

Can Causal Discovery Algorithms Help in Generating Legal Arguments?*

Soham Wasmatkar^{1,†}, Subinay Adhikary^{2,†}, Rakshit Rohan^{2,†}, Shouvik Kumar Guha³,
Saptarshi Pyne^{2,*} and Kripabandhu Ghosh^{2,*}

¹Manipal University Jaipur, Jaipur, Rajasthan, 303007, India

²Indian Institute of Science Education and Research (IISER) Kolkata, Mohanpur, 741246, India

³The West Bengal National University of Juridical Sciences, Kolkata, 700098, India

Abstract

In 2011, Judea Pearl received the Turing Award, considered the Nobel Prize in Computing, for *fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning*. It includes pioneering the development of *causal discovery* algorithms. These computer algorithms can analyze large multivariate datasets and automatically discover the causal relationships among the constituent variables. They have been widely used in many critical fields such as medicine and economics to support decisions. However, to our knowledge, they have not been leveraged in law. This paper attempts to alleviate this gap by investigating whether causal discovery algorithms can be leveraged for automated generation of legal arguments. To that end, a novel legal dataset is prepared by identifying 17 legal concepts, such as *physical assault* and *property dispute*. A curated collection of 150 homicide cases are annotated with these concepts, e.g., a case is annotated with *physical assault* only if a physical assault had been reported in that case. Subsequently, a selected set of widely-used causal discovery algorithms is applied to the annotated dataset to discover the causal relationships between the legal concepts. Additionally, the degrees of belief associated with the discovered relationships are quantified in mathematical probabilities. It is shown that some of the causal relationships help derive viable legal-argumentation patterns, e.g., if one could establish that a physical assault has not taken place during a homicide, it should be a sufficient condition (with probability 1) to establish that the homicide has not been committed due to a property-related dispute. Given the facts of a case, these argumentation patterns can be applied to automatically generate arguments specific to that case. Thus, this paper shows that causal discovery algorithms can be helpful in generating legal arguments, opening up avenues for promising future endeavors. Concurrently, a broader opportunity lies in investigating which other legal applications can benefit from the use of causal discovery algorithms.

Keywords

Argumentation, Bayesian Networks, Causal Reasoning, Causal Discovery, Explainable AI (XAI), AI and Law

1. Introduction

Causal discovery algorithms are computational methods for automating the discovery of causal relationships, among variables of interest, from large datasets [1, 2]. These algorithms are widely used in many critical fields, such as medicine, psychology, and economics, to support decisions [3, 4, 5]. However, their applications have not been leveraged in law, which is also a critical field. Instead, the legal literature only leverages the causal relationships discovered manually by legal experts [6, 7, 8, 9, 10, 11].

Causal discovery algorithms can augment the invaluable insights of the legal experts by substantially enhancing the scale (i.e. the number of prior cases that can be analyzed) and quantifying the degree of

* RELAF'26: Reasoning with Evidence in Law Enforcement and Forensics, Workshop at ICAIL 2026, Singapore.

*Corresponding author.

† Joint first authors. They contributed equally.

✉ soham.2430010342@mun.manipal.edu (S. Wasmatkar); sa21rs094@iiserkol.ac.in (S. Adhikary); rr24rs102@iiserkol.ac.in (R. Rohan); shouvikkumarguha@nujs.edu (S. K. Guha); saptarshipyne01@gmail.com (S. Pyne); kripa.ghosh@gmail.com (K. Ghosh)

🌐 <https://www.nujs.edu/faculty/dr-shouvik-kumar-guha> (S. K. Guha); <https://sap01.github.io> (S. Pyne);

<https://sites.google.com/view/kripabandhughosh-homepage> (K. Ghosh)

🆔 0009-0006-2165-2077 (S. Adhikary); 0009-0000-7725-8550 (R. Rohan); 0000-0003-1623-0444 (S. K. Guha); 0000-0001-9710-6749 (S. Pyne); 0000-0002-8130-1221 (K. Ghosh)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

belief for each causal relationship in a data-driven manner. For example, a causal discovery algorithm can be applied on a dataset that contains the files of all past homicide cases in the Supreme Court of India, the highest court of India. Suppose, the algorithm discovers a set of causal relationships between three variables as depicted in Figure 1.

