View Reviews

Paper ID

2742

Paper Title

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

Reviewer #1

Questions

1. Please enter a detailed review describing the strengths and weaknesses of the submission.

Summary:

The paper considers the problem of checking morality of undirected graphs, i.e., checking if there exists a DAG whose moral graph is the same as the given undirected graph. The main contribution of the paper is to show that checking morality of the graph is equivalent to finding an elimination ordering of vertices and subsets of edges, which when removed result in a graph where there exists at least one node for which there are no missing edges over its set of neighboring nodes. The problem of checking morality is NP-complete and the authors show that for graphs with degree at most 4 the problem is poly-time solvable while for graphs with degree 5 the problem remains NP-complete.

Review:

The main contribution of the paper is to show that an undirected graph G is moral if there exists a sequence of vertices and edges (v, E_v), where E_v is a subset of edges over the neighbors of v in G, such that on removing v and E_v from G results in a graph G' in which there exists at least one node for which the edge set over its neighbors is complete (i.e. there are no missing edges). This simple concept is presented using a plethora of definitions: simplical node, deficiency, recursively simplical, weak recursively simplical, ordering, excesses, elimination kits, perfect elimination kits, partial perfect elimination kits. The number of definitions introduced is astounding and while they serve to formalize the paper, they make the paper very unreadable. In particular, a lot of shortcut notations could be completely avoided e.g. the notation D(G) \neq \varnothing implies for all u \in V D_G(u) \neq \varnothing. There are many more examples of such shortcut notations that make it very hard to follow the paper.

Coming up with poly-time solvable cases essentially involves looking at cases where the edge set E_v is the entire set of edges over the neighbors of v --- thus avoiding checking an exponential number of subsets. The authors prove that this happens for graphs with degree at most 4. The proof for (maximum) degree 4 graphs is non-trivial. However, I have a few questions regarding the proofs:

1. In Lemma 4.3 line 202. The presence of a degree 3 node doesn't mean the graph is K_4 (i.e. G = K_4 as claimed). Do you mean the G[N_G[x]] = K_4 ?
2. Same problem in Lemma 4.4 line 260: G = K_5).

The proof of NP-completeness of degree 5 graphs essentially follows the proof of Verma and Pearl (1993).

Minor comments:

Line 30, introduction, 2nd paragraph: is various of ways => in various ways
Line 52: degree of a node is denoted by d_G(u) in the rest of the papers. The subscripts G can be dropped throughout the paper when the graph is clear from co
ntext.
Definition 2.5: Add that empty graph is WRS given that the definition is recursive.
Is Definition 2.9 really necessary?
Proof of Lemma 4.3, line 202: x is a leave => x is a leaf.

Line 233: when stucking at => when stuck at
Line 307: dealth => dealt.

Post Rebuttal:

After having read the author's responses, I still believe that the results are of limited significance given that the algorithms developed are for at most degree 4 graphs. This rules out using ideas developed in the paper elsewhere e.g. in local to global algorithms for Bayesian network structure learning. Furthermore, the task of checking morality itself is itself not very important given that 1. it is NP-complete to do so for general graphs, 2. under faithfulness assumptions neighborhood learning methods can consistently recover moral graphs from a finite number of samples in polynomial time, 3. the set of unfaithful distributions have Lebesgue measure zero. I also believe that the presentation can be significantly improved --- the paper is needlessly pedantic and some of the definitions can be dropped. For instance, given that the graph is weak recursively simplical implies that these exists an ordering of vertices and edges that can be removed in each step so that the graph remains weak recursively simplical. So definitions and examples from 2.6 to 2.9 and Theorem 1 itself are self-evident. Also, given that the most of the proofs follow by induction and the main proofs are primarily case-by-case analysis, most of the proofs can be moved to the Appendix and more intuition can be provided about the proofs in the main text. Given these concerns I stick to my rating.

2. Please provide an overall score for the submission.
Weak Reject: Borderline, tending to reject

3. Please enter a 2-3 sentence summary of your review explaining your overall score.
While the paper makes progress on the problem of checking morality of undirected graphs, I find the scope of the paper quite limited. It is fairly easy to recover the Markov blanket and the moral graph from data using neighborhood selection methods. Analyzing the presented algorithms for the finite sample case would make the paper more compelling.

5. Please rate your confidence in the score assigned.
High: Reviewer has understood the main arguments in the paper, and has made high level checks of the proofs.

