997.
6. Tom Claassen and Tom Heskes. A logical characterization of constraint-based causal discovery. *arXiv preprint arXiv:1202.3711*, 2012.
7. Diego Colombo and Marloes H Maathuis. Order-independent constraint-based causal structure learning. *The Journal of Machine Learning Research*, 15(1): 3741–3782, 2014.
8. Denver Dash and Marek J Druzdzel. A hybrid anytime algorithm for the construction of causal models from sparse data. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 142–149. Morgan Kaufmann Publishers Inc., 1999.
9. Cassio P De Campos and Qiang Ji. Efficient structure learning of bayesian networks using constraints. *The Journal of Machine Learning Research*, 12:663–689, 2011.
10. Shobeir Fakhraei, Louiqa Raschid, and Lise Getoor. Drug-target interaction prediction for drug repurposing with probabilistic similarity logic. In *ACM SIGKDD 12th International Workshop on Data Mining in Bioinformatics (BIOKDD)*, 2013.
11. Shobeir Fakhraei, Bo Huang, Louiqa Raschid, and Lise Getoor. Network-based drug-target interaction prediction with probabilistic soft logic. *Comput. Biol. Bioinf.*, 11:775–787, 2014.
12. Shobeir Fakhraei, James Foulds, Madhusudana Shashanka, and Lise Getoor. Collective spammer detection in evolving multi-relational social networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1769–1778. ACM, 2015.
13. Nir Friedman and Moises Goldszmidt. Learning bayesian networks with local structure. In *Learning in graphical models*, pages 421–459. Springer, 1998.
14. Isabelle Guyon, Constantin F Aliferis, Gregory F Cooper, André Elisseeff, Jean-Philippe Pellet, Peter Spirtes, and Alexander R Statnikov. Design and analysis of the causation and prediction challenge. In *WCCI Causation and Prediction Challenge*, pages 1–33, 2008.
15. Antti Hyttinen, Patrik O Hoyer, Frederick Eberhardt, and Matti Jarvisalo. Discovering cyclic causal models with latent variables: A general sat-based procedure. *arXiv preprint arXiv:1309.6836*, 2013.
16. Tommi Jaakkola, David Sontag, Amir Globerson, and Marina Meila. Learning bayesian network structure using LP relaxations. In *International Conference on Artificial Intelligence and Statistics*, pages 358–365, 2010.
17. Pigi Kouki, Shobeir Fakhraei, James Foulds, Magdalini Eirinaki, and Lise Getoor. Hyper: A flexible and extensible probabilistic framework for hybrid recommender systems. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 99–106. ACM, 2015.
18. Jay Pujara, Hui Miao, Lise Getoor, and William Cohen. Knowledge graph identification. In *The Semantic Web–ISWC 2013*, pages 542–557. Springer, 2013.

19. Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daume III, and Lise Getoor. Learning latent engagement patterns of students in online courses. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.
20. Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine learning*, 62(1-2), 2006.
21. Mark Schmidt, Alexandru Niculescu-Mizil, Kevin Murphy, et al. Learning graphical model structure using l1-regularization paths. In *AAAI*, volume 7, pages 1278–1283, 2007.
22. Peter Spirtes. An anytime algorithm for causal inference. In *AISTATS*. Citeseer, 2001.
23. Peter Spirtes and Clark Glymour. An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, 9(1):62–72, 1991.
24. Peter Spirtes and Christopher Meek. Learning bayesian networks with discrete variables from data. In *KDD*, volume 1, pages 294–299, 1995.
25. Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65(1): 31–78, 2006.