Chain Graph Interpretations and Their Relations Revisited

Dag Sonntag and Jose M. Peña

ADIT, IDA, Linköping University, Sweden
dag.sonntag@liu.se, jose.m.pena@liu.se

Abstract

In this paper we study how different theoretical concepts of Bayesian networks have been extended to chain graphs. Today there exist mainly three different interpretations of chain graphs in the literature. These are the Lauritzen-Wermuth-Frydenberg, the Andersson-Madigan-Perlman and the multivariate regression interpretations. The different chain graph interpretations have been studied independently and over time different theoretical concepts have been extended from Bayesian networks to also work for the different chain graph interpretations. This has however led to confusion regarding what concepts exist for what interpretation.

In this article we do therefore study some of these concepts and how they have been extended to chain graphs as well as what results have been achieved so far. More importantly we do also identify when the concepts have not been extended and contribute within these areas. Specifically we study the following theoretical concepts: Unique representations of independence models, the split and merging operators, the conditions for when an independence model representable by one chain graph interpretation can be represented by another chain graph interpretation and finally the extension of Meek's conjecture to chain graphs. With our new results we give a coherent overview of how each of these concepts is extended for each of the different chain graph interpretations.

Keywords: Chain graphs, Lauritzen-Wermuth-Frydenberg interpretation, Andersson-Madigan-Perlman interpretation, multivariate regression interpretation.

1. Introduction

Chain graphs (CGs) are hybrid graphs with two types of edges representing different types of relationships between the random variables of interest. These are the directed edges representing asymmetric relationships and a secondary type of edge representing symmetric relationships. Hence CGs extend Pearl's classical interpretation of directed and acyclic graphs (DAGs), i.e. Bayesian networks (BNs). However, there exist three different interpretations of CGs in research. These are the Lauritzen-Wermuth-Frydenberg (LWF) interpretation presented by Lauritzen, Wermuth and Frydenberg in the late eighties [9, 11], the Andersson-Madigan-Perlman (AMP) interpretation presented by Anderson, Madigan and Perlman in 2001 [2] and the multivariate regression (MVR) interpretation presented by Cox and Wermuth in the nineties [6, 7]. A fourth interpretation of CGs can also be found in a study by Drton [8] but this interpretation has not been further studied and will not be discussed in this paper.

Each interpretation has a different separation criterion and does therefore represent different independence models. Many papers have studied these independence models and extended many theoretical concepts regarding independence models from BNs to also work for CGs. Most of these papers have however only looked at one interpretation at a time, which has led to an incoherent picture of what theoretical concepts exist for the different CG interpretations. Moreover, this has caused research on some concepts to be missing.

In this paper we do therefore look into some of these concepts and study how they are extended to the different CG interpretations to give a coherent overview of the research performed. More importantly, we do also identify where the concepts have not yet been extended to certain CG interpretations and contribute in different ways within these areas. Specifically we look into four areas that in different ways connect to the independence models of CGs. The first area is what unique representations exist for the different independence models representable by the different CG interpretations. Having such unique representations is important since there might exist multiple CGs representing the same independence model even for the same CG interpretation. The second area concerns the feasible split and feasible merging operators. These operators are used for altering the structure of a CG without altering which Markov equivalence class it belongs to. The third area we look into is what the conditions are for when an independence model represented by one CG interpretation also can be represented by another CG interpretation. This is important since it allows us to see when the different CG interpretations overlap in terms of representable independence models. The fourth and final area concerns Meek's conjecture and whether it can be extended to the different CG interpretations. Meek's conjecture states that given two DAGs G and H, s.t. the independence model represented by G includes the independence model represented by H, we can transform G into H through a sequence of operations s.t. the independence model represented by G includes the independence model of H for all intermediate DAGs G. The operations consist in adding a single directed edge to G, or replacing G with a Markov equivalent DAG. The validity of the conjecture was proven by Chickering in 2002 [4] and has allowed several learning algorithms for DAGs to be constructed.

Our contribution, in addition to a study of previous research in the area, is then the following definitions, examples and algorithms, together with their proofs of correctness, that previously have been missing:

- The definitions of the feasible split and feasible merging operators for AMP CGs and proof that for any two Markov equivalent AMP CGs G and H there exists a sequence of feasible splits and mergings that transforms G into H.
- An example showing there are no unique representatives of equivalence classes of MVR CGs that are MVR CGs.
- An algorithm that from any AMP CG G outputs the Markov equivalent AMP essential CG H.
- The necessary and sufficient conditions for when an independence model represented by a MVR CG can be perfectly represented by a CG in another interpretation and vice versa.
- An example that proves that Meek's conjecture does not hold for MVR CGs.

The remainder of the article is organized as follows. In the next section we present the notation we use throughout the article. In Section 3 we discuss the unique representations and in Section 4 we define the feasible split and merging operators. Section 5 contains the necessary and sufficient conditions for when an independence model represented by a CG in one interpretation can be perfectly represented by a CG in another interpretation. In Section 6 we then discuss Meek's conjecture and prove that this does not hold for MVR CGs. Finally we do a short summary and conclusion in Section 7.

To improve readability of the article we have chosen to move most of the theorems, lemmas and proofs to appendices. The article does therefore include three appendices, Appendix A, B and C, that contain the theorems, lemmas and proofs of Sections 3, 4 and 5 respectively.

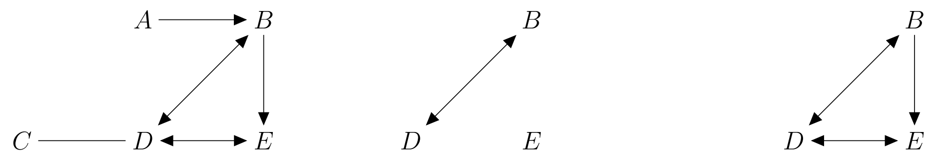
2. Notation

All graphs are defined over a finite set of discrete or continuous random variables V. If a graph G contains an edge between two nodes V_1 and V_2, we denote with $V_1{\rightarrow}V_2$ a *directed edge*, with $V_1{\leftrightarrow}V_2$ a *bidirected edge* and with $V_1{-}V_2$ an *undirected edge*. By $V_1{\circ\!\!\rightarrow}V_2$ we mean that either $V_1{\rightarrow}V_2$ or $V_1{\leftrightarrow}V_2$ is in G. By $V_1{-\!\circ}V_2$ we mean that either $V_1{\rightarrow}V_2$ or $V_1 - V_2$ is in G. By $V_1{\circ\!\!-\!\!\circ}V_2$ we mean that there exists an edge between V_1 and V_2 in G while we with $V_1{\cdots}V_2$ mean that there might or might not exist an edge between V_1 and V_2. By a *non-directed* edge we mean either a bidirected edge or an undirected edge. A set of nodes is said to be *complete* if there exist edges between all pairs of nodes in the set.

The *parents* of a set of nodes X of G is the set $pa_G(X) = \{V_1|V_1{\rightarrow}V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *children* of X is the set $ch_G(X) = \{V_1|V_2{\rightarrow}V_1$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *spouses* of X is the set $sp_G(X) = \{V_1|V_1{\leftrightarrow}V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *neighbours* of X is the set $nb_G(X) = \{V_1|V_1{-}V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *boundary* of X is the set $bd_G(X) = pa_G(X) \cup nb_G(X) \cup sp_G(X)$. The *adjacents* of X is the set $ad_G(X) = bd_G(X) \cup ch_G(X)$.

To exemplify these concepts we can study the graph G with five nodes shown in Figure 1a. In the graph we can see two bidirected edges, one between B and D and one between D and E. Hence we know the spouses of D are B and E. G also contains two directed edges between A and B and B and E and we can see that E is the only child of B and B is the only child of A. Finally G also contains one undirected edge between C and D and hence C is a neighbour of D. All and all this means that the boundary of B is A and D while the adjacents of B also contains E in addition to A and D.

A *route* from a node V_1 to a node V_n in G is a sequence of nodes $V_1, \dots, V_n$ s.t. $V_i \in ad_G(V_{i+1})$ for all $1 \leq i < n$. A *section* of a route is a maximal (w.r.t. set inclusion) non-empty set of nodes $B_1, \dots, B_n$ s.t. the route contains the subpath $B_1{-}B_2{-}\dots{-}B_n$. It is called a *collider section* if $B_1, \dots, B_n$



(a) A graph G.

(b) A subgraph of G over $\{B, D, E\}$.

(c) A subgraph of G induced by $\{B, D, E\}$.

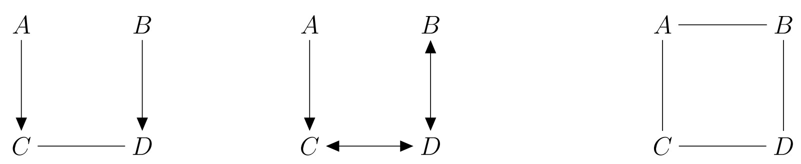
Figure 1: Three different graphs.

together with the two neighbouring nodes in the route, A and C, form the subpath $A{\rightarrow}B_1{-}B_2{-}\ldots{-}B_n{\leftarrow}C$. For any other configuration the section is a non-collider section. A *path* is a route containing only distinct nodes. The length of a path is the number of edges in the path. A path is *descending* if $V_i \in bd_G(V_{i+1})$ for all $1 \leq i < n$. A path is called a *cycle* if $V_n = V_1$. A cycle is called a *semi-directed cycle* if it is descending and $V_i{\rightarrow}V_{i+1}$ is in G for some $1 \leq i < n$. A path $\pi = V_1, \ldots, V_n$ is *minimal* if there exists no other path π_2 between V_1 and V_n s.t. $\pi_2 \subset \pi$ holds. The *descendants* of a set of nodes X of G is the set $de_G(X) = \{V_n |$ there is a descending path from V_1 to V_n in G, $V_1 \in X$ and $V_n \notin X\}$. A path is *strictly descending* if $V_i \in pa_G(V_{i+1})$ for all $1 \leq i < n$. The *strict descendants* of a set of nodes X of G is the set $sde_G(X) = \{V_n |$ there is a strict descending path from V_1 to V_n in G, $V_1 \in X$ and $V_n \notin X\}$. The *ancestors* (resp. *strict ancestors*) of X form the set $an_G(X) = \{V_1 | V_n \in de_G(V_1), V_1 \notin X, V_n \in X\}$ (resp. $san_G(X) = \{V_1 | V_n \in sde_G(V_1), V_1 \notin X, V_n \in X\}$).

To exemplify these concepts we can once again look at the graph G in Figure 1a. We can here see two paths between B and C, $B{\leftrightarrow}D{-}C$ and $B{\rightarrow}E{\leftrightarrow}D{-}C$, and that the latter of these is descending while the former is not. We can also see that the former is minimal while the latter is not since it contains one extra node E. An example of a route between B and C that is not a path is $B{\leftrightarrow}D{\leftrightarrow}E{\leftarrow}B{\leftrightarrow}D{-}C$. We can see that G contains one cycle $B{\leftrightarrow}D{\leftrightarrow}E{\leftarrow}B$ that is semi-directed. Moreover we can see that E is a strict descendant of A due to the strictly descending path $A{\rightarrow}B{\rightarrow}E$, while D is not. D is however in the descendants of A together with B, C and E. A is therefore an ancestor of all variables except itself.

A Bayesian network (BN) is a directed acyclic graph (DAG) and contains only directed edges and no semi-directed cycles. A CG under the Lauritzen-Wermuth-Frydenberg (LWF) interpretation, denoted LWF CG, contains only directed and undirected edges but no semi-directed cycles. Likewise a CG under the Andersson-Madigan-Perlman (AMP) interpretation, denoted AMP CG, is a graph containing only directed and undirected edges but no semi-directed cycles. A CG under the multivariate regression (MVR) interpretation, denoted MVR CG, is a graph containing only directed and bidirected edges but no semi-directed cycles. A *connectivity component* C in a LWF CG or an AMP CG (resp. MVR CG) is a maximal (w.r.t. set inclusion) set of nodes s.t. there exists a path between every pair of nodes in C containing only undirected edges (resp. bidirected edges). We denote the set of all connectivity components in a CG G by $cc(G)$ and the component to which a set of nodes X belong in G by $co_G(X)$. A *subgraph* of G is a subset of nodes and edges in G. A subgraph of G induced by a set of its nodes X is the graph over X that has all and only the edges in G whose both ends are in X. A *bidirected flag* is an induced subgraph of the form $X \leftrightarrow Y \leftrightarrow Z$ in a MVR CG. With the *skeleton* of a graph G we mean a graph with the same adjacencies as G but where all edges have been replaced by undirected edges. With the *moral closure graph* of a component C in a LWF CG G, denoted $(G_{cl(C)})^m$, we mean the subgraph of G induced by $C \cup pa_G(C)$ where every edge has been made undirected and every pair of nodes in $pa_G(C)$ have been made adjacent with undirected edges.

