View Reviews

Paper ID
6765

Paper Title
The Complexity of Morality: Checking Markov Blanket Consistency with DAGs via Morality

Reviewer #1

Questions

1. Contributions: Please list three things this paper contributes (e.g., theoretical, methodological, algorithmic, empirical contributions; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.
The discuss the complexity of verifying morality of an undirected graph, i.e. of determining if a graph can be obtained as the moral graph of a directed acyclic graph. They show an efficient algorithm for verifying morality for graphs with degree at most 4, and adapt a previous proof by Verma and Pearl to show that the problem is NP-hard for graphs of degree 5 or more.

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance. Hover over the "?" next to this prompt to see a brief description of these metrics.
The proofs are sound and the result is theoretically interesting. I could not follow all the last step in Lemma 4.1, but the result seems correct.

The authors motivate the problem as one of deciding if a set of Markov Blankets induced from data are consistent with some Bayes network. They then discuss in the conclusion that their algorithm could be somehow used to find an approximate DAG when the graph is not moral. This is one direction; other direction would be to simply recognize that an undirected graph is a better representation of the data. This is somehow the spirit of Verma and Pearl's original work: to decide if a graph of associations can be represented as a DAG interpreted through d-separation.

In Sec 2, please state that E'can be empty (I took me several pages to realize that this was the case, and hence to make sense of some results). Having this, it is e.g. immediate that chordal graphs are WRS.

As Algorithm 2 is the main contribution of this work, it would be better to find some space to have in the paper, in case it gets through. Algorithm 1 can easily be explained in text (so there is no need to mention it).

Edit after rebuttal and discussion: I believe this is a good paper, that touches on a interesting theoretical problem and establishes its complexity. While the proofs are somewhat simple, the overall results is important. That said, I was convinced that the paper, as it is, it is not suited for NeurIPS, as it is only weakly motivated towards effective learning DAGs.

3. Please provide an "overall score" for this submission.
6: Marginally above the acceptance threshold. I tend to vote for accepting this submission, but rejecting it would not be that bad.

4. Please provide a "confidence score" for your assessment of this submission.
4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

5. Improvements: What would the authors have to do for you to increase your score?

The work could have greater impact if the authors could show that MBs learned from real dataset violate morality, and to what consequences this leads one.

Reviewer #2

Questions

1. Contributions: Please list three things this paper contributes (e.g., theoretical, methodological, algorithmic, empirical contributions; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.
The paper shows interesting properties of DAGs using the notion of weakly recursively simplicial. significance(medium)
The paper is good written with a lot of examples. significance(medium)

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance. Hover over the "?" next to this prompt to see a brief description of these metrics.
This paper proves many interesting properties of MECs and proposes a polynomial time MEC consistency check algorithm based on their theoretical findings.

It would be more convincing if the paper introduce some algorithms that heavily relies on the MEC. My concerns are the following:
The MMHC algorithm depends on the skeleton of the DAG rather than the MECs. The ODS algorithm in Park and Raskutti (2018) or Gaussian SEM learning algorithm, Ghoshaland Honorio (2018) requires a (reasonably large) super-set of the skeleton, instead of MECs. The PC algorithm has a process of checking cycles or Meek's conjectures. Hence, it would be better to show the proposed method actually improves the performance of the moralized graph based - DAG learning algorithms.

Park, Gunwoong, and Garvesh Raskutti. "Learning quadratic variance function (qvf) dag models via overdispersion scoring (ods)." The Journal of Machine Learning Research 18.1 (2017): 8300-8342.
Ghoshal, Asish, and Jean Honorio. "Learning identifiable gaussian bayesian networks in polynomial time and sample complexity." Advances in Neural Information Processing Systems. 2017.

Some notations are unclear. I agree that many papers abuse the notations for ease of presentations, this paper still needs to clarify some notations.
For example,
(1) "Throughout, we assume each BN is faithful, that is, its joint probability distribution and the directed acyclic graph (DAG) entail exactly the same conditional independencies."
It would be "d-separations \iif conditional independences" In addition, this paper may not need faithfulness assumption.
(2) V is a vertex set. not a set of random variables in general. I know it would be minor, however I hope this paper emphasizes that any concept of randomness or probability is not involved in this paper.
(3) In order to under the main idea of simplical vertext, some notations should be clarified such as complete subgraph, u complege, N_G(u). imperfect data.