8. I agree to keep the paper and code submissions confidential, and delete any submitted code at the end of the review cycle to comply with the confidentiality requirement.
Agreement accepted

Reviewer #2

Questions

1. Please enter a detailed review describing the strengths and weaknesses of the submission.

Morality of an undirected graph U refers to the following decision problem: does any DAG G exist such that U is the moral graph of G? It has been known for a long time that deciding morality is an NP-hard. This paper now provides a tight lower bound of 5 on the maximum degree of U for which the problem remains NP-hard.

I found this an interesting paper and the theoretical work was of high quality. The authors introduce a novel graph theoretical concept called weak recursive simpliciality (WRS) and prove equivalence between this concept and morality. They then show algorithms to determine WRS for graphs with maximum degree 3 (which is rather trivial) and 4 (which is quite interesting). They then modify the NP-hardness proof of Verma and Pearl to show that max degree 5 suffices for NP-hardness. This is an important fundamental result with potential practical implications (see below).

My main criticism of this paper would be that it does not try hard enough to explain the relevance of its theoretical results (outside of the obvious contribution to the theory of graphical models) and put it into the larger context of the literature. One symptom of this problem is the reference list which contains only 5 papers. This is a pity; several opportunities to strengthen the connection of this work to the outside world have been missed. I would like to give a few examples:

- The paper starts with the concept of local causal discovery and cites Aliferis et al. This is of course a very relevant reference, but the authors could also have noted that the very popular graphical lasso algorithm (or any other UG learning algorithm, for that matter) likewise delivers an UG for which it might be interesting to check morality.

- The authors mention themselves the "immoralizing" operation in the discussion, but don't the algorithms they present already imply such an operation? Unless I am missing something, the vertices removed in each step of PEK correspond to sinks in the DAG, so can't we simply orient edges towards every node that is removed during PEK? If it's not that simple, some explanation might be in order. Also, for graphs of maximum degree higher than 4, would it at least be possible to immoralize part of the UG using similar operations?

- Simplicial vertices also play a central role in the related problem where the input undirected graph describes the marginal (rather than conditional) independencies in a distribution, and one asks whether any DAG exists that is compatible with this pattern. See Pearl and Wermuth, "When Can Association Graphs Admit A Causal Interpretation?, volume 89 of Lecture Notes in Statistics, pages 205–214. Springer, 1994; as well as https://arxiv.org/pdf/1508.00280.pdf

- Could the authors expand more on how this result could be useful for structure learning? Which specific algorithms could it be combined with and in which specific way?

Some minor suggestions:

- The introduction "beats around the bush" somewhat by introducing the symmetry property, which is obvious buy not central to the rest of the paper, and more importantly the consistency property, which appears to be trivially equivalent to morality such that one wonders why this term is introduced at all; it is also never used except in the introduction and discussion. Morality appears easy enough to explain, why this additional term?

- In section 4, it appears strange at first that separate algorithms are given for degree 3 and 4. For the theoretical result (the tight bound), the degree 3 case is not relevant in itself. The main motivation for giving both appears to be that the algorithm for degree 3 is much simpler, but this should be made explicit.

- Lemma 4.7: please don't write "the proof is trivial" when it would take 2 sentences and you have more than enough space

= After author response =

I appreciate the response to the comments, but I felt that the authors could have done a better job of answering the questions and incorporating the feedback. It appears that the authors were not willing to do more than rewrite or add a few sentences in the introduction or discussion. In particular, I would have liked an answer to my comment on the "immoralizing" algorithm. Also the comment of the other reviewer with respect to experimental evaluation should have been taken much more seriously. The theoretical analysis provides only asymptotic runtime, and the authors could for instance have offered to run a few simulations to check the graph sizes for which their algorithms would still be feasible.

Overall, I stand by my original evaluation of the paper and will keep the score unmodified.

2. Please provide an overall score for the submission.

Weak Accept: Borderline, tending to accept

3. Please enter a 2-3 sentence summary of your review explaining your overall score.

Good paper with a theoretical result of fundamental interest. Could do a better job however embedding this work into the existing literature and pointing out potential use cases.

5. Please rate your confidence in the score assigned.
High: Reviewer has understood the main arguments in the paper, and has made high level checks of the proofs.