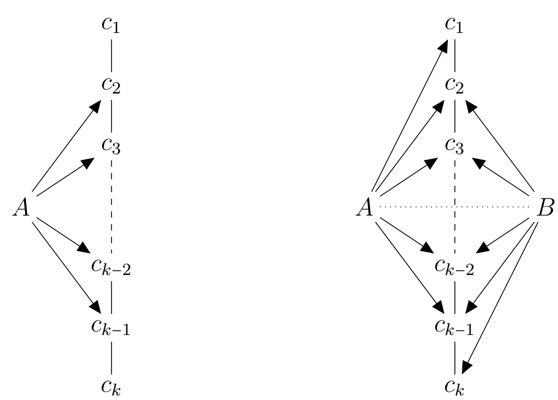
If we go back to our example in Figure 1 we can see that the graph in Figure 1b is a subgraph of G over the variables B, D and E while the graph in Figure 1c is a subgraph induced by the same variables. We can also see that G is not a CG of any of the interpretations since it contains a semi-directed cycle. An example of a LWF CG or an AMP CG H is instead shown in Figure 2a while an example of a MVR CG F is shown in Figure 2b. We can here see that H contains three connectivity components $\{A\}$, $\{B\}$ and $\{C, D\}$ and that F contains two connectivity components $\{A\}$ and $\{B, C, D\}$. An example of a bidirected flag is shown in F with the induced subgraph $C \leftrightarrow D \leftrightarrow B$ while we can see the moral closure of the component $\{C, D\}$ in H in Figure 2c.



(a) A LWF or AMP CG H. (b) A MVR CG F. (c) The moral closure of $(H_{cl(co_H(C))})^m$.

Figure 2: Three different CGs.

A *k-biflag* is an induced subgraph of either a LWF CG or AMP CG of the forms shown in Figure 3. Note that the induced subgraphs only are k-biflags if $k \geq 3$ (resp. $k \geq 2$) for the configuration seen in Figure 3a (resp. 3b).



(a) A k-biflag when $k \geq 3$. (b) A k-biflag when $k \geq 2$.

Figure 3: The possible forms of k-biflags.

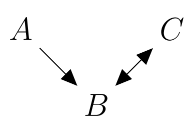
Let X, Y and Z denote three disjoint subsets of V. We say that X is *conditionally independent* from Y given Z in a probability distribution p if the value of X does not influence the value of Y when the values of the variables in Z are known, i.e. $p(X,Y|Z) = p(X|Z)p(Y|Z)$ holds and $p(Z) > 0$. We denote this by $X \perp\!\!\!\perp_p Y|Z$. When it comes to graphs we say that X is *separated* from Y given Z denoted as $X \perp\!\!\!\perp_G Y|Z$ if the following criterion is met: If G is a LWF CG then X and Y are separated given Z iff there exists no route between X and Y s.t. every node in a non-collider section on the route is not in Z and some node in every collider section on the route is in Z or $an_G(Z)$. If G is an AMP CG then X and Y are separated given Z iff there exists no S-open route between X and Y. A route is said to be *S-open* iff every non-head-no-tail node on the route is not in Z and every head-no-tail node on the route is in Z or $san_G(Z)$. A node B is said to be a *head-no-tail* in an AMP CG G between two nodes A and C on a route if one of the following configurations exist in G: $A{\rightarrow}B{\leftarrow}C$, $A{\rightarrow}B{-}C$ or $A{-}B{\leftarrow}C$. Moreover G is also said to contain a triplex $(\{A,C\},B)$ iff one such configuration exists in G and A and C are not adjacent in G. A triplex $(\{A,C\},B)$ is said to be a *flag* in an AMP CG G iff G contains one of the following subgraphs induced by A,B and C: $A{\rightarrow}B{-}C$ or $A{-}B{\leftarrow}C$. If G is a MVR CG then X and Y are separated given Z iff there exists no d-connecting path between X and Y. A path is said to be *d-connecting* iff every non-collider on the path is not in Z and every collider on the path is in Z or $san_G(Z)$. A node B is said to be a *collider* in a MVR CG G between two nodes A and C on a path if one of the following configurations exists in G: $A{\rightarrow}B{\leftarrow}C$, $A{\rightarrow}B{\leftrightarrow}C$,$A{\leftrightarrow}B{\leftarrow}C$ or $A{\leftrightarrow}B{\leftrightarrow}C$. For any other configuration the node B is a non-collider.

To exemplify these concepts we can look at the CGs in Figure 2. If we interpret the graph H in Figure 2a as a LWF CG we can see that the route $A{\rightarrow}C{-}D{\leftarrow}B$ contains one section that also is a collider section on that route. Hence we know that $A{\perp\!\!\!\perp}_H B|\varnothing$ must hold, while $A{\not\perp\!\!\!\perp}_H B|C$ also must hold since the collider section then contains a node in the given set Z. Similarly we can see that $A{\not\perp\!\!\!\perp}_H D|\varnothing$ also must hold since the route $A{\rightarrow}C{-}D$ does not contain any collider section. If we on the other hand interpret the CG H as an AMP CG we can see that $A{\perp\!\!\!\perp}_H B|\varnothing$ holds as before but that $A{\not\perp\!\!\!\perp}_H B|C$ does not hold. This is because the route $A{\rightarrow}C{-}D{\leftarrow}B$ contains two head-no-tail nodes, C in $A{\rightarrow}C{-}D$ and D in $C{-}D{\leftarrow}B$, while only C is in the given set Z. Hence the route is not S-open. Here we can also note that $A{\perp\!\!\!\perp}_H D|\varnothing$ holds since the route between A and D contains a head-no-tail node. If we finally look at the MVR CG F in Figure 2b we can note that $A{\perp\!\!\!\perp}_F B|\varnothing$ holds as before and that $A{\not\perp\!\!\!\perp}_F B|C$ does not hold, since the path between A and B contains two colliders, C and D.

The *independence model* M induced by a graph G, denoted as $I(G)$ or $I_{PGM\text{-}class}(G)$, is the set of separation statements $X{\perp\!\!\!\perp}_G Y|Z$ that hold in G according to the interpretation to which G belongs or the subscripted PGM-class. We say that two graphs G and H are *Markov equivalent* (under the same interpretation) or that they are in the same *Markov equivalence class* iff $I(G) = I(H)$. Moreover we say that G and H belong to the same *strong Markov equivalent class* iff $I(G) = I(H)$ and G and H contain the same flags. Given a probability distribution p we say that p is *Markovian* with respect to a graph G when $X{\perp\!\!\!\perp}_p Y|Z$ if $X{\perp\!\!\!\perp}_G Y|Z$ for all X, Y and Z disjoint subsets of V. Given two independence models M and N, we denote by $M \subseteq N$ that if $X \perp\!\!\!\perp_M Y|Z$ then $X \perp\!\!\!\perp_N Y|Z$ for every X, Y and Z.

An independence model can also have certain properties. Let X, Y, Z and W be four disjoint subsets of V. We say that M is a graphoid if it satisfies the following properties: *Symmetry* $X \perp\!\!\!\perp_M Y|Z \Rightarrow Y \perp\!\!\!\perp_M X|Z$, *decomposition* $X \perp\!\!\!\perp_M Y \cup W|Z \Rightarrow X \perp\!\!\!\perp_M Y|Z$, *weak union* $X \perp\!\!\!\perp_M Y \cup W|Z \Rightarrow X \perp\!\!\!\perp_M Y|Z \cup W$, *contraction* $X \perp\!\!\!\perp_M Y|Z \cup W \wedge X \perp\!\!\!\perp_M W|Z \Rightarrow X \perp\!\!\!\perp_M Y \cup W|Z$, and *intersection* $X \perp\!\!\!\perp_M Y|Z \cup W \wedge X \perp\!\!\!\perp_M W|Z \cup Y \Rightarrow X \perp\!\!\!\perp_M Y \cup W|Z$. An independence model M is also said to fulfill the *composition property* iff $X \perp\!\!\!\perp_M Y|Z \wedge X \perp\!\!\!\perp_M W|Z \Rightarrow X \perp\!\!\!\perp_M Y \cup W|Z$. Finally we do also say that p is *faithful* to G when $X \perp\!\!\!\perp_p Y|Z$ iff $X \perp\!\!\!\perp_G Y|Z$ for all X, Y and Z disjoint subsets of V.

To illustrate the last concepts we can look at the MVR CG J and the independence models in Figure 4. In Figure 4b we can see the independences that hold in J and hence the independence model of J. We can also see another independence model in Figure 4c and note that $I(J) \subseteq M$ and hence that M includes the independence model represented by J. Finally we can also see that both independence models fulfills the graphoid properties and composition property.



(a) A MVR CG J.

$A \perp\!\!\!\perp_J C|\emptyset$
$C \perp\!\!\!\perp_J A|\emptyset$

(b) The independence model of J.

$A \perp\!\!\!\perp_M C|\emptyset$
$C \perp\!\!\!\perp_M A|\emptyset$
$A \perp\!\!\!\perp_M C|B$
$C \perp\!\!\!\perp_M A|B$

(c) Another independence model M.

Figure 4: Example of independence models.

3. Unique representations

Just like many other probabilistic graphical model classes there might exist multiple CGs, in the same CG interpretation, that represent the same independence model. Sometimes it can however be desirable to have a unique graphical representation of the different representable independence models in a certain CG interpretation similarly as we have essential graphs for DAGs. Hence such unique representations have been presented by different researchers for the different interpretations. For LWF CGs these are called the largest chain graphs (LCGs) [9]. For AMP CGs we have two different unique representations, the largest deflagged graphs [17] and the AMP essential graphs [3] while for MVR CGs we have the essential MVR CGs [19]. All of these have been proven to be unique for the interpretation and Markov equivalence class they represent [3, 9, 17, 19].

Definition 1. Largest CG [9]
A LWF CG G^* is said to be the largest CG of its Markov equivalence class if it contains the maximal number of undirected edges for any LWF CG in that Markov equivalence class.

Definition 2. Largest deflagged graph [17]
An AMP CG G^* is said to be the largest deflagged graph of its Markov equivalence class iff there exists no other AMP CG H s.t. $I(G^*) = I(H)$ and either H contains fewer flags than G^* or G^* and H belong to the same strong Markov equivalence class but H contains more undirected edges.

Definition 3. AMP essential graph [3]
An AMP CG G^* is said to be the AMP essential graph of its Markov equivalence class iff for every directed edge $A{\rightarrow}B$ that exists in G^* there exists no AMP CG H s.t. $I(G^*) = I(H)$ and $A{\leftarrow}B$ is in H.

Definition 4. Essential MVR CG [19]
A graph G^* is said to be the essential MVR CG of a MVR CG G if it has the same skeleton as G and contains all and only the arrowheads common to every MVR CG in the Markov equivalence class of G.

One thing that can be noted here is that while any largest CG is a LWF CG and any largest deflagged graph or AMP essential graph are AMP CGs, an essential MVR CG does not need to be a MVR CG. Instead these graphs can contain three types of edges, undirected, directed and bidirected, and although the separation criterion defined for these graphs is close to that of MVR CGs [19], this is of course unfortunate. It can however be shown that no unique representation that is a MVR CG can exist for a Markov equivalence class of MVR CGs unless we assume some ordering of the nodes. To see this consider a system with three variables X,Y and Z for which the independence model only contains the conditional independence $X \perp\!\!\!\perp Z | Y$ and assume the contrary, i.e. that there exists a MVR CG with some unique property representing the independence model. In Figure 5 we can see the five MVR CGs representing our independence model. It can now be seen that our unique representative cannot have any bidirected edges, since we cannot distinguish between the MVR CGs shown in Figure 5b and 5c unless we assume an ordering of the nodes. Hence we can only have directed edges, but as can be seen the three remaining MVR CGs, shown in Figure 5a, 5d and 5e, all contain the same number of directed edges. Moreover, it is impossible to distinguish between the MVR CGs in Figure 5d and 5e unless we assume an ordering of the nodes. One could then imagine that we could somehow define the unique representation to contain the shortest descending path, but such an idea can easily be proven not to work for a system containing only two nodes, and no conditional independences. Hence we cannot find any representative in the Markov equivalence class with some distinguished structural property. This in turn means that we must go outside the class of MVR CGs to have a unique graph representing this Markov equivalence class.