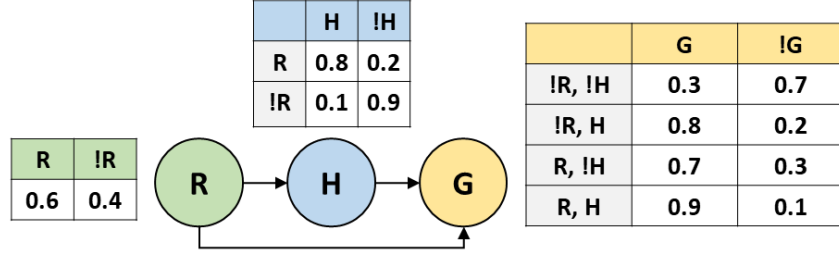


Figure 1: A Hypothetical Causal Graph between Three Variables—a Defendant being a Relative (R) of the victim, an heir (H) to the victim’s wealth, and found to be guilty (G). Each of these nodes are assumed to be binary variables, e.g., a defendant can either be a relative (R) of the victim or not (!R). Each directed edge represents a cause-and-effect relationship between two nodes, e.g., the edge $R \rightarrow H$ indicates that being a relative has a causal effect on being an heir. In this paper, the degree of belief in the causal effect is measured as a “conditional probability table” (CPT). Each node is annotated with a CPT where the states of the node constitute the columns and the states of its cause(s), if any, constitute the rows. For example, in the CPT of node “H”, the value 0.8 corresponding to row “R” and column “H” implies that, among the prior cases where the defendants were relatives of the victims, in 80% of those cases, the defendants were also heirs.

Subsequently, legal experts can examine whether the causal relationships align with their insights and make refinements, if necessary. The refined causal graph can be used for automatically generating quantitative arguments such as “If you can establish that the defendant is not an heir to the victim’s wealth, it increases the chance of proving the defendant not guilty by 59%.” The value of 59% comes from calculating the conditional probabilities $P(\text{not guilty} \mid \text{not an heir})$ (denoted by $P(!G|!H)$) and $P(\text{not guilty} \mid \text{an heir})$ (denoted by $P(!G|H)$), and computing their difference. $P(!G|!H)$ can be calculated as $P(!G|!H) = P(!G, !H) / P(!H)$ [By Bayes’ theorem]
 $= (P(!G, R, !H) + P(!G, !R, !H)) / P(!H)$ [By the law of total probability]
 $= P(!G|R, !H) * P(!H|R) + P(!G|!R, !H) * P(!H|!R)$ [By Bayes’ theorem]
 $= 0.3 * 0.2 + 0.7 * 0.9 = 0.69$. Similarly, $P(!G|H)$ calculates to 0.1. Their difference is $(0.69 - 0.1) = 0.59$ which is the improvement in the probability if it can be established that the defendant is not an heir. Thus, the quantification of the arguments generated with the help of causal discovery algorithms are completely explainable in legal and mathematical terms. Moreover, the accuracy of the quantified values depends on the magnitude of prior cases a causal discovery algorithm has analyzed, which for a computer algorithm can be all related cases in human history.

The purpose of this paper is to employ some of the widely-used causal discovery algorithms on a moderately large number of legal cases and investigate whether the resultant causal graphs can help generate legal arguments. To that end, a novel dataset consisting of 150 homicide cases is prepared in Section 2. Six widely-used causal discovery algorithms are applied on the dataset and the resultant causal graphs are examined to understand what kind of arguments can be generated from them (Section 3). The conclusions are drawn in Section 4 along with a discussion on the future directions.

2. Methodology

The preparation of the novel dataset and selection of the algorithms are described in this section.

2.1. Dataset Preparation

Identification of Legal Concepts Adhikary et al. [12] identify seven legal concepts such as “life imprisonment” and “physical assault” (Table 1). Subsequently, they annotate 200 homicide cases with

Table 1

The Legal Concepts Used for Annotating and Analyzing Homicide Cases. The concepts placed above the dashed line are identified by Adhikary et al. [12] and the rest are identified by the authors of this paper.

Legal Concept	Description
Life Imprisonment	Court punished the accused with life imprisonment for murder.
Physical Assault	Accused killed, murdered, or seriously hurt a person using a sharp weapon or firearm.
Homicide Murder	Court observed that the accused committed murder deliberately.
Evidence Inconsistency	Proper or consistent evidence of the crime was not found in the case.
Witness Testimony	Witness testimony was collected from non-expert people (e.g., a common person).
Expert Witness Testimony	Witness testimony collected from forensic, ballistic, or other subject-matter experts.
Riot	A riot was reported.
-----	-----
Political Rivalry	Incident was connected to political issues, resulting in injury or murder.
Homicide Not Murder	Court declared the offense as culpable homicide not amounting to murder.
Revenge	Murder committed as an act of vengeance.
Property Dispute	Murder committed due to a property-related dispute.
Testimony Challenged	Witness testimony presented by prosecution or defence was contested and evaluated by the court.
Prosecutorial Delay or Inability	Prosecutorial delay or inability was recorded.
Investigation Agency	Presence of investigative agencies other than the Police.
Evidence Insufficient	Insufficiency of evidence was noted during the judgment.
Rarest of the Rare Case	Court declared the case as a “rarest of the rare” case.
Death Sentence	A death penalty was handed down by the court.

these concepts, e.g., a case is annotated with the concept “life imprisonment” only if the court has punished the accused with life imprisonment. Ten new legal concepts are identified in this paper (Table 1). Among the aforementioned 200 case files, 150 are randomly selected for annotation with these $(7+10) = 17$ legal concepts.