The example in Figure 1 may not be a good choice for delivering the main idea of this paper. As you know, Figure 1 (c) is a well-known impossible moral graph because 1-2-3-4-1 without a cycle is impossible in a DAG settings. In such case, why do we need to apply the proposed method? I recommend to provide a different example that is not a trivial non-moral graph. Unfortunately, I cannot provide any specific papers, however you can find the these results in learning cyclic graphical models that does not have a sink or terminal node.

I think the conditions on the maximum degree (<=4) is too strong. For example of 5-node fully identifiable graph, 1->3<-2, 3->5<-4, the proposed algorithm cannot be applied because the maximum degree is 5. I agree that the consistency check of moral graph is NP hard problem, it is still not a big problem for small graphs p < 6. It would be better to generalized the algorithm for arbitrary k-maximum degree even when k is known.
I have read the author response and other reviewers' comments. After a short discussion, I am going to increase my score to 5 owing to the theoretical importance of the paper.

3. Please provide an "overall score" for this submission.
5: Marginally below the acceptance threshold. I tend to vote for rejecting this submission, but accepting it would not be that bad.

4. Please provide a "confidence score" for your assessment of this submission.
3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

5. Improvements: What would the authors have to do for you to increase your score?
I am willing to increase my score if (1) the proposed method is actually applied to the existing algorithms and (2) the proposed method is more generalized for arbitrary k-maximum degree.

Reviewer #4

Questions

1. Contributions: Please list three things this paper contributes (e.g., theoretical, methodological, algorithmic, empirical contributions; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.
The paper studies the complexity of checking whether an undirected, connected graph is moral, that is, whether the graph can be obtained as the moralization of a DAG.

The three main points of the paper:

1) The authors present an alternative, equivalent characterization for a graph being moral, namely that of being weakly recursively simplicial
2) The problem is polynomial-time solvable in maximum degree 4 graphs
3) The problem is NP-complete for graphs with maximum degree 5

Together, these results paint a very clear picture of the complexity of morality checking. From a practical standpoint the results are of limited interest, as the cases for which the problem is polynomial are quite restricted and unlikely to occur in practice. However, from a theoretical standpoint this completely settles the question of the importance of bounded degree and so provides very clear indication of where (not) to look for improvements.

The alternative characterization could be quite useful for a number of theoretical and practical applications. It is a very clean characterization.

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance. Hover over the "?" next to this prompt to see a brief description of these metrics.
Originality/significance:
- The NP-hardness proof is a very simple modification of an NP-hardness proof by Verma and Pearl, which creates a graph which has a single vertex of degree >5. This vertex is replaced by a long path. This progress is very incremental. It would be more honest to state in the main text what the scope of the modifications required is (and not just in the supplementary material).

- The equivalence of WRS to being moral is quite easy to see, but was not known before so this is an elegant result which may be quite useful.
- The algorithm for the maximum degree 4 case follows from quite an intricate case analysis. However, I do not think that many practical applications will have the property of bounded degree.

The significance is further discussed in field (1).

Quality/clarity:
The paper contains a few statements about complexity that are known to be incorrect, and these would have to be rectified to make the paper acceptable for publication. However, these are inconsequential to the main results and can easily be fixed. Specifically:

1) The introduction of section 4 states “the number of edges between a simplicial vertex’s neighbours can grow exponentially with the degree of the vertex”. -> the number of edges between the neighbours is at most quadratic in the degree, not exponential.
Moreover, the authors seems to erroneously equate being NP-complete with having exponential time complexity. This is incorrect, as some NP-complete problems are known to have sub-exponential (but super-polynomial) algorithms.

2) Section 4.3. “However, at some point, the problem must become NP-complete to agree with the general result.” This is not necessarily true, as the problem might be polynomial-time solvable for any fixed degree. For example, maximum clique is NP-complete in general, but polynomial-time (even linear-time) solvable for any graph of bounded maximum degree.

The writing contains many spelling and grammar mistakes, and there are further technical improvements that could be made, as explained in the detailed comments below. Overall though, the paper is quite clear and easy to read.