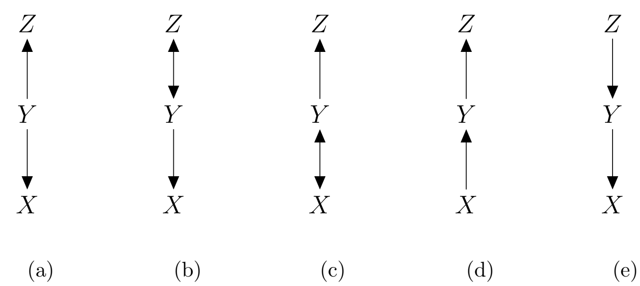


Figure 5: MVR CGs representing the independence $X \perp\!\!\!\perp Z | Y$.

To be able to identify if a graph is of a certain unique representation all the representations have been characterized [3, 17, 19, 21]. In addition to this there exist algorithms that, given a CG in a certain interpretation, outputs the unique representation of that interpretation. However, while the algorithms for the largest CGs [12, 21], the largest deflagged graphs [17] and the essential MVR CGs [19] have been proven to be correct, it does not, to the authors knowledge, exist any such proof for the algorithm of AMP essential graphs. Hence we present another algorithm, shown in Algorithm 1, that from an AMP CG G outputs the Markov equivalent AMP essential graph G^* and prove its correctness in Appendix A. The algorithm uses the notion of blocked edges. By a *block* on an edge $X - Y$ towards Y, represented as $X \multimap Y$, we mean that the edge can not be replaced by a directed edge $X \rightarrow Y$ in the final step of the algorithm. The definition of a circle on an edge is in the algorithm also extended to include such an edge ending. Hence, by

$X \multimap Y$ we mean any of the edges $X \rightarrow Y, X - Y$ or $X \dashv Y$ and with $X \circ\!\!-\!\!\circ Y$ we mean any edge between X and Y, blocked or otherwise. The algorithm also uses a set of rules shown in Figure 6. A rule is applicable if the antecedent is satisfied for an induced subgraph of G. When a rule is applied one (or two) of the non-blocked edge ends are replaced with a block as shown in the consequent of the rule while the rest of the edge ends are kept the same.

Input: An AMP CG G.
Output: The AMP essential graph G^* Markov equivalent of G.

For each ordered pair of non-adjacent nodes A and B in G
  Set $S_{AB} = S_{BA} = S$ s.t. $A \perp\!\!\!\perp_G B | S$
Let G^* denote the undirected graph that has the same adjacencies as G
Apply the rules R1-R4 to G^* while possible
Replace every edge $A - B$ in every cycle in G^* that is of length greater than three, chordless, and without blocks with $A \vdash\!\dashv B$
Apply the rules R2-R4 to G^* while possible
Replace every edge $A \vdash B$ (respectively $A \vdash\!\dashv B$) in G^* with $A \rightarrow B$ (respectively $A - B$)

Algorithm 1: Algorithm for constructing the AMP essential graph for an AMP CG G with its corresponding rules shown in Figure 6.

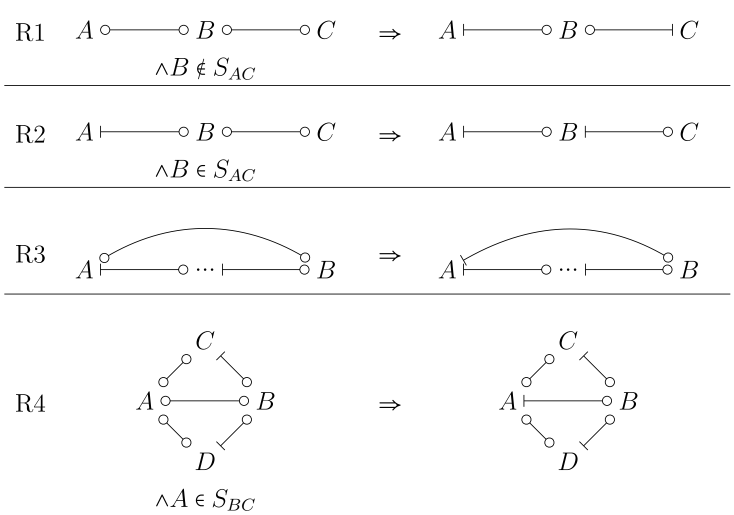


Figure 6: The rules for Algorithm 1 with the antecedent on the left hand side and the consequent on the right hand side.

We can note some things about the algorithm. In line 2, any possible S fulfilling the requirements will do. For instance, if $co_G(A) = co_G(B)$ then $S = pa_G(A \cup nb_G(A)) \cup nb_G(A)$, otherwise $S = pa_G(A)$ [14, Lemma 2]. Also note that, in line 5, a cycle with no blocks means that the ends of the edges in the cycle have no blocks. The third thing worth mentioning is that the rule R1 is not used in line 6, because it will never fire after its repeated application in line 4. Finally, note that G^* may have edges without blocks after line 6.

4. Feasible splits and mergings

Today there exist mainly two operators, the feasible split and the feasible merging, for changing the structure of a CG without changing the Markov equivalence class it belongs to, i.e. the represented independence model. More specifically the feasible merging operator defines the conditions for when the directed edges between two adjacent chain components in a CG G can be replaced by the non-directed edges in that CG interpretation without altering the represented independence model of G. The feasible split operator does the reverse, i.e. defines the conditions for when the non-directed edges between two connected sets, U and L, of nodes in G can be replaced by directed edges oriented from U towards L. An important property of the operators is that for any two Markov equivalent CGs G and H of the same interpretation there exists a sequence of feasible splits and mergings that transforms G into H.

The operators have been used in previous research in various ways such as proving theorems [18], finding the largest chain graph for a certain LWF CG [20] or exploring the Markov equivalence class in structure learning algorithms [16]. For the LWF and MVR CG interpretations these operators and their conditions have already been proven to be correct [18, 20] and hence the definitions are only repeated here. The conditions for a feasible split and feasible merging for the AMP CG interpretation have however not yet been presented, and hence we present these operators here and prove that they are sound in Appendix B. In the appendix we do also prove that for any two AMP CGs G and H of the same Markov equivalence class there must exist a sequence of feasible splits and mergings such that G is transformed into H. Note that the feasible merging operator here does not correspond to the legal merging presented in the deflagging procedure for AMP CGs by Roverato and Studený [17]. Their operation was applied to strong equivalence classes, not the more general Markov equivalence classes discussed here.

Definition 5. Feasible split for LWF CGs [20]
A connectivity component C of CG G under the LWF interpretation can be feasibly split into two disjoint sets U and L s.t. $U \cup L = C$ by replacing every undirected edge between U and L with a directed edge oriented towards L iff:
1. $\forall A \in nb_G(L) \cap U, pa_G(L) \subseteq pa_G(A)$
2. $nb_G(L) \cap U$ is complete

Definition 6. Feasible merging for LWF CGs [20]
Let U and L denote two connectivity components of G. A merging between the two components, performed by replacing every edge $X \to Y$ with $X - Y$ s.t. $X \in U$ and $Y \in L$, is feasible iff:
1. $\forall A \in pa_G(L) \cap U, pa_G(L) \setminus U \subseteq pa_G(A)$
2. $pa_G(L) \cap U$ is complete

Definition 7. Feasible split for MVR CGs [18]
A connectivity component C of CG G under the MVR interpretation can be feasible split into two disjoint sets U and L s.t. $U \cup L = C$ by replacing every bidirected edge between U and L with a directed edge oriented towards L iff:
1. $\forall A \in sp_G(U) \cap L$, $U \subseteq sp_G(A)$ holds
2. $\forall A \in sp_G(U) \cap L$, $pa_G(U) \subseteq pa_G(A)$ holds
3. $\forall B \in sp_G(L) \cap U$, $sp_G(B) \cap L$ is a complete set

---

Definition 8. Feasible merging for MVR CGs [18]
Let U and L denote two connectivity components of G. A merging between the two components, performed by replacing every edge $X \to Y$ with $X \leftrightarrow Y$ s.t. $X \in U$ and $Y \in L$, is feasible iff:
1 For all $A \in ch_G(U) \cap L$, $pa_G(U) \cup U \subseteq pa_G(A)$ holds
2 For all $B \in pa_G(L) \cap U$, $ch_G(B) \cap L$ is a complete set
3 $de_G(U) \cap pa_G(L) = \emptyset$

Definition 9. Feasible split for AMP CGs
A connectivity component C of CG G under the AMP interpretation can be feasibly split into two disjoint sets U and L s.t. $U \cup L = C$ by replacing every undirected edge between U and L with a directed edge oriented towards L iff:
1. $\forall A \in nb_G(L) \cap U, L \subseteq nb_G(A)$
2. $nb_G(L) \cap U$ is complete
3. $\forall B \in L, pa_G(nb_G(L) \cap U) \subseteq pa_G(B)$

Definition 10. Feasible merging for AMP CGs
Let U and L denote two connectivity components of G. A merging between the two components, performed by replacing every edge $X \to Y$ with $X - Y$ s.t. $X \in U$ and $Y \in L$, is feasible iff:
1. $\forall A \in pa_G(L) \cap U, L \subseteq ch_G(A)$
2. $pa_G(L) \cap U$ is complete
3. $\forall B \in L, pa_G(pa_G(L) \cap U) \subseteq pa_G(B)$
4. $de_G(U) \cap pa_G(L) = \varnothing$

Lemma 1. *A CG G in the AMP interpretation is in the same Markov equivalence class before and after a feasible split.*

Lemma 2. *A CG G in the AMP interpretation is in the same Markov equivalence class before and after a feasible merging.*

Theorem 3. *Given two AMP CGs G and H in the same Markov equivalence class there exists a sequence of feasible splits and mergings that transforms G into H.*

With these operators we can now define maximally oriented CGs which is a term used in Section 5 and various proofs in Appendix C.

Definition 11. Maximally oriented CG
A CG G (under any interpretation) is maximally oriented iff no feasible split can be performed on G.

A maximally oriented CG can be obtained from any member of its Markov equivalence class by performing feasible splits until no more feasible splits can be performed.

Theorem 4. *A CG in the AMP or MVR interpretation has the minimal set of non-directed edges for its Markov equivalence class iff no feasible split is possible.*

The following theorem shows that for the AMP and MVR CG interpretation there may exist several maximally oriented CGs in a given Markov equivalence class but all of them share the same non-directed edges.

Theorem 5. *For any Markov equivalence class of CGs in the AMP or MVR CG interpretation, there exists a unique minimal (w.r.t. inclusion) set of non-directed edges that is shared by all members of the class.*

For the AMP interpretation the correctness of Theorems 4 and 5 follow directly from Lemma 13. For the MVR CG interpretation the proofs are given previously by Sonntag and Peña [18, Theorem 1 and Theorem 2]. To see that the theorems do not hold for the LWF CG interpretation consider the LWF CGs $X \rightarrow Y \rightarrow Z - W \leftarrow X$ and $X \rightarrow Y - Z \leftarrow W \leftarrow X$. No split is feasible in either CG and even though they represent the same independence model they do not have the same set of undirected edges.

5. Translations between interpretations

As noted in the introduction most papers that have studied CGs and the independence models they represent have studied the different CG interpretations independently. There are few exceptions to this, such as the study of discrete CG models by Drton [8] and the study of CGs representing Gaussian distributions by Wermuth et al. [23].

Therefore it has not really been studied what differences and similarities that exist between the different interpretations in terms of representable independence models. Andersson et al. made a small study of this when they presented their new (AMP) interpretation and managed to show when the independence model of a CG in the AMP interpretation could be represented perfectly by a CG in the LWF interpretation [2]. They did however not show when the opposite held and did not do any comparison with CGs in the MVR interpretation. Similarly, Wermuth and Sadeghi presented the conditions for when the independence model of a CG in the MVR interpretation could be represented by a CG in the LWF or AMP interpretation when they introduced regression graphs [22]. The conditions stated were however only necessary and sufficient if the two CGs contained the same connectivity components and not the more general case when the CGs can take any form.

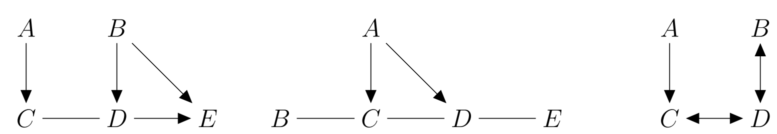
In this section we therefore identify the necessary and sufficient conditions for when a CG in one interpretation can be perfectly translated into a CG in another interpretation. By *translate*, we mean that the independence model represented by a CG in one interpretation can be represented perfectly by a CG in another interpretation. A summary of these results is presented in Table 1 while the actual theorems and their proofs are shown in Appendix C.