Annotation For the annotation task, two legal experts having expertise in crime-related documents are hired. Given a case file, they are asked to annotate spans of text with one or more legal concepts using the legal annotation tool LeDA [13] (Figure 2). Accordingly, they annotate the selected 150 case files with the 17 legal concepts. The inter-annotator agreement (IAA) is measured using the F1-score, instead of the kappa coefficient, as recommended by Wyner et al. [14]. The IAA score depends on three types of agreements: **Correct**: when two annotators mark exactly the same text span with the same concept, **Partial**: when both annotators assign the same concept but select different text spans, and **Missing**: when the annotators assign different concepts for the same text span. For each case file, the IAA score is calculated with respect to a pair of annotators (Figure 3). If there are more than two annotators, the pairwise IAA scores are averaged to obtain the final IAA score of the case. The case files with IAA scores below 0.8 are reviewed by a senior legal expert (co-author Dr. Guha), who finalize the annotations. Based on the annotations, we derive a binary data matrix encoding the presence (1) or absence (0) of the concepts in each case (Figure 4). The data matrix is used as the input to the causal discovery algorithms.

2.2. Selection of Algorithms

Six widely-used causal discovery algorithms are selected based on their methodological diversity, in order to ensure that they do not share the same biases. The algorithms are PC [15], GES [16], GRaSP [17], BOSS [18], LiNGAM [19], and ANM [20]. PC constructs the causal graph by adding one causal relationship at a time. It does so by performing a “conditional independence” test on each potential relationship and adding the relationship only if it passes the test. GES, on the other hand, proactively adds a relationship and examines whether this addition increases the likelihood-based

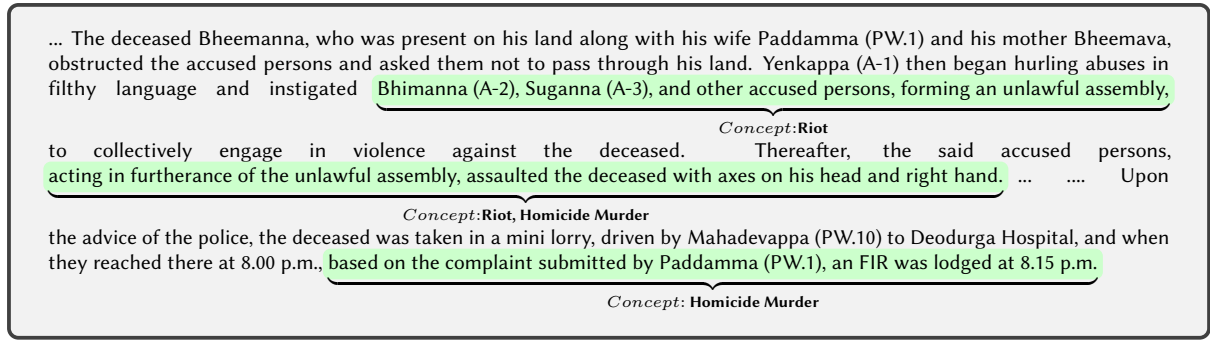


Figure 2: Examples of How Text Spans Within or Across Sentences are Annotated with Legal Concepts.

