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

The complexity of checking Markov blankets consistency with DAGs via morality

First Author · Second Author · Third Author

Received: date / Accepted: date

1 Introduction

Introduced by (peral citation) as the smallest subset of variables in a Bayesian network, given which the target variable is conditional independent from the rest of the variables, Markov blanket has became popular for scaling up learning Bayesian network structures and causal discovery (give a few citations). It naturally divides the problem of learning a global structure into learning local structures within Markov blankets that can be done in parallel. Markov blankets have also been applied during data preprocessing as a feature selection technique that looks for the smallest but most informative subset of features for predicting a target, to reduce computational complexity of having a large set of features (citation). In a faithful Bayesian network (citation), the Markov blanket of a target variable consists of its parents, children and children's other parents (a.k.a., spouses) (Figure 1).

F. Author
first address
Tel.: +123-45-678910
Fax: +123-45-678910
E-mail: fauthor@example.com

S. Author
second address

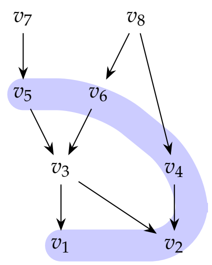

Fig. 1: A DAG, in which the Markov blanket of v_3 is $\{v_5, v_6, v_1, v_2, v_4\}$.

Since 19xx, researchers have been developing efficient algorithms for learning Markov blankets from observational data (many citations). Due to the complexity of the problem, most of these learners are heuristics based on either statistical significant tests or objective functions for optimally balancing between predictive power and complexity. A set of subsets of variables $B = \{B_1, \ldots, B_n\}$, one for each variable $v_i \in V$, which are learned from a Markov blanket learner must satisfy the symmetric and consistent properties for B being a valid set of Markov blankets. The symmetric property is a direct consequence of the graphical interpretation of Markov blankets in a directed acyclic graph (DAG). That is, $v_i \in B_j$ if and only if $v_j \in B_i$. The consistent property entails that there exists at least one DAG $G = (V, E)$ s.t. the set of parents, children and spouses of node $v_i \in V$ equals B_i for all v_i in G.

Outputs of Markov blanket learners are generally used without explicitly knowing the role of each member in B_i relative to the target variable v_i. This does not stop the symmetric property to be checked (and possibly enforced), but makes it non-trivial for checking their consistency (with a DAG). Without being consistent, a set of learned Markov blankets could lead to inconsistent local structures within Markov blankets, which has to be resolved at some stage during the transition from local to global structure learning. In this paper, we drew a connection between Markov blankets' consistency and graph morality, and presented algorithms for checking morality for undirected graphs with various of maximum degrees. The contribution of this paper is threefold. First, we introduced the concepts of *weak recursively simplicial* and *perfect elimination kit* to help checking morality and prove important proerties of moral graphs that will be studied in future investigation. Second, we developed polynomial time algorithms for checking morality for maximum degree 3 and 4 graphs. And we proved that checking morality for graphs with maximum degree 5 and above is NP-complete. Third, all the algorithms for checking morality, including a backtracking algorithm for graphs with maximum degree 5 and above, give a way of *immoralizing* moral graphs to obtain DAGs.

2 Preliminary

In this section, we introduced the a few graph theory concepts that lead to Bayesian networks, moral graphs and chordal graphs. We then introduced several new concepts that are proved to be equivalent as graph morality in the next section, and consequently are used for proving the complexity in Section **??**.

Definition 1 A **graph** is a pair $G = (V, E)$ comprising a set V of vertices (or nodes) together with a set E of edges (or arcs).

Definition 2 A **directed graph** is a graph $G = (V, E)$, where E is a set of ordered pairs of distinct vertices in V.

Definition 3 A **hybrid graph** is a graph consisting of both directed and undirected edges.

Later in this section, we introduced moral graphs, which require the direction of a directed edge to be dropped. Hence, if $G = (V, E)$ is a directed graph, we use $U(G)$ to denote the undirected version of G and $U(E)$ to denote the undirected version of G's edges.

Definition 4 A **path** is a graph $P = (V, E)$, whose vertex set $V = \{v_1, v_2, \ldots, v_{n-1}, v_n\}$ and edge set $E = \{v_1v_2, \ldots, v_{n-1}v_n\}$, where $v_i \neq v_j$ for $i, j \in [1, n]$.

We use $P = v_1 \ldots v_k$ to denote a path between v_1 and v_k. The path P has length k, denoted by $|P| = k$.

Definition 5 A **cycle** is a closed path $P = v_1 \ldots v_{k+1}$ where $v_i \neq v_j$ for $i, j \in [1, k]$ and $v_1 = v_{k+1}$.

A cycle of length m is called an m-cycle, denoted by C_m.

Definition 6 A graph is **connected** if every pair of vertices are connected by a path.

The vertex and edge set of G is denoted by $V(G)$ and $E(G)$ respectively. Throughout this paper, we use u, v to represent vertices in V and uv to represent an (undirected) edge in E. A graph refers to a connected undirected graph, unless it is said otherwise.

Definition 7 A directed graph $G = (V, E)$ is called a **directed acyclic graph** if it contains no directed cycles.

In a directed acyclic graph $G = (V, E)$, u is a *parent* of v, denoted by $u \in P(v)$ (or v is a *child* of u), if there is a directed edge from u to v. And u is an *ancestor* of v (or v is a *descendent* of u) if there is a directed path from u to v. Furthermore, v is a *nondescendent* of u if v is not a descendent of u.

Definition 8 Let $\mathcal{P}$ be a joint probability distribution of the random variables in V, and $G = (V, E)$ be a directed acyclic graph. We say $< G, \mathcal{P} >$ satisfies the **Markov condition** if for every variable $v_i \in V$, it is conditionally independent of its non-descendants given its parents set.

Definition 9 Let $\mathcal{P}$ be a joint probability distribution of the random variables in V, and $G = (V, E)$ be a directed acyclic graph. We say $< G, \mathcal{P} >$ forms a **Bayesian network** if it satisfies the Markov condition.