8. I agree to keep the paper and code submissions confidential, and delete any submitted code at the end of the review cycle to comply with the confidentiality requirement.
Agreement accepted

Reviewer #3

Questions

1. Please enter a detailed review describing the strengths and weaknesses of the submission.
Summary and Decision: The main contribution of this paper is a characterization of moral graphs (i.e. those undirected graphs that arise as moralizations of DAGs) via an extension of the notion of perfect elimination orderings (PEOs) for chordal graphs. They then use this characterization of moral graphs to prove that the morality of an undirected graph with maximum degree at most D for D = 2,3,4 can be decided in polynomial time, however, the problem becomes NP-complete for D = 5. The decidability of the graph morality problem was previously considered by Verma and Pearl (1993), who proved that it was in general NP-complete. So these results really show that restricting by degree bounds only improves the complexity story for extremely sparse graphs. While the mathematical characterization of moral graphs is nice, its consequences for the graph morality problem seem only incremental. Moreover, the authors claim that their main motivation is actually the problem of determining Markov blanket consistency. However, the connections to this problem are only mentioned in the introduction and conclusion of this paper, and are never explicitly drawn in any other place in the paper. More so, even if this connection was made more explicit in the paper, the complexity results in section 4 suggest that these methods could only be useful in the Markov Blanket consistency problem for extremely sparse graphs, and are otherwise unhelpful. So again the results feel incremental, thereby weakening the sense of potential impact for this paper. Additionally, the paper is unfortunately not so well-written, suffering particularly from a lack of helpful referencing, the use of undefined terminology, the use of inappropriate mathematical and phonetic abbreviations, as well as a general sense of hastiness. Based on all of these things, I am afraid that I cannot recommend this paper for acceptance at ICML, unless the authors can make major improvements to the quality of the paper and a much more convincing argument for the non-incrementality of their results. Below I give a list of some other major/minor concerns about the paper.

Major Concerns:

This paper could benefit from the addition of helpful references throughout. For instance, in the introduction, the authors could already give references about where the symmetry and consistency properties for Markov blankets were first considered (or even a helpful summary). Note that the referenced Aliferis paper does not use these terms. Additionally, if Aliferis is to serve as a comprehensive reference, why is it that you chose only to cite part II of this paper? Also, in the introduction, where you claim that the Markov blanket consistency problem has been only sparsely considered, it would be appropriate to include the few references to where it has actually been considered. Later on in this paragraph, you could also cite the papers which used/developed the local to global methods (note that there is nothing wrong with citing Aleferis more than once!). Additionally, it would be helpful if you would explain a bit more about what you actually mean here. For example, why are these methods insufficient? How will your methods improve our ability to solve this problem?

In general, the introduction could really benefit from more detail as I began to mention in my above comment. Another example occurs in the final paragraph of the intro: This paragraph states that the problem of graph morality is going to be a key focus, but you have only spoken about Markov blanket consistency up until this point. Details on the graph morality problem, its relevance/history, and how it relates to the Markov blanket problem are all absent. Moreover, the connection between the Markov blanket problem and the graph morality problem is never made explicit, nor is it mentioned again until the conclusion of the paper. This feels extremely strange, since you claim that it is the central contribution of your paper.

In regards to the previous two comments, a general problem is that the paper lacks sufficient detail about a number of fundamental ideas claimed to be central to the paper. This is also strange since the paper is quite short. You have the space to improve your exposition with more detail.

The Background Material section (section 2), is also strangely organized and seems to be full of holes. For instance, in regards to organization, it seems completely random which definitions end up in a definition environment and which ones end up in a paragraph. I would recommend reserving the definition environment for definitions that are newly introduced in this paper, and then defining all other terms in line. This would also have the effect of shortening this section, which would make it more pleasant for the reader. In regards to holes, some things that seem to go undefined include the following:
-directed cycle (which is used to define a DAG)
-faithfulness of a Bayesian Network (used without clarification, or remark of its relevance, in the intro)
-moral graph (the moral graph of a DAG is defined, but what it means for an undirected graph to be moral is never defined. This notion is key in this paper.)
-A stack of graphs (a key notion in the proofs in section 4)
Additionally, some of the definitions given seem to be wrong. For instance, an ordering of a node set is more than just a bijection of sets; it uses the natural order on [n]. Why not just assume your nodes are already labeled and then define an ordering as a permutation of their labels? The ``closed neighbors of u in G'' is normally called the ``closed neighborhood of u in G.'' Also, you define this using the open neighborhood which goes undefined (until perhaps afterwards where you then define a much more general concept that you never seem to use in its full generality). Additionally, are you sure that you want the neighbors of the subgraph to be all nodes that are cone points over the subgraph? Or do you simply want the set of nodes that are adjacent to any vertex of H? When H is a point, these two things coincide, but they do not do so in general. Additionally, on line 73, col 1, you let P be a joint probability distribution on the random variables in V, but V is a collection of nodes. P should really be a distribution on a collection of random variables indexed by V. Another definition problem: on lines 105-106, col 1, doesn't your definition of G+E' just equal G since E'\subset E? Please clarify what you want here.