	LWF	AMP	MVR
LWF	-	?	$(G_{cl(K)})^m$ is chordal for all $K \in cc(G)$.
AMP	G contains no k-biflag where $k \geq 2$ [2]	-	G' does not contain any induced subgraph of the form $X - Y - Z$
MVR	G' contains no bidirected edge	G' contains no bidirected flag	-

Table 1: Given a CG G in the interpretation denoted in the row, and a maximally oriented CG G' in the Markov equivalence class of G, there exists a CG H in the interpretation denoted in the column s.t. G and H are Markov equivalent iff the condition in the intersecting cell is fulfilled.

From the table two things can be noted. First that the conditions given in the table may include a maximally oriented CG G' in the same Markov equivalence class as G. This is done for several reasons. First, such a graph is easy and computationally simple to find. Secondly, this allows the proofs to be based on the idea that no feasible split is possible for the interpretation in mind. Third and last, the search space of CGs is smaller and more assumptions can be made on the CG. This in turn allows for more efficient algorithms when calculating if the condition holds for some CG. The second note that can be made is that there still does not exist any necessary and sufficient condition for when a translation of a LWF CG G into an AMP CG H is possible. Andersson et al. gave a necessary condition but also showed that this condition was not sufficient [2]. We have managed to prove the necessity of more elaborate conditions but still been unable to prove sufficiency for these. Hence this condition is left for future work.

To exemplify the conditions we can look at the CGs shown in Figure 7. We can here see that the LWF CG G shown in Figure 7a contains five components, $\{A\}$, $\{B\}$, $\{C, D\}$ and $\{E\}$. To see if G is transferable to the MVR CG interpretation we can now check whether the moral closure for each component is chordal or not. It is then clear that the moral closure of the component containing E has the structure $B-D-E-B$ and hence is chordal. However, if we look at the moral closure of the component containing C and D it has the structure $A-C-D-B-A$ which is a non-chordal cycle. This means that G is not transferable to the MVR CG interpretation. For the AMP CG H shown in Figure 7b we can immediately see that it is not transferable to the LWF CG interpretation since it contains a k-biflag with $k = 4$. To see whether it can be transferable to the MVR CG interpretation we first have to find a maximally oriented version of H. In this case H is such a graph and since H contains an induced subgraph of the form $B-C-D$ we know that it is not transferable. Unlike the previous graphs the MVR CG F shown in Figure 7c is not maximally oriented. To see this we can note that a split is feasible with B in U and $\{C, D\}$ in L. The resulting MVR CG F', which is maximally oriented, does then have the following structure $A \rightarrow C \leftrightarrow D \leftarrow B$. This means that F is not transferable to the LWF CG interpretation since F' contains a bidirected edge, but that it is transferable to the AMP CG interpretation since F' contains no bidirected flag.



(a) A LWF CG G. (b) An AMP CG H. (c) A MVR CG F.

Figure 7: Examples of transferability.

6. Extension of Meek's conjecture

Meek's conjecture states that given two DAGs G and H, s.t. $I(H) \subseteq I(G)$, G can be transformed into H through a sequence of operations s.t., after each operation, G is a DAG and $I(H) \subseteq I(G)$. The operations consist in adding a single directed edge to G, or replacing G with a Markov equivalent DAG. The conjecture was proven to be valid by Chickering in 2002 [4, Theorem 4] who gave a constructive proof, i.e. an algorithm that constructs a valid sequence of operations for any DAGs G and H. Hence, strictly speaking Meek's conjecture is really a theorem, but since the statement is known as

Meek's conjecture we will use that term in this article. Using the conjecture the correctness could be proven for several structure learning algorithms for DAGs that only require the probability distribution p of the data to fulfill the graphoid properties and the composition property [5, 13]. These algorithms can be seen as to consist of two phases: A first phase that starts from the empty graph H and adds single edges to it until p is Markovian with respect to H, and a second phase that removes single edges from H until p is Markovian with respect to H and p is not Markovian with respect to any DAG F s.t. $I(H) \subseteq I(F)$. The success of the first phase is guaranteed by the composition property assumption, whereas the success of the second phase is guaranteed by the validity of Meek's conjecture.

Having similar structure learning algorithms for CGs is of course desirable. Hence the conjecture was extended to LWF CGs by Peña et al. in 2014 [16]. The authors stated the following: Given two LWF CGs G and H, s.t. $I(H) \subseteq I(G)$, G can be transformed into H through a sequence of operations s.t., after each operation, G is a LWF CG and $I(H) \subseteq I(G)$. The operations do in this case consist of adding a single directed edge to G, adding a single non-directed edge to G, or replacing G with a Markov equivalent LWF CG. The authors then proved that this conjecture held through a constructive proof. Moreover, they showed that this extended conjecture allowed for the construction of structure learning algorithms that only require the data to fulfill the graphoid properties and the composition property. This was done by introducing and proving the correctness of such an algorithm [16].

Given the definition of the extended Meek's conjecture for LWF CGs it is easy to see what it would look like for the AMP and MVR CG interpretations, the only thing that changes is that AMP resp. MVR CGs are considered instead of LWF CGs. In 2012 Peña did however show that such an extension of Meek's conjecture does not hold for AMP CGs [14]. For MVR CGs the conjecture has to our knowledge not been studied previously though we can here show an example that proves it does not hold. Consider the conjecture to be: Given two MVR CGs G and H, s.t. $I(H) \subseteq I(G)$, G can be transformed into H through a sequence of operations s.t., after each operation, G is a MVR CG and $I(H) \subseteq I(G)$. The operations do in this case consist of adding a single directed edge to G, adding a single bidirected edge to G, or replacing G with a Markov equivalent MVR CG. We can then study the independence models of the MVR CGs G and H shown in Figure 8.

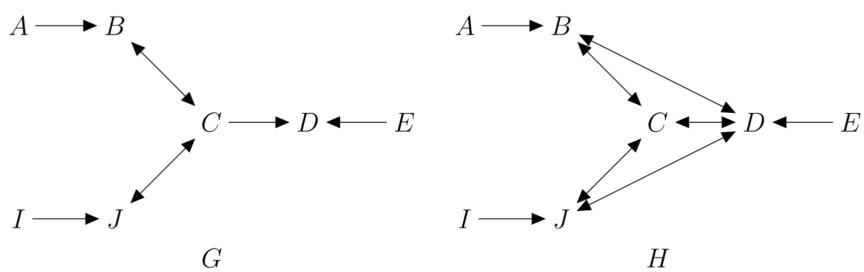


Figure 8: Two MVR CGs G and H s.t. $I(H) \subset I(G)$.

To see that $I(H) \subseteq I(G)$ holds we can list all separators between any pair of distinct nodes for the two CGs. By $\mathcal{S}^{Z}_{XY}$ we mean all the sets of nodes S for which $X \perp\!\!\!\perp_Z Y | S$ holds in an AMP CG Z. Specifically,

- $\mathcal{S}^{H}_{AB} = \mathcal{S}^{G}_{AB} = \mathcal{S}^{H}_{BC} = \mathcal{S}^{G}_{BC} = \mathcal{S}^{H}_{CD} = \mathcal{S}^{G}_{CD} = \mathcal{S}^{H}_{DE} = \mathcal{S}^{G}_{DE} = \mathcal{S}^{H}_{IJ} = \mathcal{S}^{G}_{IJ} = \mathcal{S}^{H}_{JC} = \mathcal{S}^{G}_{JC} = \emptyset$,
- $\mathcal{S}^{H}_{AC}$ = all the node sets that do not contain $\{B\} = \mathcal{S}^{G}_{AC}$,
- $\mathcal{S}^{H}_{AD}$ = all the node sets that do not contain $\{B\} \subset \mathcal{S}^{G}_{AD}$ = all the node sets that do not contain $\{B\}$ or that contain $\{C\}$,
- $\mathcal{S}^{H}_{AE}$ = all the node sets that do not contain $\{B, D\} \subset \mathcal{S}^{G}_{AE}$ = all the node sets that do not contain $\{B, D\}$ or that contain $\{C\}$,
- $\mathcal{S}^{H}_{AI}$ = all the node sets that contain neither $\{B, C, J\}$ nor $\{B, D, J\} \subset \mathcal{S}^{G}_{AI}$ = all the node sets that do not contain $\{B, C, J\}$,
- $\mathcal{S}^{H}_{AJ}$ = all the node sets that contain neither $\{B, C\}$ nor $\{B, D\} \subset \mathcal{S}^{G}_{AJ}$ = all the node sets that do not contain $\{B, C\}$,
- $\mathcal{S}^{H}_{BD} = \emptyset \subset \mathcal{S}^{G}_{BD}$ = all the node sets that contain $\{C\}$
- $\mathcal{S}^{H}_{BE}$ = all the node sets that do not contain $\{D\} \subset \mathcal{S}^{G}_{BE}$ = all the node sets that do not contain $\{D\}$ or that contain $\{C\}$,
- $\mathcal{S}^{H}_{BI}$ = all the node sets that contain neither $\{C, J\}$ nor $\{D, J\} \subset \mathcal{S}^{G}_{BI}$ = all the node sets that do not contain $\{C, J\}$,
- $\mathcal{S}^{H}_{BJ}$ = all the node sets that contain neither $\{C\}$ nor $\{D\} \subset \mathcal{S}^{G}_{BJ}$ = all the node sets that do not contain $\{C\}$,
- $\mathcal{S}^{H}_{CE}$ = all the node sets that do not contain $\{D\} = \mathcal{S}^{G}_{CE}$,
- $\mathcal{S}^{H}_{CI}$ = all the node sets that do not contain $\{J\} = \mathcal{S}^{G}_{CI}$,
- $\mathcal{S}^{H}_{DI}$ = all the node sets that do not contain $\{J\} \subset \mathcal{S}^{G}_{DI}$ = all the node sets that do not contain $\{J\}$ or contain $\{C\}$,

- $\mathcal{S}^H_{DJ} = \varnothing \subset \mathcal{S}^G_{DJ}$ = all the node sets that contain $\{C\}$
- $\mathcal{S}^H_{EI}$ = all the node sets that do not contain $\{D, J\} \subset \mathcal{S}^G_{EI}$ = all the node sets that do not contain $\{D, J\}$ or contain $\{C\}$,
- $\mathcal{S}^H_{EJ}$ = all the node sets that do not contain $\{D\} \subset \mathcal{S}^G_{EJ}$ = all the node sets that do not contain $\{D\}$ or contain $\{C\}$.

Then, since $\mathcal{S}^H_{XY} \subseteq \mathcal{S}^G_{XY}$ for all $X, Y \in \{A, B, C, D, E, I, J\}$ with $X \neq Y$ we know that $I(H) \subseteq I(G)$.

Let $\mathcal{G}$ resp. $\mathcal{H}$ denote the Markov equivalence class of G resp. H. We then know that any CG in $\mathcal{G}$ resp. $\mathcal{H}$ must take the form of the corresponding CG shown in Figure 9, where a circle at the end of an edge represents an unspecified end, i.e. an arrowhead or nothing.

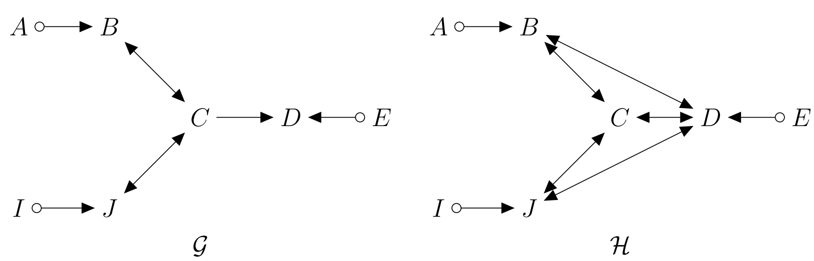


Figure 9: All MVR CGs in $\mathcal{G}$ resp. $\mathcal{H}$.

However, we cannot transform any CG in $\mathcal{G}$ into a MVR CG in $\mathcal{H}$ as required by Meek's conjecture. To see this, note that adding any edge to any CG in $\mathcal{G}$ between two non-adjacent nodes in $\mathcal{H}$ gives that $I(H) \subseteq I(G)$ does not hold. Hence the only modifications that we can perform to any MVR CG in $\mathcal{G}$ is adding the edge $B \to D$ or the edge $J \to D$. This does however imply that $A \not\perp\!\!\!\perp D$ or $I \not\perp\!\!\!\perp D$ hold in the resulting MVR CG, whereas $A \perp\!\!\!\perp D$ and $I \perp\!\!\!\perp D$ hold in any MVR CG in $\mathcal{H}$.

7. Summary and conclusion

In this article we have covered different concepts of CGs that in different ways connect to their representable independence models. We have studied what results there exist for the different CG interpretations in previous research and contributed in different ways when results have been missing. All the areas we have covered do now have results for all three CG interpretations. Hence our hope is that this article can work as a coherent overview of how the concepts discussed here are applied for the different CG interpretations.