Definition 10 A Bayesian network $< G, \mathcal{P} >$ satisfies the **faithfulness** condition if the only conditional independencies in $\mathcal{P}$ are those entailed by the Markov condition.

Definition 11 Let $< G = (V, E), \mathcal{P} >$ be a Bayesian network. The **Markov blanket** of u in the Bayesian network, denoted by $B(u)$, is the minimum subset of variables s.t. $u \perp\!\!\!\perp_{\mathcal{P}} v \mid B(u)$ for each $v \in V \setminus B[u]$, where $B[u] = B(u) \cup \{u\}$.

Definition 12 The **moral graph** of a directed acyclic graph $G = (V, E)$ is an undirected graph $H = (V, U(E) \cup F)$, where $F = \{uv \notin U(E) \mid u, v \in P_G(x), \forall x \in V\}$ is the set of filled edges.

The above definition implicitly states a way of obtaining a moral graph from a DAG. That is, by joining all pairs of non-adjacent parents in the DAG, then dropping all the directions. The process of obtaining a moral graph from a DAG is also known as *moralization*.

Example 1 Figure 2 shows a DAG G and its moral graph H that is obtained by joining v_3 and v_4 then dropping all the directions in the hybrid graph.

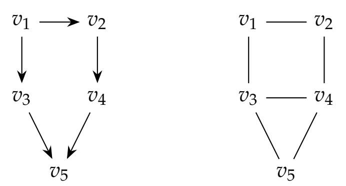

Fig. 2: A DAG $G = (V, E)$ and its moral graph $H = (V, U(E) \cup F)$, in which $v_3v_4 \in F$ is a filled edge.

If a given Bayesian network $< G = (V, E), \mathcal{P} >$ is faithful, the Markov blanket of a variable u consists of its parents, children and children's other parents. For any spouse v of u that is neither a parent nor child of u, the edge uv is a filled edge to make the moral graph H of G. Hence, for each node $u \in V$, there is a one-to-one correspondance between its Markov blanket in G and its neighbourhood in H. For example, in Figure 2 $B_G(v_3) = \{v_1, v_5, v_4\} = N_H(v_3)$.

Definition 13 A **collider** in a hybrid graph is a node with at least two parents.

Definition 14 A directed acyclic graph $G = (V, E)$ is a **consistent extension** of a hybrid graph $H = (V, F)$ if $U(E) = U(F)$ and G and H have the same set of colliders.

Definition 15 Let $G' = (V', E')$ and $G = (V, E)$ be two graphs. If $V' \subseteq V$ and $E' \subseteq E$, then G' is a **subgraph** of G, written as $G' \subseteq G$.

Definition 16 Let $G' = (V', E')$ and $G = (V, E)$ be two graphs. If $G' \subseteq G$ and $uv \in E$ for all $u, v \in V'$, then G' is an **induced subgraph** of G, written as $G' = G[V']$.

For simplicity, if $V' \subset V$ then we use $G - V'$ to denote $G[V \setminus V']$. If $V' = \{u\}$, then we use $G - u$. If $V' = V(H)$, then we use $G - H$. Similarly, if $E' \subset E$ then we use $G + E'$ and $G - E'$ to denote $(V, E \cup E')$ and $(V, E \setminus E')$ respectively. If $E' = \{uv\}$ then we use $G + uv$ or $G - uv$ instead.

Definition 17 Let $G = (V, E)$ be a graph. The set of **neighbours** of u in G is $N(u) = \{v \in V \mid uv \in E\}$. The closed neighbourhood of u in G is $N[u] = N(u) \cup \{u\}$.

It is also useful to define the neighbours of a subgraph $H \subset G$ as $N_G(H) = \{u \in V \setminus V(H) \mid uv \in E, \forall v \in V(H)\}$.

Definition 18 Let $G = (V, E)$ be a graph. The **maximum degree** of the graph is $\Delta(G) = \max\{d(u) \mid u \in V\}$, where $d(u) = |N(u)|$ is the degree of u.

Definition 19 A **clique** is a subset of nodes in a graph where every two distinct nodes are adjacent.

Definition 20 A **simplicial node** in a graph is a node whose neighbours form a clique.

Definition 21 Let $G = (V, E)$ be a graph. The **deficiency** of a node x in G is $D(x) = \{uv \notin E \mid u, v \in N(x)\}$.

A node u is simplicial in G if and only if $D(u) = \emptyset$. That is, no edge needs to be filled in to make the neighbours of u a clique. We write $D(G) \neq \emptyset$ if $D_G(u) \neq \emptyset, \forall u \in V$. And $D(G) = \emptyset$ if $\exists x \in V$ s.t. $D(x) = \emptyset$.

Example 2 In the moral graph H as shown in Figure 2, $D_H(v_1) = \{v_2v_3\}$ and $D_H(v_5) = \emptyset$.

Definition 22 A **chord** in a m-cycle $C_m = v_1 \dots v_m v_1$ is an edge $v_i v_j \notin E(C_m)$ for $i, j \in [1, m]$.

Definition 23 A graph is **chordal** if each m-cycle for $m \geq 4$ has a chord.

Definition 24 An **ordering** of a graph $G = (V, E)$ with n vertices is a bijection $\alpha : \{1, \dots, n\} \leftrightarrow V$.

Another equivalent property as being chordal is that G has a *perfect elimination ordering*. That is, for each $x \in V$ with $\alpha^{-1}(x) = i + 1$, we have $D_{G^i}(x) = \emptyset$, where $G^i = G - \{\alpha(1), \dots, \alpha(i)\}$ and $G^0 = G$.

Definition 25 A graph $G = (V, E)$ is **recursively simplicial** if $\exists x \in V$ with $D_G(x) = \emptyset$ s.t. the induced subgraph $G - x$ is recursively simplicial.

Being recursively simplicial is equivalent as being chordal. It makes chordality a hereditary property. The following concept is not as strong as recursively simplicial, so it requires some edges to be removed in addition to the removal of a simplicial node.