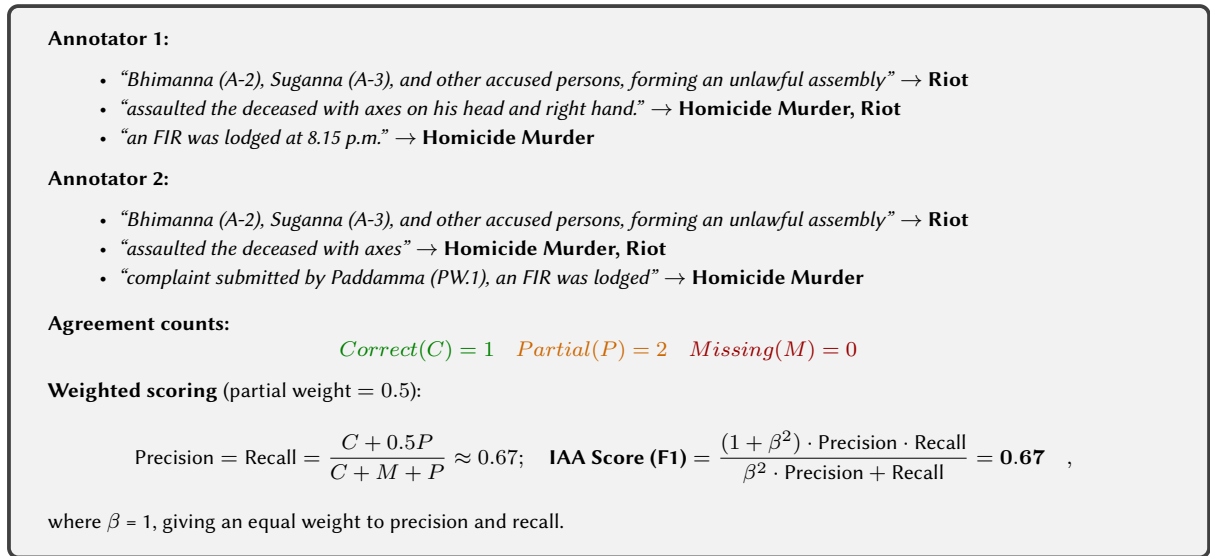


Figure 3: An Example of How the Inter-Annotator Agreement (IAA) Score is Calculated.

“score” of the whole graph. The relationship is removed only if it does not increase the score. GRASP and BOSS start with the variables (here, legal concepts) arranged in a random order (suppose, “R H G” for the Figure 1 example). They construct all causal graphs that can be represented by this order s.t. for every causal relationship $X \rightarrow Y$ in the causal graph, X should appear before Y in the given order, e.g., the causal graph in Figure 1 is one of the valid causal graphs for the order “R H G”. The algorithms calculate different permutations of the given order (e.g., “H R G”), calculates the scores of all causal graphs for each permuted order, and selects a causal graph with the maximal score. LiNGAM and ANM formulate the causal discovery task as linear and non-linear regression problems, respectively, where a legal concept can be modeled as the target variable and its potential causes i.e. the remaining legal concepts can be modeled as predictor variables.

3. Results

Each of the six causal discovery algorithms outputs a causal graph that represents the causal relationships discovered by that algorithm. To figure out which causal relationships are true causal patterns hidden in the dataset and not sensitive to the choice of an algorithm, a “consensus causal graph” is constructed (Figure 5). It consists of the causal relationships that are independently discovered by more than one algorithm. Each relationship is assigned an integer weight (e.g., 3) which represents how many algorithms agree on the relationship.

There are six causal relationships in the consensus causal graph and all of them are agreed on by half