Lines 82-83: col 2: A graph is chordal if and only if it is recursively simplicial. You do not state this here, but you use it later on.

In the complexity section, the consequence of your results for graphs with maximum degree at most D for D greater than five should also be noted.

In theorems 4.1 and 4.2 more detail is needed. For example, give some references to the polynomial time results you are using (even if they are just standard texts). Additionally, it could be nice to just compute the complexities here. You have the space.

The first sentence of the conclusion (section 5) claims that this paper drew a connection between checking Markov blanket consistency and graph morality. This connection was never made sufficiently explicit. Here would also be a good spot to discuss the consequences of your complexity results in relation to the Markov blanket consistency problem. You really need to make a case at some point that your observations offer less than incremental contributions to solving this problem.

Minor Comments:

every ``s.t.'' in the paper should really be ``such that''

every instance of ``\exists'' that happens in line and not in a set (including sentences in algorithm and theorem environments) should be replaced with ``there exists'' (similarly for any such ``\forall'')

Line 187-188: col 1: the sentence ``Some of these lemmas are proved by contradiction'' sounds strange. Why do the authors feel compelled to note this?

Line 233: col 1: ``stucking'' --> ``it gets stuck at'' or something similar...

Line 249: col 1: are --> will be

The sentence in between the statement of Theorem 4.3 and the proof environment should really just be included in the proof environment as the first sentence of the proof.

2. Please provide an overall score for the submission.
Reject: Clearly below the acceptance threshold

3. Please enter a 2-3 sentence summary of your review explaining your overall score.
As noted in my review above, my choice of overall score is based on my opinions that the results appear to be at best incremental, and hence not likely to have a major impact, together with the fact that the overall writing quality of the paper is low to the point that it obscures the value of the results. In its current form, I do not think the paper meets the significance or clarity standards for acceptance at ICML. However, if the authors can make major improvements to the presentation of their results and a more convincing argument for the non-incrementality of their results, I am happy to reconsider my position.

5. Please rate your confidence in the score assigned.
High: Reviewer has understood the main arguments in the paper, and has made high level checks of the proofs.

8. I agree to keep the paper and code submissions confidential, and delete any submitted code at the end of the review cycle to comply with the confidentiality requirement.
Agreement accepted

Reviewer #4

Questions

1. Please enter a detailed review describing the strengths and weaknesses of the submission.
The paper mainly aims to check the consistency of MB, which is a less studied issue. Authors show the a way to check MB consistency with DAGs with morality and provide proofs.

Strength:
Rigorous proof.

Weakness:
1. the paper is hard to read with many new concepts, even though with many examples.
2. No empirical evaluation of the proposed algorithm.

Major Comments:
1. MB consistency is definitely an important issue to resolve, as authors mentioned, it is vital for local-to-global BN structure learning. However, contradictories usually only occurs in finite samples (for example, score-based MB discovery algorithm do enjoy consistency of score-based methods). This paper does not really address this practical issue, which diminishes the main motivation of this work and limits the novelty of this paper.
2. Empirical evaluation on the proposed data would be helpful to test whether consistency and complexity holds as claimed, even on some synthetic data.

2. Please provide an overall score for the submission.
Weak Accept: Borderline, tending to accept

3. Please enter a 2-3 sentence summary of your review explaining your overall score.
Many results are not surprising, and the practicality of the paper is very limited.

5. Please rate your confidence in the score assigned.

High: Reviewer has understood the main arguments in the paper, and has made high level checks of the proofs.

8. I agree to keep the paper and code submissions confidential, and delete any submitted code at the end of the review cycle to comply with the confidentiality requirement.

Agreement accepted