Definition 26 A graph $G = (V, E)$ is **weak recursively simplicial** if $\exists x \in V$ with $D_G(x) = \emptyset$ and $\exists E' \subseteq \{uv \in E \mid u, v \in N_G(x)\}$ s.t. the subgraph $G' = G - x - E'$ is weak recursively simplicial.

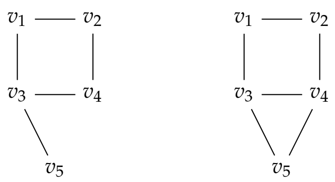

Fig. 3: An example of a non-weak recursively simplicial graph G (left) and a weak recursively simplicial graph H (right).

Example 3 H is a weak recursively simplicial (WRS) graph that can be justified by the following two criterion:

1. recursively removing $\{v_5, v_3v_4\}$, $\{v_3\}$, $\{v_4\}$, $\{v_1\}$, $\{v_2\}$ to eliminate the graph completely,
2. each node removed is a simplicial node in the current graph.

On the other hand, the graph G is not WRS, because there is no way G can be eliminated completely whilst the node removed at each step is simplicial in the current graph.

If a graph is recursively simplicial (i.e., chordal), it is also weak recursively simplicial with $E' = \emptyset$ for each simplicial node x. The converse, however, is not true. For example, the graph H in Figure 3 is WRS but not chordal.

Definition 27 A set of **excesses** of a graph $G = (V, E)$ according to an ordering α is a bijection $\epsilon_\alpha : \alpha \leftrightarrow \{\epsilon_\alpha(v_1), \ldots, \epsilon_\alpha(v_n)\}$, where each $\epsilon_\alpha(v_i) \subseteq E(G[N(v_i)])$ consists of some edges between the neighbours of v_i.

The composition $\kappa = (\alpha, \epsilon_\alpha)$ of an ordering and a set of excesses is called an elimination kit of a graph G. We use $\kappa(1)$ to denote the node $\alpha(1)$ and its excess $\epsilon_\alpha(\alpha(1))$. By using the concept of elimination kit, we extend the definition of the subgraph $G^i = G - \{\kappa(1), \ldots, \kappa(i)\} \subset G$ for $i \in [1, n]$, with the same convention $G^0 = G$.

Example 4 An ordering $\alpha = (v_5, v_3, v_4, v_1, v_2)$ and a set of excesses $\epsilon_\alpha = (\emptyset, \emptyset, \emptyset, \emptyset, \emptyset)$ form an elimination kit of H in Figure 2.

Definition 28 Let $G = (V, E)$ be a graph and $\kappa = (\alpha, \epsilon_\alpha)$ be an elimination kit of G. It is a **perfect elimination kit** if for each $x \in V$ with $\alpha^{-1}(x) = i + 1$, we have $D_{G^i}(x) = \emptyset$.

If a graph G admits a perfect elimination kit (pek), there may exists $x \in V$ that is simplicial in a subgraph G^i, but not simplicial in G. We call such x a *local simplicial* node. In general, a graph may have none, one or more than one perfect elimination kit.

Example 5 The elimination kit in Example 4 is not perfect, because $D_{G^1}(v_3) \neq \emptyset$. The only pek for H is $\alpha = (v_5, v_3, v_4, v_1, v_2)$ and $\epsilon_\alpha = (\{v_3v_4\}, \emptyset, \emptyset, \emptyset, \emptyset)$.

Definition 29 Let $G = (V, E)$ be a graph and $\kappa = (\alpha, \epsilon_\alpha)$ be an elimination kit of G. It is a **partial perfect elimination kit** if there exists a non-empty subgraph $G^i \subset G$ s.t. $D(G^i) \neq \emptyset$ and $D_{G^{j-1}}(\alpha(j)) = \emptyset$ for $j \in [1, i]$.

A 4-cycle has no partial pek, because it has no simplicial node. A graph that has a pek may also has a partial pek.

Example 6 Example 4 is a partial pek, because $D(G^1) \neq \emptyset$ and $D_{G^0}(v_5) = \emptyset$.

3 Weak recursively simplicial graphs

The first main task of this section is to prove there is a one-to-one correspondance between moral graphs and weak recursively simplicial graphs. To do this, we prove the following two lemmas.

Lemma 1 *Let $G = (V, E)$ be a DAG and H be the moral graph of G. Then H is weak recursively simplicial.*

Proof The lemma is proved by induction on the number of nodes. Let $G(n)$ and $H(n)$ denote a DAG and its moral graph over a set of n nodes. The lemma is true for $n \leq 3$, because all graphs contain three nodes or less are WRS. Assuming $H(n)$ is WRS for an arbitrary $n \geq 4$. We want to show that the moral graph $H(n + 1)$ is also WRS. It is known that each DAG contains at least one sink x, and x becomes a simplicial node in the DAG's moral graph because its parents form a clique after moralization. Hence, the moral graph $H(n + 1)$ contains a simplicial node x. By removing x and the edges that were introduced by moralization to make x's neighbourns a clique, the resulting graph $H(n)$ is the moral graph of the DAG $G(n)$ obtained by removing x from $G(n + 1)$. The inductive hypothesis assumes that each moral graph $H(n)$ is WRS. Hence, $H(n + 1)$ is also WRS. □

Lemma 1 suggests that the moral graph of any DAG is WRS. To prove there is a one-to-one correspondance it is remaining to show that the converse is also true. That is, a WRS graph is the moral graph of a DAG.