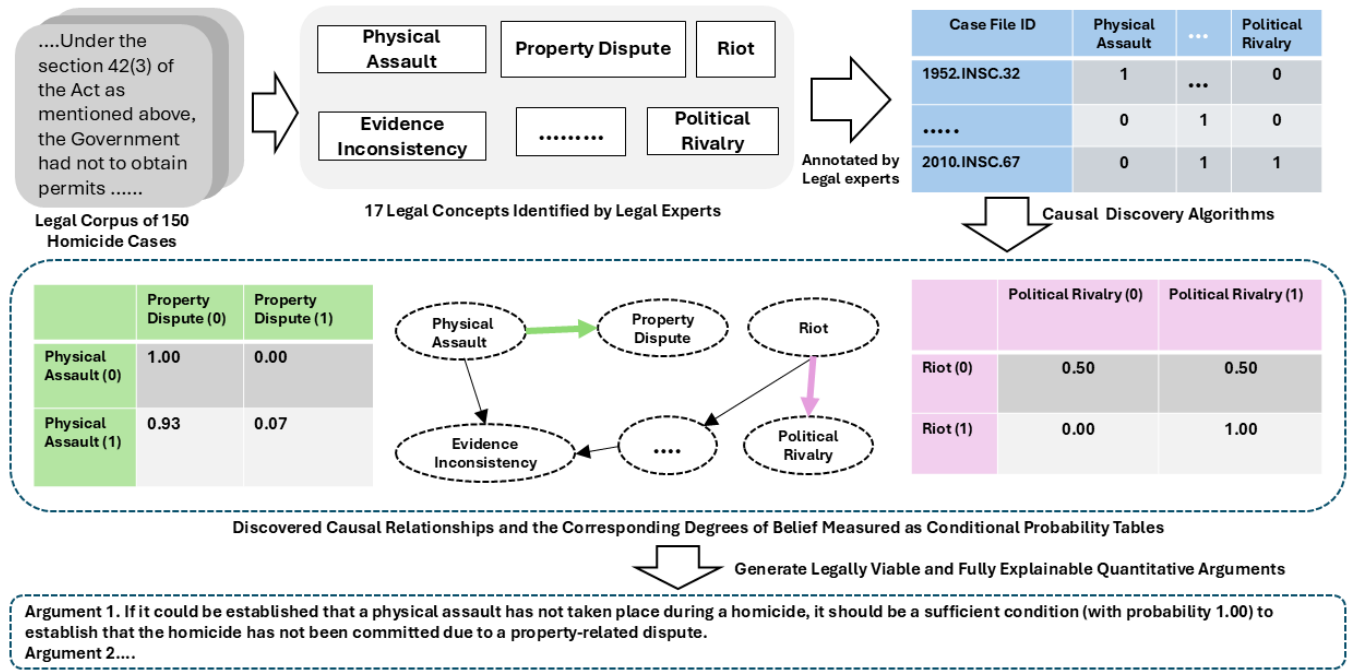


Figure 4: The Causal Discovery Algorithm-Aided Workflow for Automated Generation of Legal Arguments. A curated legal corpus of 150 homicide cases are utilized to identify 17 legal concepts. Subsequently, each case is annotated with a legal concept only if the concept is applicable to the case. Based on the presence (0) or absence (1) of the 17 legal concepts in 150 homicide cases, a 150-by-17 binary data matrix is prepared. The matrix is inputted to an array of widely-used causal discovery algorithms for discovering causal relationships among the legal concepts. The degrees of belief associated with the discovered causal relationships are measured as conditional probability tables (only two tables are shown). Finally, the causal relationships along with the associated conditional probability tables are used for generating legal arguments.

or more of the algorithms, indicating how robust the relationships are. They cater to four “effect” nodes: political rivalry (3 causal relationships) and testimony challenged, rarest of the rare case, property dispute (1 causal relationship each). The degrees of belief in their discovered causes are measured in CPTs. Some of them are examined below.

The Causal Relationships of Political Rivalry Political rivalry has three “predictive causes”—evidence inconsistency, prosecutorial delay or inability, and riot (see node N10 in Figure 5). They are called predictive causes since the combinations of their values can predict what value political rivalry would take [21]. These are different from the true cause of political rivalry, which involves sociopolitical history. Three algorithms, namely, PC, GGrASP, and BOSS, agree on all three of these causes (Table 2) while the GES algorithm agrees on only two of these causes (Table 3). From Table 2, it can be seen that, for all the prior cases where consistent evidence was not found ($N_8 = 1$) and riots were reported ($N_4 = 1$), the homicides were connected to political rivalry. In other words, if consistent evidence is not found and a riot is reported in a homicide case, it can be argued that the homicide(s) is connected to political rivalry with certainty (probability = 1.0). For all other combinations of their values, the connection to political rivalry is either non-existent (probability = 0.0) or equivalent to a random chance (probability = 0.5). Although, the value of prosecutorial delay or inability (N_2) is irrelevant for reaching a certainty, it plays a crucial role in increasing the probability from non-existence to a random chance. GES, being a local optimization algorithm, could not capture the role of prosecutorial delay or inability in determining political rivalry. However, it does agree on the roles of evidence inconsistency and riot in ascertaining political rivalry (Table 3).

The proposed framework is substantially different from those based on pairwise correlation measures, such as the Pearson correlation coefficient [22]. Firstly, the causal relationships are directed. Secondly,