Acknowledgments

This work is funded by the Center for Industrial Information Technology (CENIIT) and a so-called career contract at Linköping University, by the Swedish Research Council (ref. 2010-4808), and by FEDER funds and the Spanish Government (MICINN) through the project TIN2010-20900-C04-03.

Appendix A: Lemmas and proofs for Section 3

In this appendix we prove the correctness of Algorithm 1.

Lemma 6. *After line 3, G and G^* have the same adjacencies.*

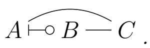
Lemma 7. *After line 6, all the blocks in G^* are on edge ends that are not arrowheads in G.*

Proof. It has been proven that any of the rules R1-R4 only blocks edge ends that are not arrowheads in G [15, Lemma 3]. Of course, for this to hold, the blocks in the antecedent of the rule must be on edge ends that are not arrowheads in G. This implies that, after line 4, all the blocks in G^* are on edge ends that are not arrowheads in G, because G^* has no blocks before line 4. However, to prove that this result also holds after line 6, we have to prove that line 5 only blocks edge ends that are not arrowheads in G. To do so, consider any cycle ρ^* in G^* that is of length greater than three, chordless, and without blocks. Let ρ denote the cycle in G corresponding to the sequence of nodes in ρ^*. Note that no (undirected) edge in ρ^* can be directed in ρ because, otherwise, a subroute of the form $A \rightarrow B \circ\!\!- C$ must exist in ρ, which implies that G contains the triplex $A \rightarrow B \circ\!\!- C$ because A and C cannot be adjacent in G since ρ^* is chordless, which implies that $A \vdash\!\!\circ B \circ\!\!\dashv C$ is in G^* by R1 in line 4, which contradicts that ρ^* has no blocks. Therefore, every edge in ρ^* is undirected in ρ and, thus, line 5 only blocks edge ends that are not arrowheads in G. □

Lemma 8. *After line 7, G and G^* have the same triplexes. Moreover, G^* has all the immoralities that are in G.*

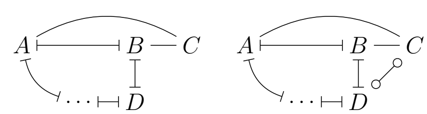
Proof. The proof is essentially the same as that of Lemma 4 in [15]. □

Lemma 9. *After line 6, G^* does not have any induced subgraph of the form*



Proof. The proof is essentially the same as that of Lemma 5 in [15]. It just suffices to add the following case:

Case 0 Assume that $A \vdash\!\!\dashv B$ is in H due to line 5. Then, after line 5, H had an induced subgraph of one of the following forms, where possible additional edges between C and internal nodes of the route $A \vdash\!\!\dashv \ldots \vdash\!\!\dashv D$ are not shown:



case 0.1 case 0.2

Note that C cannot belong to the route $A \vdash\!\circ \ldots \vdash\!\circ D$ because, otherwise, the cycle $A \vdash\!\circ \ldots \vdash\!\circ D \vdash\!\circ B \multimap A$ would not have been chordless.

Case 0.1 If $B \notin S_{CD}$ then $B \dashv C$ is in H by R1, else $B \vdash C$ is in H by R2. Either case is a contradiction.

Case 0.2 Recall from line 5 that the cycle $A \vdash\!\!\dashv \ldots \vdash\!\!\dashv D \vdash\!\!\dashv B \vdash\!\!\dashv A$ is of length greater than three and chordless, which implies that there is no edge between A and D in H. Thus, if $C \notin S_{AD}$ then $A \vdash C$ is in H by R1, else $B \dashv C$ is in H by R4. Either case is a contradiction.

□

Lemma 10. *After line 6, every chordless cycle $\rho^* : V_1, \ldots, V_n = V_1$ in G^* that has an edge $V_i \vdash V_{i+1}$ also has an edge $V_j \dashv V_{j+1}$.*

Proof. The proof is essentially the same as that of Lemma 6 in [15]. □

Theorem 11. *After line 7, G^* is the essential graph in the class of triplex equivalent CGs containing G.*

Proof. Using Theorem 1 stated by Peña [15] it follows that after line 7, G^* is Markov equivalent to G and it has no semi-directed cycles. Moreover, the directed edges in G^* after line 7 must be directed in the essential graph in the class of triplex equivalent CGs containing G by Lemma 7. For the same reason, the undirected edges in G^* after line 7 that correspond to $A \vdash\!\dashv B$ edges when line 7 was to be executed must be undirected in the essential graph in the class of triplex equivalent CGs containing G. We show below that every other undirected edge in G^* after line 7 (i.e. those that correspond to edges without blocks when line 7 was to be executed) must also be undirected in the essential graph in the class of triplex equivalent CGs containing G.

Let H denote the graph that contains all and only the edges of G^* resulting from the replacements in line 7, and let U denote the graph that contains the rest of the edges of G^* after line 7. Note that all the edges in U are undirected and they had no blocks when line 7 was to be executed. Therefore, U has no cycle of length greater than three that is chordless by line 5. In other words, U is chordal. Then, we can orient all the edges in U without creating immoralities nor directed cycles by using, for instance, the maximum cardinality search (MCS) algorithm [10, p. 312]. Consider any such orientation of the edges in U and denote it D. Now, add all the edges in D to H. As we show below, this last step does not create any triplex or semi-directed cycle in H:

- It does not create a triplex $(\{A, C\}, B)$ in H because, otherwise, $A{-}B \circ\!\dashv C$ must exist in G^* when line 7 was to be executed, which implies that $A \vdash\!\circ B$ or $A \circ\!\dashv B$ was in G^* by R1 or R2 when line 7 was to be executed, which contradicts that $A{-}B$ is in U.

- Assume to the contrary that it does create a semi-directed cycle in H. Note that this cycle cannot have any $\vdash$ edge by Lemma 10 when line 7 was to be executed and, thus, it must have some $\vdash\!\dashv$ edge when line 7 was to be executed. However, this implies that $A{-}B \vdash\!\dashv C$ must exist in G^* when line 7 was to be executed, which implies that A and C are adjacent in G^* because, otherwise, $A \vdash\!\circ B$ or $A \circ\!\dashv B$ was in G^* by R1 or R2 when line 7 was to be executed, which contradicts that $A{-}B$ is in U. Then, $A \vdash\!\circ C$ or $A \circ\!\dashv C$ exist in G^* by Lemma 9 when line 7 was to be executed, which implies that $A \vdash\!\circ B$ or $A \circ\!\dashv B$ was in G^* by R3 when line 7 was to be executed, which contradicts that $A{-}B$ is in U.

Consequently, H is a CG that is triplex equivalent to G. Finally, let us recall how the MCS algorithm works. It first unmarks all the nodes in U and, then, iterates through the following step until all the nodes are marked: Select any of the unmarked nodes with the largest number of marked neighbors and mark it. Finally, the algorithm orients every edge in U away from the node that was marked earlier. Clearly, any node may get marked firstly by the algorithm because there is a tie among all the nodes in the first iteration, which implies that every edge may get oriented in any of the two directions in D and thus in H. Therefore, either orientation of every edge of U occurs in some CG H that is triplex equivalent to G. Then, every edge of U must be undirected in the essential graph in the class of triplex equivalent CGs containing G. □

Appendix B: Lemmas and proofs for Section 4

In this appendix we prove that the feasible split and feasible merging for AMP CGs are sound.

Lemma 1. A CG G in the AMP interpretation is in the same Markov equivalence class before and after a feasible split.

Proof. Assume the contrary. Let G be a CG under the AMP interpretations and G' a graph s.t. G' is G with a feasible split performed upon it. G and G' are in different Markov equivalence classes or G' is not a CG under the AMP interpretation iff (1) G and G' does not have the same adjacencies, (2) G and G' does not have the same triplexes or (3) G' contains semi-directed cycles.

First it is clear that G and G' contain the same adjacencies since a feasible split does not change the adjacencies of any node in G. We do also know that conditions 1, 2 and 3 in Definition 9 must be fulfilled for G. Secondly let us assume G and G' do not have the same triplexes. First let us assume that G' contains a triplex $(\{X, Y\}, Z)$ that does not exist in G. Such a triplex can only occur if $Z \in L$ since the only difference between G and G' is that G' contains some directed edges oriented towards L where G contains undirected edges. If the triplex is a flag then the one of the nodes X or Y, let us say X, must be in U and the other one, let us say Y, must be in L. However, according to condition 1 for the feasible split Y must be adjacent to X which gives a contradiction. If the triplex is not a flag both X and Y must be in U. They also have to be in $nb_G(L)$, which, together with condition 2, contradicts that they are not adjacent. Hence we have a contradiction for that G' contains a triplex that does not exist in G.

Secondly assume G contains a triplex $(\{X, Y\}, Z)$ that does not exist in G'. This new triplex cannot be over a node in L since these nodes only have edges oriented towards them. Hence we have that $Z \in U$. This gives that one of the nodes X or Y, let us say X, must be a parent of Z and the other, let us say Y, must be in L. This does however contradict condition 3, since every parent of Z also must be a parent of Y, and hence X and Y must be adjacent. This gives us a contradiction.

Finally assume that a semi-directed cycle is introduced. This can happen iff we have two nodes X and Y s.t. $X \in de_{G'}(Y), X \in U$ and $Y \in L$. We know no semi-directed cycle existed in G before the split and that $de_{G'}(Y) \subseteq de_G(Y) \setminus U$. This together with the fact that $X \notin de_G(Y) \setminus U$ then gives a contradiction. □

Lemma 2. A CG G in the AMP interpretation is in the same Markov equivalence class before and after a feasible merging.

Proof. Assume the contrary. Let G be a CG under the AMP interpretation and G' a graph s.t. G' is G with a feasible merging performed upon it. G and G' are in different Markov equivalence classes or G' is not a CG under the AMP interpretation iff (1) G and G' does not have the same adjacencies, (2) G and G' does not have the same triplexes or (3) G' contains semi-directed cycles.

First it is clear that G and G' contain the same adjacencies since a feasible merging does not change the adjacencies of any node in G. It must also be the case that the conditions in Definition 10 must hold in G. Secondly let us assume G and G' do not have the same triplexes. First let us assume that G contains a triplex $(\{X, Y\}, Z)$ that does not exist in G'. Such a triplex can only occur if $Z \in L$ since the only difference between G and G' is that G contains some directed edges oriented towards L where G' contains undirected edges. If the triplex is a flag in G then one of the nodes X or Y, let us say X, must be in U and the other one, let us say Y, must be in L. However, according to condition 1 for the feasible merging Y must be adjacent to X which gives a contradiction. If the triplex is not a flag both X and Y must be in U. They also have to be in $pa_G(L)$, which, together with condition 2, contradicts that they are not adjacent. Hence we have a contradiction for that G contains a triplex that does not exist in G'.

Secondly assume G' contains a triplex $(\{X, Y\}, Z)$ that does not exist in G. This new triplex cannot be over a node in L since any new undirected edge with a node in L as endnode must previously have been a directed edge oriented towards L. Hence we know that $Z \in U$. This gives that one of the nodes X or Y, let us say X, must be a parent of Z and the other, let us say Y, must be in L. This does however contradict condition 3, since every parent of Z also must be a parent of Y, and hence X and Y must be adjacent. This gives us a contradiction.

Finally assume G' contains a semi-directed cycle. This means that we have three nodes X, Y and Z s.t. $X \in L$, $Z \in U$, $Y \notin U \cup L$, $Z{-}X{\leftarrow}Y$ is in G' and $Y \in de_{G'}(Z)$. However, this means that $Y \in pa_G(X)$ and $Y \in de_G(U)$ must hold which violates condition 4 in Definition 10 and hence we have a contradiction. □

Lemma 12. *If an AMP CG G is transformed into an AMP CG H through a feasible split then there exists a feasible merging that transforms H into G and vice versa.*

Proof. We will start by proving that a merging transforming H into G must be feasible if G was transformed into H with a feasible split. Assume the contrary. Then we know that G and H have the same structure with the exception that some undirected edges $X{-}Y$ in G are replaced by directed edges $X{\rightarrow}Y$ in H. Let U and L be the same set of nodes as during the feasible split. Then we have that U and L must belong to the same connectivity component C in G but to different connectivity components in H. We also have that $C = U \cup L$. Now, if a merging is feasible in H with this U and L we have a contradiction. Hence one of the conditions in Definition 10 must fail in H. Now, since $pa_H(L) \cap U = nb_G(L) \cap U$, it is straightforward to see that condition 2 must hold since we know that $nb_G(L) \cap U$ is complete from condition 2 in the previous split. For condition 3 we also know from condition 3 in the previous split that $\forall B \in L, pa_G(nb_G(L) \cap U) \subseteq pa_G(B)$. Hence, since $pa_H(L) \cap U = nb_G(L) \cap U$, $pa_G(nb_G(L) \cap U) = pa_H(pa_H(L) \cap U)$ and that $\forall B \in L, pa_G(B) = pa_H(B) \setminus U$ we must have that $\forall B \in L, pa_G(pa_H(L) \cap U) \subseteq pa_H(B)$, i.e. condition 3, must hold. For condition 1 we must have that $\forall A \in nb_G(L) \cap U, L \subseteq nb_G(A)$ must hold or the previous split would not have been feasible. This together with $nb_G(L) \cap U = pa_H(L) \cap U$ and that $\forall A \in nb_G(L) \cap U, nb_G(A) \cap L = ch_H(A) \cap L$ then gives that $\forall A \in pa_H(L) \cap U, L \subseteq ch_H(A)$ must hold. Finally, condition 4 must hold or a semi-directed cycle exists in G. Hence all conditions must be fulfilled and the merging must be feasible.