Lemma 2 *Let $H = (V, E)$ be a weak recursively simplicial graph. Then H is the moral graph of a DAG.*

Proof The lemma is proved by induction on the number of nodes n. The statement is true for $n = 1$, because a single node graph $H(1)$ is both the moral graph of $G(1)$ and a WRS graph. Assuming the lemma is true for an arbitrary $n \geq 2$. That is, any WRS graph $H(n)$ with $n \geq 2$ is the moral graph of a DAG $G(n)$. Each WRS graph has a simplicial node. Assuming x is a simplicial node of a WRS graph $H(n+1)$ and x is the first in the ordering α (i.e., $\alpha^{-1}(x) = 1$). When orienting $H(n + 1)$, we first let all the edges connect to x be directed towards it. Furthermore, some edges between the neighbours of x are removed so that the resulting graph $H(n)$ is still WRS. By the inductive assumption, $H(n)$ is the moral graph of a $G(n)$. Hence, by attaching x and the edges connect to it onto $G(n)$ we obtain a DAG $G(n+1)$, whose moral graph is $H(n+1)$. □

Theorem 1 *A graph is weak recursively simplicial if and only if it is the moral graph of a DAG.*

Proof The theorem follows from the Lemma 1 and Lemma 2. □

Moralization from a DAG to an undirected graph is trivial, whilst orienting a moral graph to obtain a DAG with the same Markov blankets is not trivial. We call such a process *immoralization*. It is beyond the focus of this paper but is an interesting topic that worth further investigating. It has a potential to get from a (symmetric and consistent) set of Markov blankets to a DAG that may not likely to be the generating Bayesian network structure, but could be used as an initial structure for structure learning and causal discovery.

Chodality is considered to be a *hereditary* property of a graph, because the subgraph is still chordal after removing a simplicial node from a chordal graph. Similarly, morality is also a hereditary property, but in the sense that the subgraph is still moral after removing a simplicial node and some edges between its neighbours from a moral graph.

Next, we present a backtracking algorithm for checking whether or not a given graph $G = (V, E)$ is WRS. If it is, the algorithm will return TRUE and orient it into a hybrid graph that always has a consistent DAG extension [Dor and Tarsi, 1992], whose Markov blankets are identical to G's.

Algorithm 1 Checking morality using backtracking

```
Require: a graph G = (V, E)
  function φ(F)
    if G = ∅ then return TRUE
    end if
    if D(G) = ∅ then
      H = G                                        ▹ cache G
      for each x ∈ V(G) s.t. D_G(x) = ∅ do         ▹ for each simplicial node
        E′ = {uv ∈ E(G) | u, v ∈ N_G(x)}
        for each ϵ(x) ⊆ E′ do                      ▹ for each excess
          G = H                                    ▹ restore G
          G = G − x − ϵ(x)
          if φ(F) = TRUE then return TRUE          ▹ apply recursion
          end if
        end for
      end for
    end if
    return FALSE                                   ▹ return FALSE if no simplicial node
  end function
```

So far, we have proved an one-to-one correspondance between moral graphs and WRS graphs. The following theorem proves that being WRS is the same as having a pek.

Theorem 2 *A graph is weak recursively simplicial if and only if has a perfect elimination kit.*

Proof If $G = (V, E)$ is WRS, the simplicial node x and the edges $E' \subset E(G[N_G(x)])$ removed at each step of the recursion form an ordering and a set of excesses, because the x at each step of the recursion is local simplicial. Hence, G admits a pek. The converse is also true because if G admits a pek, it can be eliminated recursively to the empty graph. □

Verma and Pearl [1993] made the following two remarks that could increase the speed of checking morality.

Remark 1 If a graph is moral, it has at least one simplicial node.

Remark 2 If a graph is moral, each cycle in it shares an edge with a k-clique.

4 Complexity

To check whether a graph G is WRS (or has a pek), it involves removing a subset of edges between the neighbours of a simplicial node. Not only because the number of subsets increases exponentially in the degree of a simplicial node, but a selection of subset could affect future options, which can not be anticipated at the time of selection, the prlblem of checking morality is difficult in general. Verma and Pearl[1993] proved that this problem is NP-complete for arbitrary graphs. In this section, we proved that for graphs with maximum degree less than 5, their morality can be checked in polynomial time. We presented algorithms for graphs with maximum degree 3 and 4. Furthermore, we revised the proof by Verma and Pearl so that the NP-completeness argument holds true for graphs with maximum degree 5.

If a graph is not connected, the morality of every connected component of it will be checked separately. Without loss of generality, we assume all graphs are connected. If a graph G has $\Delta(G) = 0$, it is a single node. If $\Delta(G) = 1$, it is an single edge. If $\Delta(G) = 2$, it is either a path or a simple cycle. Therefore, the morality of a graph G with $\Delta(G) \leq 2$ can be checked efficiently. To prove a polynomial time algorithm for checking maximum degree 3 graphs, we prove the following lemmas first.

Lemma 3 *If $G = (V, E)$ is not moral, then $H = G + uv$ is not moral for any pair of non-adjacent $u, v \in V$ s.t. $N_G(u) \cap N_G(v) = \emptyset$.*

Proof G is not moral implies $D(G) \neq \emptyset$ or it has only partial peks. $N_G(u) \cap N_G(v) = \emptyset$ implies uv is not part of any 3-clique. So if $D(G) \neq \emptyset$, then $D(H) \neq \emptyset$. Assume G has only partial peks. For each partial pek $(\alpha, \epsilon_\alpha)$, if x satisfies that $D_{G^{\alpha^{-1}(x)-1}}(x) \neq \emptyset$, then it also satisfies $D_{H^{\beta^{-1}(x)-1}}(x) \neq \emptyset$ where (β, ϵ_β) is a partial pek of H. Futhermore, $N_G(u) \cap N_G(v) = \emptyset$ entails that for any $x \in V$ its excess $\epsilon_\alpha(x) = \epsilon_\beta(x)$. Hence, H has only partial peks and consequently is not moral. □

Lemma 4 *Let $G = (V, E)$ be a moral graph. If $\exists x \in V$ s.t. $D_G(x) = \emptyset$ and $\forall u, v \in N_G(x)$, $N_G(u) \cap N_G(v) \subset N_G[x]$, then $G' = G - x - E(G[N_G(x)])$ is moral.*