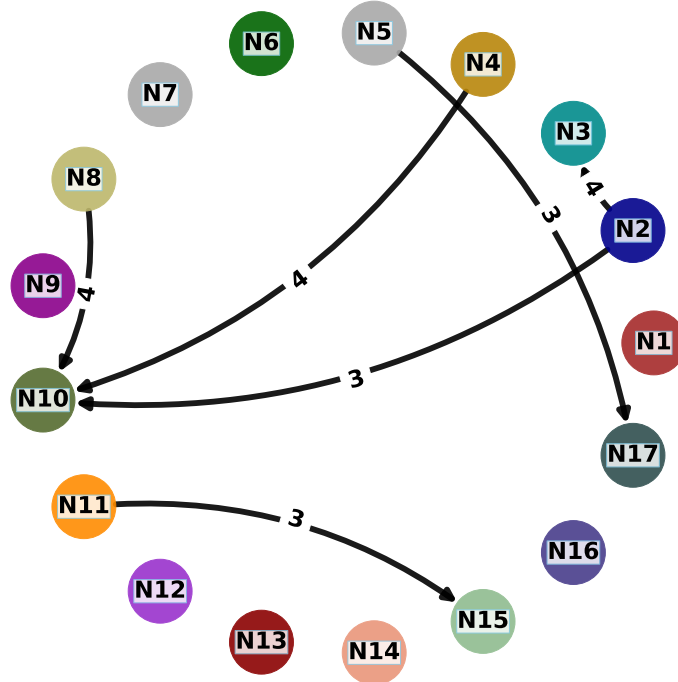


Figure 5: The Consensus Causal Graph. The nodes represent the legal concepts. *N1*: witness testimony, *N2*: prosecutorial delay or inability, *N3*: testimony challenged, *N4*: riot, *N5*: death sentence, *N6*: life imprisonment, *N7*: homicide murder, *N8*: evidence inconsistency, *N9*: expert witness testimony, *N10*: political rivalry, *N11*: physical assault, *N12*: evidence insufficient, *N13*: investigation agency, *N14*: revenge, *N15*: property dispute, *N16*: homicide not murder, and *N17*: rarest of the rare case. The edges represent the causal relationships agreed upon by multiple algorithms. An edge weight (e.g., 3 or 4) implies the count of algorithms that agree on the edge.

it examines a fine-grained association between a distinct value-combination of potential causes and a distinct value of the effect. Even though there are only two cases involving political rivalry (Column “*N10* = 1”, Table 2) in the dataset, it could detect that evidence inconsistencies and riots are simultaneously reported in exactly those cases. Pearson correlation, on the other hand, measures the association between political rivalry and a single potential cause, which aggregates over their different combinations of values. Since there are much larger number of cases involving evidence inconsistencies (12+1+0+1 = 14 cases) and riots (9+1+1+1 = 12 cases), their Pearson correlations with political rivalry are reported to be low, 0.36 and 0.39, respectively. Hence, these predictive causes could have never been discovered using such correlation measures.

The Causal Relationship between Physical Assault and Property Dispute Half of the causal discovery algorithms agree that there exists a causal relationship between physical assaults (*N11*) and motives related to property disputes (*N15*) in homicide cases (Figure 5). They concur that establishing a physical assault has not taken place during a homicide should be a sufficient condition (with probability = 1.00) to establish that the homicide has not been committed due to a property-related dispute (Table 4). Such a strong causation could not have been captured from their extremely low Pearson correlation (0.2). On the other hand, establishing that a physical assault has taken place does not suffice to establish that the homicide is related to a property dispute. It only increases the chance by 7% (from a probability value of 0.00 to 0.07). The causal factors that can explain the remaining 93% of uncertainty in establishing a motive related to property dispute might be beyond the seventeen legal concepts considered in this paper. Identifying more legal concepts and annotating prior cases with them can help causal discovery algorithms mitigate this gap. The dataset and code used in this section can be found at <https://github.com/spynegroup/LegalCausation>.

Table 2

The Conditional Probability Table of “Political Rivalry” in the Causal Graphs Separately Discovered by the PC, GRaSP, and BOSS Algorithms. The rows where the causes predict the presence of political rivalry ($N_{10} = 1$) with certainty i.e. with a probability of 1.0 are boldfaced. The number of cases corresponding to each probability value is given in parentheses.

Causes	Effect		
Evidence Inconsistency (N8), Prosecutorial Delay or Inability (N2), Riot (N4)	Political Rivalry (N10) = 0	N10 = 1	Total No. of Cases for Each Combination of N8,N2,N4
0, 0, 0	1.0 (125)	0.0 (0)	(125)
0, 0, 1	1.0 (9)	0.0 (0)	(9)
0, 1, 0	1.0 (1)	0.0 (0)	(1)
0, 1, 1	1.0 (1)	0.0 (0)	(1)
1, 0, 0	1.0 (12)	0.0 (0)	(12)
1, 0, 1	0.0 (0)	1.0 (1)	(1)
1, 1, 0	0.5 (0)	0.5 (0)	(0)
1, 1, 1	0.0 (0)	1.0 (1)	(1)
Total no. of cases for each value of N10	(148)	(2)	(150)

Table 3

The Conditional Probability Table of “Political Rivalry” in the Causal Graph Discovered by the GES Algorithm. The row where the causes predict the presence of political rivalry ($N_{10} = 1$) with certainty (probability = 1.0) is boldfaced.