Detailed comments

The consistency property guarantees that there exists at least one […] -> is this the definition of the consistency property, or is this a consequence of the consistency property? Perhaps replace “guarantees that there exists” with “states that there must exist”

In the proof of lemma 3.2, please make explicit where (adjacent to which vertices) the sink is added.

Lemma 3.3. In the introduction, you state that you consider only simple connected graphs. However, G-u need not be connected. I do not think this is a consequential error, but the statement that you consider only connected graphs seems false.

You use (in the proof of Lemma 4.2) both the notation \subseteq, \subset and \subsetneq. What does \subset mean? Is it the same as \subseteq, or the same as \subsetneq, or something else entirely?

“In addition, their common neighbours [formula] implies [formula]” -> how does a set of neighbours imply anything? Perhaps you should write “In addition, the fact that for their common neighbours it holds that [formula], implies […]”

Theorem 4.1 it seems that the running time could easily be improved to O(n), by initially checking for each vertex whether it is simplicial and keeping a queue of simplicial vertices. As each vertex deletion affects the simpliciality at most a constant number of neighbours, the queue can be updated in constant amortized time (if a vertex becomes simplicial due to the deletion of a vertex it can be added on the queue, and if a vertex loses it simplicial status it can be skipped when erroneously extracted from the queue). Note that in the supplementary material the running time is stated to be O(n^3), but a stronger statement could be made as the main text proves O(n^2).

Theorem 4.2 is a strictly stronger version of theorem 4.1. Perhaps you could state the running time in the theorem (rather than merely “polynomial”), which would make the strengths of the theorems incomparable.

Proof of lemma 4.4. I do not understand why we can assume (wlg) that every neighbor of x has degree 3.
Does G always remain connected under the operations in Algorithm 2? Otherwise, the algorithm, as stated, would be incorrect (but could easily be fixed).

The running time of Algorithm 2 could be improved to O(n^2) by using BFS in place of Dijkstra to compute the distance.

Experimentally testing the running time of an algorithm on a graph of size 100 is highly misleading. Even a sufficiently slowly growing exponential function could look linear on that scale. Some argument could be made that testing graphs of size up to 100 gives some confidence for the correctness, but for the running time much larger samples should be used. Especially considering that the theoretical analysis shows the running time is quadratic or even cubic, seeing linear behavior in practice suggests the used test cases are not triggering the worst case. Moreover, the theoretical analysis of the running time is obviously correct, so I see little added value in the experiments. For the correctness analysis, do the random inputs have a reasonable chance of giving a “yes”-instance?

Spelling & grammar
Abstract: without relying on Bayesian networkS
Introduction: has became -> has become
Page 2, last sentence of second paragraph: the sentence is hard to read due to three sentences being connected by “if and only if” without any punctuation. I suggest “if and only if the nbhs are consistent with a DAG, which is the case if and only if […]”
“Last but not the least” -> “Last but not least”
“shortlist a few important definitions” -> “shortlist” is a strange verb to use in this case.
“is equivalent as” -> “is equivalent to”
“In section 4, we present a polynomial time algorithm” -> you seem to present multiple algorithms
The deficiency is […] the set of edges, whose addition make the neighbourhood -> make should be makeS, after the formula and before “the set of edges”, insert some punctuation or “[…], that is, the set of edges”
Lemma 4.5: The deficiency [formula] and the 3-clique path is maximal imply that -> The fact that the deficiency [formula] and that the 3-clique path is maximal imply that
Page 8: The source code are available -> the source code is available
this is equivalent as -> equivalent to […] in polynomial time too. -> in polynomial time.
We drew a connect -> drew a connection
We have got some positive results indicate -> positive results THAT indicate
Figure 9: it consists OF three isomorphic subgraphs
“then the subgraph corresponds to the variable [...]” -> […] THAT corresponds
Page 12: “as what did [13]” -> “as what [13] did”

3. Please provide an "overall score" for this submission.
6: Marginally above the acceptance threshold. I tend to vote for accepting this submission, but rejecting it would not be that bad.

4. Please provide a "confidence score" for your assessment of this submission.
3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

5. Improvements: What would the authors have to do for you to increase your score?
Guidance on improvements is included in the detailed comments.