Proof Assume G' is not moral. The removal of $E(G[N_G(x)])$ implies every pair of neighbours $u, v \in N_G(x)$ are non-adjacent in G'. In addition, $N_G(u) \cap N_G(v) \subset N_G[x]$ implies $N_{G'}(u) \cap N_{G'}(v) = \emptyset$. By Lemma 3, the subgraph $G'' = G - x - S$ is not moral for any proper subset $S \subsetneq E(G[N_G[x]])$. Without loss of generality, assuming $\alpha^{-1}(x) = 1$, so $\nexists \epsilon_\alpha(x) \subset E(G[N_G(x)])$ s.t. the subgraph $G - x - \epsilon_\alpha(x)$ is moral. This contradicts with G being moral. □

Lemma 5 *Let $G = (V, E)$ be a moral graph with $\Delta(G) = 3$. If $\exists x \in V$ with $D_G(x) = \emptyset$, then $G' = G - x - E(G[N_G(x)])$ is moral.*

Proof When $d_G(x) = 3$, $G = K_4$ and $E(G') = \emptyset$, so G' is moral. When $d_G(x) = 1$, x is a leaf. Hence, $E(G[N_G(x)]) = \emptyset$ and consequently G' is moral.

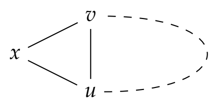

Fig. 4: A graph G with $\Delta(G) = 3$ and $D_G(x) = \emptyset$.

When $d(x) = 2$, if $uv \notin E(C_m)$ for $m \geq 3$, then G' is moral. Assuming $uv \in E(C_m)$ (Figure 4). For $m = 3$, $\exists w \neq x$ s.t. $w \in N_G(u) \cap N_G(v)$, then $H = G[\{x, u, v, w\}])$ is triangulated. Since $d_G(u) = d_G(v) = \Delta(G)$, $N_G(G - H) = w$. That is, the rest of the graph is connected to H via a single node w only. Hence, G' is moral. For $m \geq 4$, G' is also moral by Lemma 4. □

Algorithm 2 Checking morality for maximum degree 3 graphs

Require: a graph $G = (V, E)$ with $\Delta(G) = 3$
  while $\exists x$ s.t. $D_G(x) = \emptyset$ **do**
    $G = G - x - E(G[N_G(x)])$
  end while
  if $G = \emptyset$ **then**
    return TRUE
  else
    return FALSE
  end if

Theorem 3 *The morality of maximum degree 3 graphs can be check in polynomial time.*

Proof A straightfoward algorithm (Algorithm 2) for checking morality for maximum degree 3 graphs can be deduced directly from Lemma 5. If G is moral, Algorithm 2 stops till G is empty and returns TRUE, else it stops at a non-empty subgraph of G and returns FALSE.

A graph G with $|V| = n$ nodes an be represented by an adjacency list. Then it takes constant time to find $N(x)$ for $x \in V$. $\Delta(G) = 3$ implies $|N(x)| \leq 3$, so its constant time to verify $D(x) = \emptyset$. In the worst case scenario, the only simplicial node is always at the end of the list, so it takes $O(n)$ time to find it. The operations of removing x, $\{xy \in E \mid \forall y \in N_G(x)\}$ and $E(G[N(x)])$ take constant time. The while loop repeats at most n times, hence Algorithm 2 is a polynomial time algorithm and has computational complexity $O(n^2)$. □

The complexity of Algorithm 2 may be reduced further by using a different data structure, but it is beyond the focus of this paper. Rose *et al.* [1976] proved chordality can be checked in polynomial time using a specially-designed data structure. For graphs with maximum degree 4, there are more cases to be considered.

Lemma 6 *Let $G = (V, E)$ be a moral graph with $\Delta(G) = 4$. If $\exists x \in V$ with $d_G(x) = \{1, 4, 3\}$ and $D_G(x) = \emptyset$, then $G' = G - x - E(G[N(x)])$ is moral.*

Proof The proof is similar as that of Lemma 5. □

Lemma 7 *Let $G = (V, E)$ be a moral graph with $\Delta(G) = 4$. If $\exists x \in V$ with $d_G(x) = 2$ and $D_G(x) = \emptyset$ s.t. $N_G(u) \cap N_G(v) = \{x\}$, then $G' = G - x - E(G[N(x)])$ is moral.*

Proof It follows from Lemma 4. □

We use K_3^m to denote a stack of more than one K_3. For example, Figure 5 contains a K_3^2.

Lemma 8 *Let $G = (V, E)$ be a moral graph with $\Delta(G) = 4$ and $K_3^2 \subset G$. If $\exists x \in V(K_3^2)$ with $d_G(x) = 2$ s.t. $D_G(x) = \emptyset$, then $G' = G - x$ is moral.*

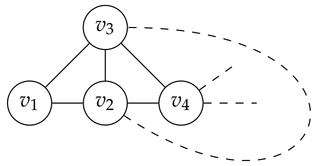

Fig. 5: A graph G with $\Delta(G) = 4$ and $K_3^2 \subset G$.

Proof Without loss of generality, assuming $K_3^3 \not\subset G$ and G is labelled as shown in Figure 5 with $x = v_1$. If $\exists u \in N_G(v_2) \cap N_G(v_3)$ s.t. $u \notin \{v_1, v_4\}$, then $H = G[\{v_1, v_2, v_3, v_4, u\}] \subset G$ is triangulated. $\Delta(v_2) = \Delta(v_3) = \Delta(G)$ implies $N_G(G-H) = \{v_4\}$. Hence, if G is moral then G' is moral. If there exists no such a node u, assuming G' is not moral. To reach a contradiction with G being moral, it is sufficient to show the subgraph $G'' = G - x - v_2v_3$ is not moral.