Causes	Effect	
Evidence Inconsistency (N8), Riot (N4)	Political Rivalry (N10) = 0	N10 = 1
0, 0	1.0	0.0
0, 1	1.0	0.0
1, 0	1.0	0.0
1, 1	0.0	1.0

Table 4

The Conditional Probability Table of “Property Dispute” in the Consensus Causal Graph (Figure 5)

Cause	Effect		
Physical Assault (N11)	Property Dispute (N15) = 0	N15 = 1	Total No. of Cases for Each Value of N11
0	1.00 (80)	0.00 (0)	(80)
1	0.93 (65)	0.07 (5)	(70)
Total no. of cases for each value of N15	(145)	(5)	(150)

4. Conclusions and Future Work

This paper asks the question, “Can causal discovery algorithms help in generating legal arguments?”. To answer the question, a novel legal dataset is prepared by annotating 150 prior homicide cases with 17 legal concepts such as the *life imprisonment*. A case is annotated with *life imprisonment* only if the court imposed life imprisonment on the accused in that case. The annotated case files are separately inputted to six widely-used causal discovery algorithms. The algorithms are specifically selected to represent different sets of assumptions and methodological paradigms. Hence, it is expected that they do not share the same biases. Each of these algorithms discovers a distinct set of causal relationships. To identify robust causal relationships, only those that are discovered by more than one algorithm

are considered for further analysis. Many of these causal relationships are found to be useful for generating argumentation patterns that are quantifiable, fully explainable, and legally viable. For example, a consensus is achieved by half of the algorithms that establishing a physical assault has not taken place during a homicide should be a sufficient condition (with probability 1) to establish that the homicide has not been committed due to a property-related dispute. Given the facts of a case, these argumentation patterns can be applied to automatically generate arguments specific to that case.

Therefore, future directions lie in answering the question “How can we maximize the utilization of causal discovery algorithms for automated legal argument generation?”. Increasing the number of legal concepts and annotating thousands of cases with them can be a mammoth task; however, it can enhance the quality and quantity of legal arguments produced with the help of causal discovery algorithms. Large language models [23] can be utilized to make this mammoth annotation task feasible.

At the same time, more advanced levels of causal reasoning techniques can be applied on the discovered causal relationships to generate more nuanced and persuasive legal arguments. The causal reasoning technique applied in this paper is known as *association* based reasoning which forms the lowest level in *the ladder of causation* [24]. The highest level is formed by the *counterfactual* reasoning. A counterfactual is a scenario that did not happen but could have happened. For example, if person “X” keeps a couch in front of a fire exit and person “Y” dies in a fire, is X legally responsible for Y’s death? It necessitates answering the causal question of whether the couch blocking the fire exit is a “but-for” cause of Y’s death [25]; the answer lies in calculating the counterfactual probability that Y would have survived if the couch had not been blocking the fire exit. If the probability is reasonably high, one could argue that X should be held legally responsible for Y’s death. In 2011, Judea Pearl received the *Turing Award*, considered the Nobel Prize in Computing, for “fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning” [26] which includes the development of a theoretical framework for counterfactual reasoning in conjunction with causal discovery. Pearl is profoundly optimistic about its applications in law and believes “In principle, counterfactuals should find easy application in the courtroom” [24].

For accused persons, witnesses, and victims, causal reasoning provides transparency about how decisions affecting their liberty and interests are made. For victims’ families in death sentence cases, understanding the causal structure may help them evaluate whether the sentence reflects the crime’s severity or whether procedural factors such as prosecutorial inability were determinative.

The causal discovery algorithms can be used to automatically discover a causal structure from prior related cases in a fast and fully transparent manner. This paper envisages a pipeline where the automatically discovered causal structure is refined through the human supervision of legal experts. The refined causal structure can help the litigating parties understand the causal reasoning behind the judgment affecting their lives and, if required, appeal if the said causal reasoning diverges from the refined causal structure. Moreover, many causal discovery algorithms can incorporate expert knowledge to improve efficiency and reduce false discovery [27]. This feature can be utilized to incorporate causal relationships known a priori such as the ones specified by legal doctrines. At the same time, expert knowledge can be leveraged to impose legal constraints on possible edges i.e. to specify which causal relationships are not expected. One may therefore opine that incorporating the usage of causal discovery algorithms in legal analysis represents not merely a technical upgrade, but a conceptual advancement toward law’s fundamental commitment to timely justice and transparency.

Acknowledgments

Thanks to the legal experts who were involved in the annotation process, namely, Ms. Bipasha Kundu and Ms. Anjana Raghunath. We are also indebted to the anonymous reviewers for their invaluable