Secondly assume that a split transforming H into G is not feasible when G was transformed into H with a feasible merging. Let U and L denote the same U and L as for the previous feasible merging. Then we know that a split with these sets of nodes will transform H into G and hence that the split cannot be feasible. Hence, to not have a contradiction one of the conditions in Definition 9 must fail. To see that condition 2 must hold we once again have that $pa_G(L) \cap U$ is complete or condition 2 would have failed for the previous merging. We also have that $pa_G(L) \cap U = nb_H(L) \cap U$ and hence that $nb_H(L) \cap U$ is complete which means condition 2 must hold for the split. For condition 3 we can note that $\forall B \in L, pa_G(pa_G(L) \cap U) \subseteq pa_G(B)$ must hold due to the previous merging. Hence, since $pa_G(L) \cap U = nb_H(L) \cap U$ it follows that $pa_G(pa_G(L) \cap U) = pa_H(nb_H(L) \cap U)$. Moreover it must be the case that $\forall B \in L, pa_H(B) = pa_G(B) \setminus U$ and $pa_H(nb_H(L) \cap U) \cap U = \varnothing$. This means that $\forall B \in L, pa_H(nb_H(L) \cap U) \subseteq pa_H(B)$ must hold. Hence condition 3 must hold. Finally for condition 1 we have that $\forall A \in pa_G(L) \cap U, L \subseteq ch_G(A)$ or the previous merging would not have been feasible. This together with $pa_G(L) \cap U = nb_H(L) \cap U$ and $\forall A \in nb_H(L) \cap U, ch_G(A) \cap L = nb_H(A) \cap L$ then gives that $\forall A \in nb_H(L) \cap U, L \subseteq nb_H(A)$ must hold. Hence all conditions must be fulfilled and the split must be feasible. □

Theorem 3. Given two AMP CGs G and H in the same Markov equivalence class there exists a sequence of feasible splits and mergings that transforms G into H.

Proof. Since we know that any merging is reversible with a split and vice versa we only have to show that the largest deflagged graph G^* is reachable from any AMP CG G. From Theorems 4 and 5 we get that a maximally oriented AMP CG G' is reachable from G through a sequence of feasible splits and that G' contains the unique minimal set of undirected edges. Hence we know that G' must be in the largest strong Markov equivalence class in the Markov equivalence class of G. Now, to reach G^* from G' we only have to make a set of legal mergings [17]. A legal merging replaces directed edges between two components with undirected edges similarly as a feasible split. The conditions for a merging of two components U and L to be legal are the following: (1) $\forall A \in L, pa_G(L) = pa_G(A)$, (2) $pa_G(L) \cap U$ is complete in G and (3) $\forall B \in pa_G(L) \cap U, pa_G(L) \setminus U = pa_G(B)$. We can now see that if a merging is legal for two components U and L, then this implies that the same merging must also be feasible. Hence the largest deflagged graph G^* is reachable from two AMP CGs G and H in the same Markov equivalence class. From G^* we then know that there exists a sequence of feasible splits and mergings to reach H. This sequence is simply the reverse inverse sequence of feasible splits and mergings to reach G^* from H, i.e. where all splits have been replaced by mergings and vice versa and the order has been reversed. Hence there must exist a sequence of feasible splits and mergings that transforms G into H. □

Lemma 13. *If no split is feasible in an AMP CG G then there exists no other AMP CG H s.t. $I(H) = I(G)$ and H has a directed edge $X \rightarrow Y$ where G has an undirected edge $X - Y$.*

Proof. Assume this is not the case. We then know X and Y are in two different components in H and that G and H have the same adjacencies. Let us choose X and Y s.t. there exist no nodes Z, W s.t. $Z \in de_H(Y) \cup Y$ and $Z - W$ exists in G but $Z \rightarrow W$ exists in H. Since H contains no semi-directed cycles we know that we must be able to choose such X and Y. We can then let C be the component of X and Y in G and $L = co_H(Y) \cap C$. We can also let $U = C \setminus L$.

For a split not to be feasible in G with this U and L we know that one of the conditions in Definition 9 must fail in G. Assume conditions 1 fails. It must then in G exist an induced subgraph of the form $u_i - l_j - l_k$ s.t. $u_i \in U$ and $l_j, l_k \in L$. Hence $u_i \perp\!\!\!\perp_G l_k | l_j \cup nb_G(l_k) \cup pa_G(C)$ must hold. For the same to hold in H, where the edge $u_i \rightarrow l_j$ exists, we also must also have that $l_j \rightarrow l_k$ exists, which is a contradiction since we then would have chosen l_j as X and l_k as Y. Assume condition 2 fails. It must then in G exist an induced subgraph of the form $u_i - l_j - u_k$ s.t. $u_i, u_k \in U$ and $l_j \in L$. Once again we have that $u_i \perp\!\!\!\perp_G u_k | l_j \cup nb_G(u_k) \cup pa_G(C)$ must hold. For the same to hold in H, where the edge $u_i \rightarrow l_j$ exists, we also must also have that $l_j \rightarrow u_k$ exists, which is a contradiction since we then would have chosen l_j as X and u_k as Y. Hence condition 3 must fail. We then must have that G must contain an induced subgraph of the form $P \rightarrow u_i - l_j$ s.t. $P \in pa_G(C)$, $u_i \in U$ and $l_j \in L$. Hence a triplex $(\{P, l_j\}, u_i)$ must exist in G. We can then see that for H, which instead contains the induced subgraph $P \multimap\!\!\!\!\circ u_i \rightarrow l_j$, no such triplex can exist, which is a contradiction. Hence the split must be feasible in G with the defined U and L which contradicts the assumption. □

Appendix C: Theorems, lemmas and proofs for Section 5

In this Appendix we prove that the conditions given in Table 1 are necessary and sufficient with the exception of when an AMP CG can be translated to a LWF CG since this has been shown before [2].

Translation of MVR CGs to AMP CGs

Theorem 14. *Given a MVR CG G, and a maximally oriented MVR CG G' in the Markov equivalence class of G, there exists an AMP CG H s.t. $I_{MVR}(G) = I_{AMP}(H)$ iff G' contains no bidirected flag.*

Proof. Sufficiency follows from from Lemmas 17 and 18 and necessity follows from Lemma 15. □

Lemma 15. *A MVR CG G and an AMP CG H with the same structure, except that every bidirected edge in G is replaced by an undirected edge in H and where G contains no bidirected flag, represent the same independence model.*

Proof. Assume to contrary that there exist two CGs, G under the MVR interpretation and H under the AMP interpretation, s.t. G does not contain any bidirected flag, i.e. induced subgraph of the form $X \leftrightarrow Y \leftrightarrow Z$, G and H contain the same directed edges, and for every bidirected edge in G H has an undirected edge instead (and only contains those undirected edges) but $I_{MVR}(G) \neq I_{AMP}(H)$. Clearly we must have $V_G = V_H$ and that $ad_G(X) = ad_H(X), pa_G(X) = pa_H(X)$ and $co_G(X) = co_H(X)$ holds for all $X \in V_G$. Given the definition of strict descendants $san_G(X) = san_H(X)$ must also hold. Moreover note that H cannot contain any induced subgraph of the form $X - Y - Z$. Finally note that both G and H contains the same paths between any pair of nodes X and Y.

For $I(G) \neq I(H)$ to hold there has to exist a path π in G (resp. H) that is d-connecting (resp. S-open) s.t. there exists no path in H (resp. G) that is S-open (resp. d-connecting). Let π be a minimal d-connecting (resp. S-open) path in G (resp. H). Note that π cannot contain any subpath of the form $V_1 \leftrightarrow V_2 \leftrightarrow V_3$ (resp. $V_1 - V_2 - V_3$) since the edge $V_1 \leftrightarrow V_3$ (resp. $V_1 - V_3$) must exist in G (resp. H) or G contains a bidirected flag or semi-directed cycle. This in turn would mean that π is not minimal since the path $\pi \setminus V_2$ also must be d-connecting and shorter than π. For π to be both d-connecting and S-open for any set of nodes Z it must contain the same colliders and head-no-tail nodes. A node $W \in \pi$ is a collider if it is part of the following configurations of edges in π (1) $\rightarrow W \leftarrow$, (2) $\leftrightarrow W \leftarrow$, (3) $\rightarrow W \leftrightarrow$ and (4) $\leftrightarrow W \leftrightarrow$. Clearly the fourth case cannot occur. Case 1-3 would be translated into (1) $\rightarrow W \leftarrow$, (2) $- W \leftarrow$, (3)$\rightarrow W -$ in H which are all (and the only) head-no-tail configurations. Hence π must be d-connecting in G iff π is S-open in H which contradicts the assumption. □

Lemma 16. *If a maximally oriented CG G in the MVR interpretation contains a bidirected flag $X \leftrightarrow Y \leftrightarrow Z$ then G also contains an induced subgraph of the form shown in (1) Figure 10a, (2) 10b, (3) $P \circ\!\!\rightarrow Q \leftrightarrow Y \leftrightarrow Z$, (4) $P \circ\!\!\rightarrow Q \leftrightarrow W \leftrightarrow Z$ s.t. $bd_G(Q) \subseteq bd_G(Y) \cup Y$ and $Y \in sp_G(Q)$ hold.*

Proof. Assume the contrary, that no such induced subgraph exists in G even though G contains a bidirected flag and G is maximally oriented. Let C be the component to which X, Y and Z belong. Let A be the set of nodes A_k s.t. $A_k \in sp_G(Y)$ but $A_k \notin sp_G(Z)$. We know that X fulfills these criteria and hence $|A| \geq 1$.

First note that if there exists a node $A_k \in A$ s.t. $bd_G(A_k) \nsubseteq bd_G(Y) \cup Y$ then there exists an induced subgraph $P \circ\!\!\rightarrow A_k \leftrightarrow Y \leftrightarrow Z \cdots P$ in G for some node $P \in bd_G(A_k) \setminus bd_G(Y) \setminus Y$. Hence we have a contradiction since G either contains an induced subgraph of the form shown in Figure 10b ($P \in bd_G(Z)$) or of the form $P \circ\!\!\rightarrow Q \leftrightarrow Y \leftrightarrow Z$ ($P \notin bd_G(Z)$). Therefore we must have that $bd_G(A_k) \subseteq Y \cup bd_G(Y)$ holds for all $A_k \in A$, i.e. that $bd_G(A) \subseteq Y \cup bd_G(Y)$ holds.

Secondly note that we can let B be a subset of A s.t. B consists of the nodes in one connected subgraph in the subgraph of G induced by A (any connected subgraph will do). Let D be the set of nodes s.t. $D = sp_G(Y) \cap sp_G(Z) \cap sp_G(A)$. With these sets we know that the spouses of Y can be either adjacent of Z or not, hence $sp_G(Y) \setminus Z = D \cup A$ must hold. This in turn gives that $sp_G(A) = D \cup Y$ and $bd_G(A) \subseteq D \cup Y \cup pa_G(Y)$ since $\forall A_k \in A \; bd_G(A_k) \subseteq Y \cup bd_G(Y)$ holds. Moreover $sp_G(B) \subseteq D \cup Y$ and $bd_G(B) \subseteq D \cup Y \cup pa_G(Y)$ must also hold. Hence, if D is empty then $sp_G(B) = \{Y\}$ and $bd_G(B) \subseteq Y \cup pa_G(Y)$ must hold. This does however lead to a contradiction because a split then is possible s.t. U consists of B and L consists of $C \setminus U$. Hence there has to exists at least one node in D.