v_2 and v_3 are the only two nodes in G' that have the potential to become simplicial by the removal of v_2v_3 from G'. If G' is not moral, then $D(G') \neq \emptyset$ or it has only partial peks. Assuming $D(G') \neq \emptyset$, then $d_{G'}(v_2) = d_{G'}(v_3) = 3$. By $K_3^3 \not\subset G$, it entails $N_G(v_3) \cap N_G(v_4) = \{v_2\}$. Hence, $D_{G''}(v_2) \neq \emptyset$ and $D_{G''}(v_3) \neq \emptyset$. Therefore, G'' is not moral. Assuming G' has only partial peks. The partial peks $(\alpha, \epsilon_\alpha)$, in which $D_{G'^{\alpha^{-1}(v_2)-1}}(v_2) = D_{G'^{\alpha^{-1}(v_3)-1}}(v_3) = \emptyset$, remains to be partial peks for G''. For any other $(\alpha, \epsilon_\alpha)$ where $D_{G'^{\alpha^{-1}(v_2)-1}}(v_2) \neq \emptyset$ or $D_{G'^{\alpha^{-1}(v_3)-1}}(v_3) \neq \emptyset$, it implies $D_{G'^{\alpha^{-1}(v_4)-1}}(v_4) \neq \emptyset$ and $d_{G'}(v_2) = d_{G'}(v_3) = 3$. This implies $D_{G''}(v_2) \neq \emptyset$ or $D_{G''}(v_3) \neq \emptyset$. Hence, the set of all partial peks of G'' is a proper subset of the set of all partial peks of G'. Therefore, G'' is not moral. □

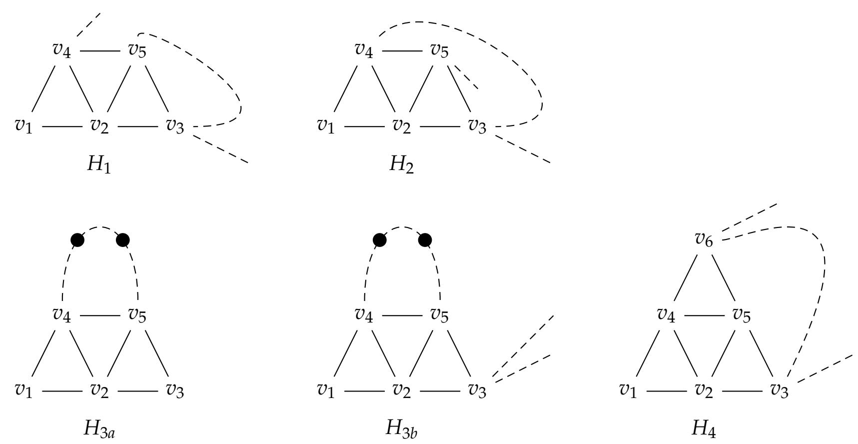

Fig. 6: A list of all possible induced subgraphs $H \subset G$ with $\Delta(G) = 4$, where $K_3^3 \subset H$. Each dotted edge connects a node to a subgraph (possibly empty) of G. H_4 is a special case of H_3 when $N_G(v_4) \cap N_G(v_5) = \{v_2, v_6\}$. Without loss of generality, assume $D_G(v_1) = \emptyset$.

Lemma 9 *Let $G = (V, E)$ be a moral graph with $\Delta(G) = 4$ and $K_3^3 \subset G$ as shown in Figure 6. If* $\max\{|P| \mid P = v_4 \dots v_5 \in G[V - \{v_1, v_2, v_3\}]\} \leq 2$ *and* $D_G(v_1) = \emptyset$, *then* $G' = (V - \{v_1\}, E - E(G[N[v_1]]))$ *is moral.*

Proof The lemma assumes G contains a subgraph that is identical to either H_1, H_2 or H_4. It is then safe to remove v_1 and v_2v_4, because $v_4v_5 \notin C_m \subset G[V \setminus \{v_1, v_2, v_3\}]$ for $m \geq 4$. Hence, $K_3 \subset G'$ is moral by Lemma 7. □

Algorithm 3 presents an efficient way of checking morality for graphs with maximum degree 4. Its correctness and complexity are proved in Theorem 4. The alogirhtm gets rid of simplicial nodes with degree 4, 1 and 3 first. Then it deals with simplicial nodes with degree 2, whose neighbours can adjacent to at most two other nodes. Within degree 2 simplicial nodes cases, it identifies a long stack of K_3s (if

Algorithm 3 Checking morality for maximum degree 4 graphs

Require: a graph $G = (V, E)$ with $\Delta(G) = 4$
  if $\exists x$ s.t. $D_G(x) = \emptyset, d_G(x) = 4$ **then**
    return TRUE
  end if
  while $D(G) = \emptyset$ **do**
    if $\exists x$ s.t. $D_G(x) = \emptyset, d_G(x) = 1$ **then**
      $G = G[V - \{x\}]$
    else if $\exists x$ s.t. $D_G(x) = \emptyset, d_G(x) = 3$ **then**
      $G = (V - \{x\}, E - E(G[N_G[x]]))$
    else if $\exists x \in K_3^m$ s.t. $D_G(x) = \emptyset$ for $m \geq 4$ **then** ▷ reduce long stack
      $G = (V - \{x\}, E - E(G[N_G[x]]))$
    else if $\exists x \in K_3$ s.t. $D_G(x) = \emptyset$ **then**
      $G = (V - \{x\}, E - E(G[N_G[x]]))$
    else if $\exists x \in K_3^2$ s.t. $D_G(x) = \emptyset$ **then**
      $G = G[V - \{x\}]$
    else ▷ all simplicial $x \in K_3^3$
      if $\max\{|P| \mid P = v_4 \dots v_5 \in G[V - \{v_1, v_2, v_3\}]\} \leq 2$ **then**
        $G = (V - \{x\}, E - E(G[N_G[x]]))$
      else
        if $\exists y \in K_3^3$ s.t. $D_G(y) = \emptyset$ and $|N_G(x) \cap N_G(y)| = 1$ **then** ▷ $H_{3a} \in G$
          $G = G[V - \{x, y\}]$
        else ▷ $H_{3b} \in G$
          $G = (V - \{x\}, E - E(G[N_G[x]]))$
        end if
      end if
    end if
  end while
  return FALSE

there is any) and reduces its length recursively. Each time a simplicial node and the edge between its neighbours are removed, the length of the stack is reduced by 2. Depending on the length of the original stack, it will be reduced to a stack of $1, 2$ or 3. It is worth mentioning that if $x \in K_3^3$ is in a stack of 3 K_3s, the cases are dealt in a fixed order. The reason behind it can be seen in the examples in Figure 7.