comments.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] A. Zanga, E. Ozkirimli, F. Stella, A survey on causal discovery: theory and practice, *International Journal of Approximate Reasoning* 151 (2022) 101–129.
- [2] C. K. Assaad, E. Devijver, E. Gaussier, Survey and evaluation of causal discovery methods for time series, *Journal of Artificial Intelligence Research* 73 (2022) 767–819.
- [3] X. Shen, S. Ma, P. Vemuri, G. Simon, Challenges and opportunities with causal discovery algorithms: application to alzheimer’s pathophysiology, *Scientific reports* 10 (2020) 2975.
- [4] J. J. Anker, E. Kummerfeld, A. Rix, S. J. Burwell, M. G. Kushner, Causal network modeling of the determinants of drinking behavior in comorbid alcohol use and anxiety disorder, *Alcoholism: clinical and experimental research* 43 (2019) 91–97.
- [5] P. M. Addo, C. Manibialoa, F. McIsaac, Exploring nonlinearity on the co2 emissions, economic production and energy use nexus: A causal discovery approach, *Energy Reports* 7 (2021) 6196–6204.
- [6] R. Liepiņa, T. de Lima, E. Lorini, G. Pisano, G. Sartor, A causal model checker for legal cases, in: *Proceedings of the Twentieth International Conference on Artificial Intelligence and Law*, 2025, pp. 248–257.
- [7] A. P. Dawid, F. Dotto, M. Graves, J. B. Kadane, J. Mortera, G. Robertson, J. Q. Smith, A. L. Wilson, A comparison of graphical methods using the case of the murder of meredith kercher as an example, *Law, Probability and Risk* 24 (2025) mgaf002.
- [8] C. Dahlman, E. Kolflaath, Causal models versus reason models in bayesian networks for legal evidence, *Synthese* 200 (2022) 477.
- [9] A. Liefgreen, D. Lagnado, The role of causal models in evaluating simple and complex legal explanations, in: *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 43, 2021.
- [10] R. Liepiņa, G. Sartor, A. Wyner, Arguing about causes in law: a semi-formal framework for causal arguments, *Artificial intelligence and law* 28 (2020) 69–89.
- [11] R. Liepina, G. Sartor, A. Wyner, Causal models of legal cases, in: *International Workshop on AI Approaches to the Complexity of Legal Systems*, Springer, 2015, pp. 172–186.
- [12] S. Adhikary, P. Sen, D. Roy, K. Ghosh, A case study for automated attribute extraction from legal documents using large language models, *Artificial Intelligence and Law* (2024) 1–22.
- [13] S. Adhikary, D. Roy, D. Ganguly, S. Kumar Guha, K. Ghosh, Leda: a system for legal data annotation, *Frontiers in Artificial Intelligence and Applications* (2023) 370–367.
- [14] A. Z. Wyner, W. Peters, D. Katz, A case study on legal case annotation., in: *JURIX*, 2013, pp. 165–174.
- [15] P. Spirtes, C. N. Glymour, R. Scheines, *Causation, prediction, and search*, MIT press, 2000.
- [16] D. M. Chickering, Optimal structure identification with greedy search, *Journal of machine learning research* 3 (2002) 507–554.
- [17] W.-Y. Lam, B. Andrews, J. Ramsey, Greedy relaxations of the sparsest permutation algorithm, in: *Uncertainty in Artificial Intelligence*, PMLR, 2022, pp. 1052–1062.
- [18] B. Andrews, J. Ramsey, R. Sanchez Romero, J. Camchong, E. Kummerfeld, Fast scalable and accurate discovery of dags using the best order score search and grow shrink trees, *Advances in neural information processing systems* 36 (2023) 63945–63956.
- [19] S. Shimizu, P. O. Hoyer, A. Hyvärinen, A. Kerminen, M. Jordan, A linear non-gaussian acyclic model for causal discovery., *Journal of Machine Learning Research* 7 (2006).

- [20] P. Hoyer, D. Janzing, J. M. Mooij, J. Peters, B. Schölkopf, Nonlinear causal discovery with additive noise models, *Advances in neural information processing systems* 21 (2008).
- [21] F. X. Diebold, *Elements of forecasting*, South-Western College Pub. Cincinnati, OH, USA, 1998.
- [22] J. Lee Rodgers, W. A. Nicewander, Thirteen ways to look at the correlation coefficient, *The American Statistician* 42 (1988) 59–66.
- [23] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Akhtar, N. Barnes, A. Mian, A comprehensive overview of large language models, *ACM Trans. Intell. Syst. Technol.* 16 (2025). URL: <https://doi.org/10.1145/3744746>.
- [24] J. Pearl, D. Mackenzie, *The book of why: the new science of cause and effect*, Basic books, 2018.
- [25] B. A. Spellman, A. Kincannon, The relation between counterfactual (but for) and causal reasoning: Experimental findings and implications for jurors' decisions, *Law and Contemp. Probs.* 64 (2001) 241.
- [26] S. J. Russell, Judea pearl, 2011. URL: https://amturing.acm.org/award_winners/pearl_2658896.cfm.
- [27] Y. Yu, L. Hou, X. Liu, S. Wu, H. Li, F. Xue, A novel constraint-based structure learning algorithm using marginal causal prior knowledge, *Scientific Reports* 14 (2024) 19279.