Thirdly note that $D \cup Y$ must be complete or the induced subpath $A_k \leftrightarrow D Y_i$

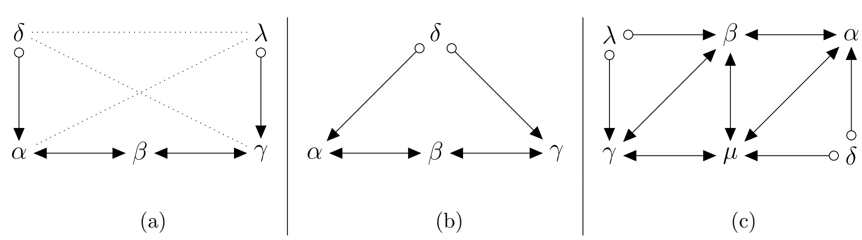


Figure 10: MVR subgraph forms.

$\leftrightarrow Z \leftrightarrow DY_j \leftrightarrow A_l$ exists in G for some nodes $A_k, A_l \in A$ and $DY_i, DY_j \in D \cup Y$. This means that G contains an induced subgraph of the form shown in either Figure 10a ($A_k \neq A_l$) or 10b ($A_k = A_l$).

Fourth and finally note that there must exist a node P s.t. $P \in bd_G(B) \cup B$ but $P \notin bd_G(D_j)$ for some $D_j \in D \cup Y$ or a split is feasible where U consists of B and L consists of $C \setminus U$. Note that $D_j \neq Y$ must hold since $bd_G(B) \cup B \subseteq bd_G(Y) \cup Y$. This means that there must exist two nodes B_i, D_j s.t. $P \in bd_G(B_i), P \notin bd_G(D_j), B_i \in B$, $B_i \in sp(D_j)$ and $D_j \in D$ s.t. the induced subgraph $P \circ\!\!\rightarrow B_i \leftrightarrow D_j \leftrightarrow Z \cdots P$ exists in G. This is a contradiction either because G contains an induced subgraph of the form shown in Figure 10b ($P \in bd_G(Z)$) or $P \circ\!\!\rightarrow B_i \leftrightarrow D_j \leftrightarrow Z$ ($P \notin bd_G(Z)$) where $bd_G(B_i) \subseteq bd_G(Y) \cup Y$ and $Y \in sp_G(B_i)$ holds. □

Lemma 17. *If a maximally oriented CG G in the MVR interpretation contains a bidirected flag then at least one of the induced subgraphs shown in Figure 10 exist in G.*

Proof. Assume the contrary, that no such induced subgraph exists in G even though G contains a bidirected flag and G is maximally oriented. Since G contains a bidirected flag we do with Lemma 16 get that G must contain an induced subgraph $X \leftrightarrow Y \leftrightarrow Z \leftarrow\!\!\circ W$ or a contradiction directly follows. If we now apply Lemma 16 to $X \leftrightarrow Y \leftrightarrow Z$ we get that, since for G to contain any induced subgraph of the form shown in Figure 10a or 10b is a contradiction, there exists a set of nodes (that can be renamed to) c_1, c_2, c_3 s.t. the induced subgraph $c_1 \circ\!\!\rightarrow c_2 \leftrightarrow c_3 \leftrightarrow Z$ exists in G and $c_3 = Y$ holds or $bd_G(c_2) \subseteq bd_G(Y) \cup Y$ and $Y \in sp_G(c_2)$ hold. If $c_3 = Y$, G must contain the subgraph $c_1 \circ\!\!\rightarrow c_2 \leftrightarrow Y \leftrightarrow Z \leftarrow\!\!\circ W$ where $c_1 \notin ad_G(Y)$ and $W \notin ad_G(Y)$ must hold and $c_1 = W$ might hold. Clearly this subgraph takes the form of either Figure 10a ($c_1 \neq W$) or 10b ($c_1 = W$) which is a contradiction. Hence $c_3 \neq Y$, $bd_G(c_2) \subseteq bd_G(Y) \cup Y$ and $Y \in sp_G(c_2)$ must hold.

Since $W \notin ad_G(Y)$ holds and $bd_G(c_2) \subseteq bd_G(Y) \cup Y$ it is clear that $c_1, c_3 \in bd_G(Y)$ must hold. Hence $W \neq c_2$ holds since $W \notin ad_G(Y) \cup Y$. This in turn means that $W \notin bd_G(c_2)$ holds since $bd_G(c_2) \subseteq bd_G(Y) \cup Y$ and $W \notin bd_G(Y) \cup Y$. Finally we can see that $W \in bd_G(c_3)$ holds or the induced subgraph $c_1 \circ\!\!\rightarrow c_2 \leftrightarrow c_3 \leftrightarrow Z \leftarrow\!\!\circ W$ takes the form shown in Figure 10a ($c_1 \neq W$) or 10b ($c_1 = W$). However, if $W \in bd_G(c_3)$ holds G contains an induced subgraph of the form shown in Figure 10c (where $\delta = W$, $\lambda = c_1$, $\mu = c_3$, $\gamma = c_2$, $\beta = Y$ and $\alpha = Z$) and we have a a contradiction. □

Lemma 18. *The independence model of a CG G in the MVR interpretation which contains an induced subgraph of one of the forms shown in Figure 10 cannot be perfectly represented as a CG H in the AMP interpretation.*

Proof. Assume the contrary, that there exists a CG H under the AMP interpretation that can represent these independence models.

First assume that the independence model of the graph shown in Figure 10a can be represented in a CG H in the AMP interpretation. It is clear that H must have the same skeleton, or some separations or non-separations that hold in G would not hold in H. The following independence statements hold in G: $\delta \perp\!\!\!\perp_G \beta | pa_G(\beta)$, $\alpha \perp\!\!\!\perp_G \gamma | pa_G(\alpha)$ and $\beta \perp\!\!\!\perp_G \lambda | pa_G(\beta)$. $\delta \perp\!\!\!\perp_G \beta | pa_G(\beta)$ gives us that a triplex $(\{\delta, \beta\}, \alpha)$ must exist in H, since $\alpha \notin pa_G(\beta)$ i.e. that (1) $\delta \rightarrow \alpha - \beta$, (2) $\delta - \alpha \leftarrow \beta$ or (3) $\delta \rightarrow \alpha \leftarrow \beta$ exists in H. $\alpha \perp\!\!\!\perp_G \gamma | pa_G(\alpha)$ does however also state that a triplex $(\{\alpha, \gamma\}, \beta)$ must exist in H, since $\beta \notin pa_G(\alpha)$. For this to happen the edge between α and β cannot be oriented towards α hence the subgraph $\delta \rightarrow \alpha - \beta \leftarrow \gamma$ must exist in H. The orientation of the edge between β and γ does however contradict the third independence statement $\beta \perp\!\!\!\perp_G \lambda | pa_G(\beta)$ which implies that the triplex $(\{\beta, \lambda\}, \gamma)$ must exist in H, since $\gamma \notin pa_G(\beta)$. Hence we have a contradiction if G contains the induced subgraph shown in Figure 10a.

Secondly assume that the independence model of the graph shown in Figure 10b can be represented in a CG H in the AMP interpretation. It is clear that H must have the same skeleton, or some separations or non-separations that hold in G would not hold in H. The following independence statements must then hold in G: $\delta \perp\!\!\!\perp_G \beta | pa_G(\beta)$ and $\alpha \perp\!\!\!\perp_G \gamma | pa_G(\alpha)$. $\delta \perp\!\!\!\perp_G \beta | pa_G(\beta)$ gives us that two triplexes must exist in H, first $(\{\delta, \beta\}, \alpha)$ and secondly $(\{\delta, \beta\}, \gamma)$, since $\alpha, \gamma \notin pa_G(\beta)$. $(\{\delta, \beta\}, \alpha)$ gives that one of the following configurations must occur in H: (1) $\delta - \alpha \leftarrow \beta$, (2) $\delta \rightarrow \alpha - \beta$ or (3) $\delta \rightarrow \alpha \leftarrow \beta$. However, the independence statement $\alpha \perp\!\!\!\perp_G \gamma | pa_G(\alpha)$ implies that the triplex $(\{\alpha, \gamma\}, \beta)$ must exist in H since $\beta \notin pa_G(\alpha)$. If the triplex $(\{\alpha, \gamma\}, \beta)$ should hold in H the edge between α and β cannot be oriented towards α hence the subgraph $\delta \rightarrow \alpha - \beta \leftarrow \gamma$ must exist in H. The orientation of the edge between β and γ does however contradict the triplex $(\{\delta, \beta\}, \gamma)$ and hence we have a contradiction for the G shown in Figure 10b.

Third and last assume that the independence model of the graph shown in Figure 10c can be represented in a CG H in the AMP interpretation. From the Figure we can read the following independence statements: $\lambda \perp\!\!\!\perp_G \mu | pa_G(\mu)$, $\alpha \perp\!\!\!\perp_G \gamma | pa_G(\alpha), \beta \perp\!\!\!\perp_G \delta | pa_G(\beta)$. It is clear that H must have the same skeleton, or some separations or non-separations that hold in G would not hold in H. $\lambda \perp\!\!\!\perp_G \mu | pa_G(\mu)$ and $\alpha \perp\!\!\!\perp_G \gamma | pa_G(\alpha)$ gives that the triplexes $(\{\lambda, \mu\}, \beta)$ and $(\{\alpha, \gamma\}, \mu)$ must exist in H since $\beta \notin pa_G(\mu)$ and $\mu \notin pa_G(\alpha)$. As seen above this gives that $\lambda \rightarrow \gamma - \mu \leftarrow \alpha$ must exist in H. Similarly $\beta \perp\!\!\!\perp_G \delta | pa_G(\beta)$ and $\lambda \perp\!\!\!\perp_G \mu | pa_G(\mu)$ gives that $\lambda \rightarrow \beta - \mu \leftarrow \delta$ must exist in H. Finally $\alpha \perp\!\!\!\perp_G \gamma | pa_G(\alpha)$ and $\beta \perp\!\!\!\perp_G \delta | pa_G(\beta)$ gives that the triplexes $(\{\alpha, \gamma\}, \beta)$ and $(\{\beta, \delta\}, \alpha)$ must hold in H, since $\beta \notin pa_G(\alpha)$ and $\alpha \notin pa_G(\beta)$, which in turn gives that $\gamma \rightarrow \beta - \alpha \leftarrow \delta$ must exist in H. This does however contradict that H is a CG since the semi-directed cycle $\gamma \rightarrow \beta - \mu - \gamma$ exists in H. Hence we have a contradiction. □

Translation of AMP CGs to MVR CGs

Theorem 19. *Given an AMP CG G, and a maximally oriented AMP CG G' in the Markov equivalence class of G, there exists a CG H s.t. $I_{AMP}(G) = I_{MVR}(H)$ iff G' does not contain any induced subgraph of the form $X - Y - Z$.*

Proof. Sufficiency follows from Lemma 15 while necessity follows from 20. □

Lemma 20. *If a maximally oriented CG G in the AMP interpretation contains an induced subgraph of the form $X - Y - Z$ then there exists no CG H in the MVR interpretation s.t. $I_{AMP}(G) = I_{MVR}(H)$.*

Proof. Assume to the contrary that the lemma does not hold. G and H must then have the same skeleton or some separations in H do not hold in G or vice versa. Let H have a component ordering ord for its components $c_1, ...c_k$ s.t. $ord(c_i) < ord(c_j)$ if c_i is a parent of c_j. Let C be the component of X in G. From the assumption we know that $X \perp\!\!\!\perp_G Z | nb_G(X) \cup pa_G(X \cup nb_G(X))$ holds, where $Y \in nb_G(X)$, and hence that H must contain one of the following induced subgraphs: $X \circ\!\!\rightarrow Y \rightarrow Z$, $X \leftarrow Y \leftarrow\!\!\circ Z$ or $X \leftarrow Y \rightarrow Z$. For any other configuration of edges $X \perp\!\!\!\perp_H Z | nb_G(X) \cup pa_G(X \cup nb_G(X))$ does not hold. Moreover we can generalize the configurations to $X \circ\!\!\rightarrow Y \rightarrow Z$ and $X \leftarrow Y \rightarrow Z$ simply by choosing the nodes to represent X and Z accordingly. Both these structures are included in $X \circ\!\!-\!\!\circ Y \rightarrow Z$ and we will now show that this structure leads to a contradiction if a split is not feasible in G.