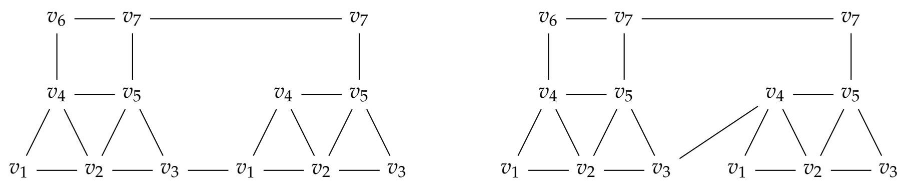

Fig. 7: Examples of simplicial nodes in K_3^3. According to Algorithm 3, in the left graph, v_3 and $v_2 v_5$ need to be removed before v_1 and $v_2 v_4$; in the right graph, v_1 and v_3 need to be removed before v_1 and $v_2 v_4$. For otherwise, neither of these two graphs can be eliminated completely.

Theorem 4 *The morality of maximum degree* 4 *graphs can be checked by Algorithm 3.*

Proof A part of the algorithm has been proven correct by Lemma 6, Lemma 7, Lemma 8 and Lemma 9. If $K_3^m \subset G$ for $m \geq 4$, removing x and $E(G[N_G[x]])$ reduces m by 2 at a time. It remains to show that the two operations for the cases of $x \in H_{3a}$ and $x \in H_{3b}$ are correct.

Assume G is moral but G' is not. If $x \in H_{3a}$, without loss of generality, assume $H \in \{H_1 = G[V \setminus \{v_1, v_2, v_3\}], H_2 = G[V \setminus \{v_1, v_2, v_3, v_4, v_5\}]\}$ s.t. $D_H(w) \neq \emptyset, \forall w \in V(H)$. It is easy to see that for any $\emptyset \neq S \subset \{v_2v_4, v_2v_5\}$, the subgraph $G'' = (V \setminus \{v_1, v_3\}, E(G[V \setminus \{v_1, v_3\}]) - S)$ satisfies $H_1 \subset G''$ and $E(G'' \setminus H_1) \cap E(H_1) = \emptyset$. Hence, G'' is not moral. If $x \in H_{3b}$, $D_{G'}(v_2) = \emptyset$. Without loss of generality, assume $H = (V(G') \setminus \{v_2\}, E(G[V(G') \setminus \{v_2\}]) \setminus S)$ where $S \subset \{v_3v_5\}$ s.t. $D_H(w) \neq \emptyset, \forall w \in V(H)$. Let $G'' = (V \setminus \{v_1\}, E(G[V \setminus \{v_1\}]))$, then $D_{G''}(v_2) \neq \emptyset$, so G'' is not moral. Hence, if G and G' both are moral.

The complexity of Algorithm 3 ...

Theorem 5 *Let $G = (V, E)$ be an undirected graph with $\Delta(G) = 5$. The problem of checking morality for G is NP-complete.*

The proof is contained in the appendix. It is a revised version of the original proof presented in [Verma and Pearl, 1993] to avoid nodes with arbitrary high degrees.

5 Conclusion

In this paper, we have drawn an one-to-one correspondance between Markov blankets consistency and graph morality. We have presented polynomial time algorithms for checking graph morality for graphs with maximum degree 3 and 4. We have also built a polynomial time reduction based upon [Verma and Pearl, 1993] from the 3-CNF problem to graph morality for graphs with maximum degree 5 and above, and consequently proved that the problem is NP-complete. Furthermore, we introduced two new concepts-weak recursively simplicial graphs and perfect elimination kit-to help proving the complexity of checking morality and for future research on related toipics.

6 Appendix

6.1 Reduction of 3-CNF to graph morality

The following is the proof of Proposition **??** that states the problem of checking morality for maximum degree 5 graphs is NP-complete.

Proof We use the same argument that Verma and Pearl used. That is, present a polynomial time reduction from 3-CNF to graph morality. The nodes that have degree greater than 5 in their construction are $\{v_i^{15}, \bar{v}_i^{15}, S^7\}$ ([Verma and Pearl, 1993] Figure 4), which have arbitrary high degrees depending on the number of clauses t in a 3-CNF problem. To reduce their high degrees, we revise Verma and Pearl's construction based on the same variable gedget and clause gedget, but a different auxiliary gedget (Figure 8, 9, 10 respectively) and different connection rules.

The detailed construction rules are in [Verma and Pearl, 1993]. The high level description is that put the envelope graph in the auxiliary gedget at the top, so that it initializes downward directions. Since each gedget contains an envelope subgraph, it can be partially directed. The K_4 in each of the clause gedget ensures that network flow does not go through the envelop subgraph that corresponds to the negation of a variable in a variable gedget. The key differences between the revised construction and the original construction are as the following. The clause gedgets are connected by a chain to get rid of the high degree node that connects to all clause gedgets. Only the first two clause gedgets are connected to a variable gedget directly and hence form a K_3. For example, the K_3 formed by $\{F_2^1, F_3^1, \bar{X}^{15}\}$. Any additional clause gedget is connected to a variable gedget via the formed K_3, not directly onto the variable gedget. For example, node F_4^1 is connected to F_3^1 and hence $\{F_4^1, F_3^1, F_3^3\}$ form another K_3. The final constructed graph is shown in Figure 11. □

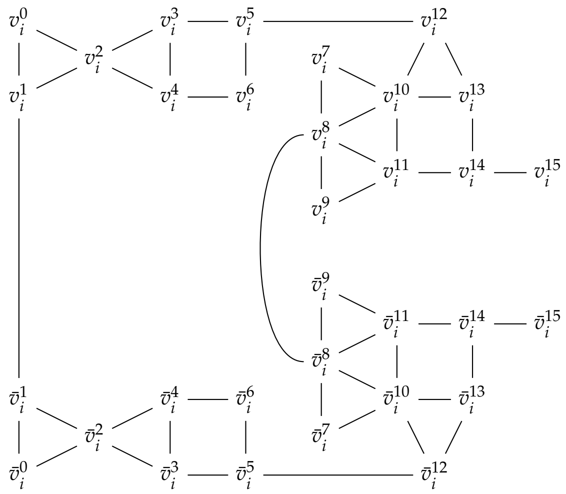

Fig. 8: A variable gedget for v_i. It contains 32 nodes, where the top half corresponding to the variable and the bottom half corresponding to the negation of the variable.

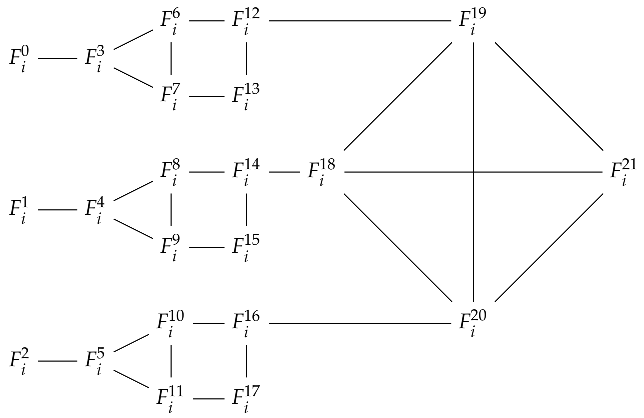

Fig. 9: A clause gedget for F_i. It contains 22 nodes. Each of the three identical subgraphs corresponding to a variable in the clause F_i.

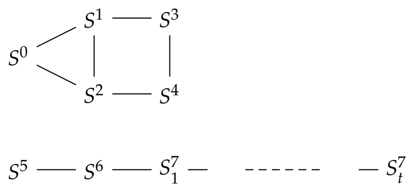

Fig. 10: An auxiliary gedget that contains $7 + t$ nodes, where t is the number of clauses in the 3-CNF.

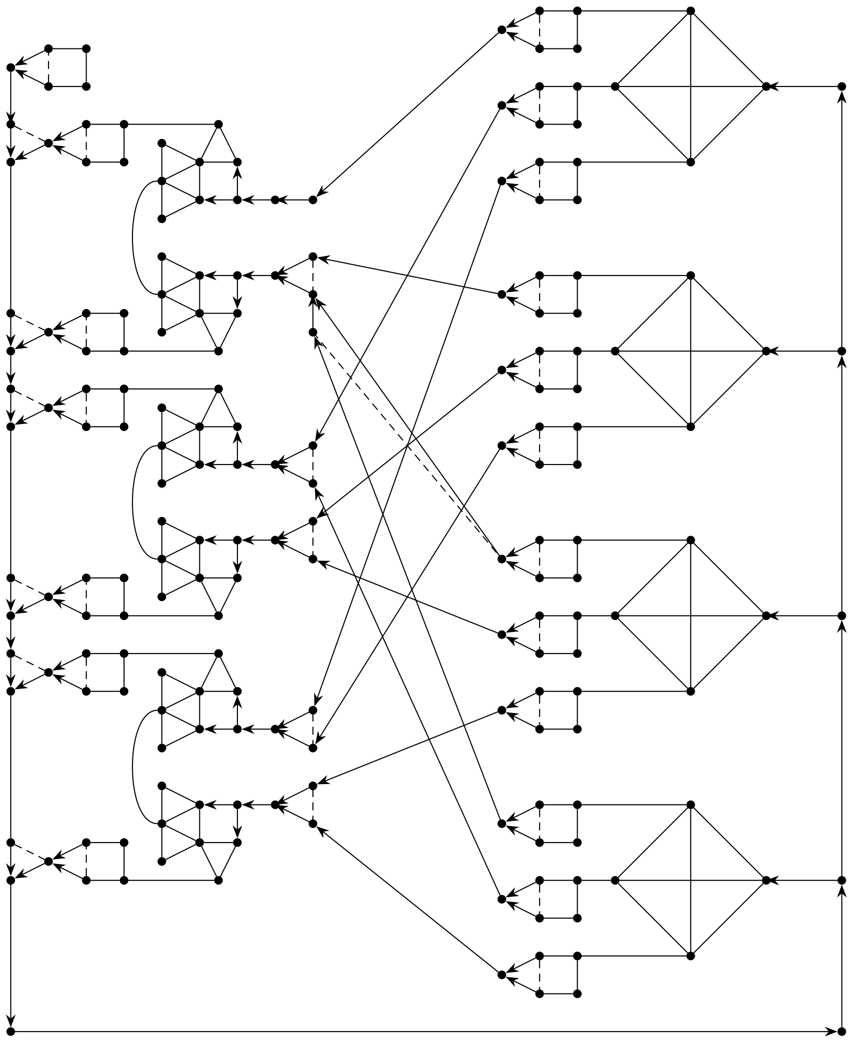

Fig. 11: A polynomial time construction that reduces 3-CNF to graph morality. The 3-CNF problem is $(X \vee Y \vee Z) \wedge (\bar{X} \vee \bar{Y} \vee Z) \wedge (\bar{X} \vee \bar{Y} \vee \bar{Z}) \wedge (\bar{X} \vee Y \vee \bar{Z})$, where $X = 1$ and $Y = Z = 0$. This construction ensures that the resulting graph has maximum degree no more than 5.

References

[Dor and Tarsi, 1992] D. Dor and M. Tarsi. A simple algorithm to construct a consistent extension of a partially oriented graph. *Technicial Report R-185, Cognitive Systems Laboratory, UCLA*, 1992.

[Rose *et al.*, 1976] D. J. Rose, R. E. Tarjan, and G. S. Lueker. Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on computing*, 5(2):266–283, 1976.

[Verma and Pearl, 1993] T. S. Verma and J. Pearl. Deciding morality of graphs is NP-complete. In *Uncertainty in Artificial Intelligence, 1993*, pages 391–399. Elsevier, 1993.