The proof is iterative and when a restart is noted this is where the proof restarts. For each restart it will be shown that there must exist a triplet of nodes X, Y, Z s.t. an induced subgraph of the form $X - Y - Z$ exists in G and $X \circ\!\!-\!\!\circ Y \rightarrow Z$ in H. Apart from this we also know that $I_{AMP}(G) = I_{MVR}(H)$ holds and that no split is feasible in G. Let the set U consist of Y and every node connected by a path to Y in the subgraph of G induced by $C \setminus Z$ and the set L consist of $C \setminus U$. This separation of sets gives that $nb_G(U) = Z$. For a split not to be feasible with these sets one of the conditions in Definition 9 must fail:

Case 1, condition 1 or 2 fails. This means that there exist two nodes $W \in C$ and $P \in C$ s.t. the induced subgraph $P{-}Z{-}W$ exists in G. Note that one of the nodes might be Y. This means that $P \perp\!\!\!\perp_G W | nb_G(W) \cup pa_G(W \cup nb_G(W))$ holds, where $Z \in nb_G(W)$ and hence that $P {\circ\!\!\rightarrow} Z {\rightarrow} W$, $P {\leftarrow} Z {\leftarrow\!\!\circ} W$ or $P {\leftarrow} Z {\rightarrow} W$ must exist in H as described above. Without losing generality we can say that either $P {\circ\!\!\rightarrow} Z {\rightarrow} W$ or $P {\leftarrow} Z {\rightarrow} W$ exists in H and that $W \neq Y$ by choosing P and W appropriately. This means that we can restart the proof with the structure $P {\circ\!\!-\!\!\circ} Z {\rightarrow} W$ in H (and $P{-}Z{-}W$ in G). The number of restarts is bounded since (1) the number of nodes in V is bounded and that $ord(co_H(Z)) > ord(co_H(Y))$.

Case 2, condition 1 and 2 hold but condition 3 fails. This means that there exist two nodes $W \in U$ and $P \notin C$ s.t. the induced subgraph $Z{-}W{\leftarrow}P$ exists in G. First let us cover the case where $W = Y$. This means that $Z \perp\!\!\!\perp_G P | pa_G(Z)$ holds. Since H have the same skeleton as G this means that H must contain an induced subgraph of the form $P {\circ\!\!\rightarrow} Y {\leftarrow\!\!\circ} Z$ since $Y \notin pa_G(Z)$. At the same time we know that H contains the edge $Y {\rightarrow} Z$ which causes a contradiction and hence $Y \neq W$ must hold. Therefore, $P \notin pa_G(Y)$ holds which generalized means that $pa_G(Y) \subseteq pa_G(Z)$ must hold. For $Z \perp\!\!\!\perp_G P | pa_G(Z)$ to hold in H there must exist an unshielded collider between Z and P over W and hence that the induced subgraph $Z {\circ\!\!\rightarrow} W {\leftarrow\!\!\circ} P$ exists in H. Similarly we have that $Y \perp\!\!\!\perp_G P | pa_G(Y)$ gives that H contains an induced subgraph of the form $Y {\circ\!\!\rightarrow} W {\leftarrow\!\!\circ} P$. Note that $Y \in ad_G(W)$ must hold since condition 2 holds. Moreover for H not to contain a semi-directed cycle over $Y {\rightarrow} Z {\circ\!\!\rightarrow} W {\leftarrow\!\!\circ} Y$ we can see that $Y {\rightarrow} W {\leftarrow\!\!\circ} P$ must exist in H. Finally note that $X \neq W$ must hold since $X \notin ad_G(Z)$ holds.

Now assume $X \in nb_G(W)$. For $X \perp\!\!\!\perp_H Z | nb_G(X) \cup pa_G(X \cup nb_G(X))$ to hold, together with $W \in nb_G(X)$ and $Z {\circ\!\!\rightarrow} W$, it is easy to see that the induced subgraph $Z {\circ\!\!\rightarrow} W {\rightarrow} X$ must be in H. We can now see that $P \in pa_G(X)$ must hold or the induced subgraph $X {\leftarrow} W {\leftarrow\!\!\circ} P$ in H contradicts that $X \perp\!\!\!\perp_G P | pa_G(X)$ holds in H. Moreover, for $X {\circ\!\!-\!\!\circ} Y {\rightarrow} W {\rightarrow} X$ not to form a semi-directed cycle in H the edge between X and Y must be oriented to $X {\leftarrow} Y$. We can therefore restart the proof by replacing X with Z, i.e. with the induced subgraph $X {\leftarrow} Y {\circ\!\!-\!\!\circ} Z$ in H (and $X{-}Y{-}Z$) in G. Since we know that $Z {\circ\!\!\rightarrow} W {\rightarrow} X$ exists in H we know that $ord(co_H(X)) > ord(co_H(Z))$. Hence we cannot get back to this subcase again (or we would have that $ord(co_H(X)) < ord(co_H(Z))$ which is a contradiction). This, together with that Y is kept the same and that $|V|$ is finite gives that the number of restarts is bounded. Hence $X \notin nb_G(W)$ must hold.

Now assume that $pa_G(Z) \subseteq pa_G(W)$. We can now restart the proof with $X \circ\!\!-\!\!\circ Y \to W$. The number of iterations is then bounded since $|V|$ is finite and case 2 cannot occur with Z as W again, or $pa_G(Z) \nsubseteq pa_G(W)$ would have to hold which is a contradiction. Hence $pa_G(Z) \nsubseteq pa_G(W)$ must hold. Let Q be the parent of Z not shared by W. Since $W \perp\!\!\!\perp_G Q | pa_G(W)$ holds, and we know that H contains the induced subgraph $Q \circ\!\!-\!\!\circ Z \circ\!\!\to W \leftarrow\!\!\circ P$, we can draw the conclusion that H must contain the induced subgraph $Q \circ\!\!\to Z \leftrightarrow W \leftarrow\!\!\circ P$ since $Q \notin pa_G(W)$. Note that if there exist two different nodes W_1 and W_2 s.t. both have the properties described for W in case 2 W_1 and W_2 must be adjacent. If this were not the case we would have that both $W_1 \perp\!\!\!\perp_G W_2 | nb_G(W_1) \cup pa_G(W_1 \cup nb_G(W_1))$ and $W_1 \not\perp\!\!\!\perp_H W_2 | nb_G(W_1) \cup pa_G(W_1 \cup nb_G(W_1))$ would hold, since $Z \in nb_G(W_1)$. Also note that since $W_1 \leftrightarrow Z$ and $W_2 \leftrightarrow Z$ exists in H the edge between W_1 and W_2 must be bidirected or H contains a semi-directed cycle. Let D be a set of nodes containing Z as well as all nodes that have the properties described for W. From the description above we can see that D must be complete and that the subgraph induced by D in H must only contain bidirected edges. We will now show that a split must be feasible in G with D as L and $C \setminus D$ as U. For a split not to be feasible one of the constraints in Definition 9 must fail.

Assume condition (1) or (2) fails. Then there exist three nodes $R \in C$, $T \in C$ and $D_j \in D$ s.t. the induced subgraph $T - D_j - R$ exists in G. Since $T \perp\!\!\!\perp_G R | nb_G(R) \cup pa_G(R \cup nb_G(R))$ holds we must, without losing generalization, have that H contains the induced subgraph $T \circ\!\!-\!\!\circ D_j \to R$, since $D_j \in nb_G(R)$. If this is the case we can however restart the proof with this induced subgraph and know that the number of iterations is bounded since $|V|$ is finite and $ord(co_H(D_j)) > ord(co_H(Y))$.

Assume condition (1) and (2) hold but (3) fails. Then there exist two nodes $R \in U$ and $T \notin C$ s.t. the induced subgraph $D_i - R \leftarrow T$ exists in G for some $D_i \in D$. First note that R must be adjacent of all nodes in D or condition 1 would have failed in this split. Secondly note that $R - Y$ must exist in G or condition 2 would fail if we restart the proof with $X \circ\!\!-\!\!\circ Y \to D_i$ and a contradiction follows from there. Thirdly note that $R \notin ad_G(X)$ must hold or the proof could be restarted with $X \circ\!\!-\!\!\circ Y \to D_i$, for which condition 3 would fail with R as W and a contradiction would follow as shown above.

Finally note that $pa_G(R) \subseteq pa_G(Z)$ must hold or R would be in D. This means that $pa_G(W) \notin pa_G(R)$, and hence that $P \notin pa_G(R)$, holds. Moreover we know that the edge $D_i \leftrightarrow W$ exists in G. For $D_i \perp\!\!\!\perp_G T | pa_G(D_i)$ to hold in H it is clear that H must contain the induced subgraph $D_i \circ\!\!\rightarrow R \leftarrow\!\!\circ T$ since $R \notin pa_G(D_i)$. Similarly we have that for $R \perp\!\!\!\perp_H P | pa_G(R)$ to hold H must contain an induced subgraph of the form $R \circ\!\!\rightarrow W \leftarrow\!\!\circ P$ since $W \notin pa_G(R)$. This means that for $R \circ\!\!\rightarrow W \leftrightarrow D_i \circ\!\!\rightarrow R$ not to form a semi-directed cycle in H the edge $R \leftrightarrow D_i$ must exist in H. Moreover, since $\forall D_m \in D \setminus D_i \; R \in ad_G(D_m)$ and $R \leftrightarrow D_i$ and $D_i \leftrightarrow D_m$ hold, $R \leftrightarrow D_m$ must also hold or G contains a semi-directed cycle. Hence the subgraph of H induced by $D \cup R$ is complete and contains only bidirected edges. This in turn means that for $Y \rightarrow D_i \leftrightarrow R \circ\!\!-\!\!\circ Y$ not to form a semi-directed cycle $Y \rightarrow R$ must exist in H. Hence we can move R into D and redo the last split again. The number of restarts are bounded since $|V|$ is finite.

Hence each condition in Definition 9 must hold and we have a contradiction. □

Translation of MVR CGs to LWF CGs

Theorem 21. *Given a MVR CG G, and a maximally oriented MVR CG G' that is in the same Markov equivalence class as G, there exists a LWF CG H s.t. $I_{MVR}(G) = I_{LWF}(H)$ iff G' contains no bidirected edge, i.e. can be represented as a BN.*

Proof. From Lemma 22 it follows that a maximally oriented CG G' in the MVR interpretation with a bidirected edge must have a subgraph of the form shown in Figure 11. If it does not contain any bidirected edge in the maximally oriented model it trivially follows that it is a BN (and hence it can be represented as a CG in the LWF interpretation). From Lemma 23 it then follows that no CG G in the MVR interpretation which contains a subgraph of the form shown in Figure 11 can be represented as a CG in the LWF interpretation. □

Lemma 22. *If a bidirected edge exists in a maximally oriented CG G in the MVR interpretation then G must contain an induced subgraph of the form shown in Figure 11.*

Proof. Assume to the contrary that a CG G in the MVR interpretation exists where (1) no induced subgraph of the form shown in Figure 11 exists, (2)

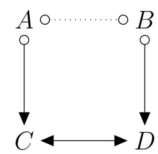


Figure 11: Included subgraph in Lemma 22 and 23.

no split is feasible and (3) at least one bidirected edge exists. From this assumption we can see that there has to exist at least two nodes X and Y s.t. $X \leftrightarrow Y$ exists in G. Let C be the connectivity component to which X and Y belong. Separate the nodes of C into two sets U and L s.t. X and every node connected by a path to X in the subgraph of G induced by $C \setminus Y$ belongs to L and $C \setminus L$ belongs to U. This separation of nodes allows us to know that $sp_G(L)$ only contains Y. For a split not to be feasible at least one condition in Definition 7 has to fail.

Case 1. Assume constraint 1 fails. This means a node $Z \in L$ exists s.t. $Z \leftrightarrow Y \leftrightarrow X$ occurs in G where $Z \notin ad_G(X)$ must hold, or Z would be in U. Now remove Y from U and add it to L as well as all nodes not connected by a path with Z in the subgraph of G induced by $U \setminus Y$ and attempt another split. This separation of nodes allows us, since we previously had $nb_G(L) = Y$ and Y now have changed sets, to say that $sp_G(U) = Y$ must hold and hence that constraint 3 cannot fail. However, if constraint 1 or 2 fails we know there exists a node W s.t. $W \circ\!\!\rightarrow Z \leftrightarrow Y \leftrightarrow X$ is a subgraph of G but where $W \notin ad_G(Y)$, and $Z \notin ad_G(X)$ and $W \notin ad_G(X)$ by definition of the initial split, which implies a contradictory induced subgraph. Hence constraint 1 cannot fail in the initial split.

Case 2. If constraint 2 or 3 fails in the initial split we know there exist two nodes V_1 and P_1 s.t. $P_1 \circ\!\!\rightarrow Y \leftrightarrow V_1$ exists in G but where $V_1 \notin ad_G(P_1)$ (note that V_1 might be X). Now let L consist of every node connected by a path to Y in the subgraph of G induced by $C \setminus V_1$ and the nodes $C \setminus L$ belong to U. This separation of nodes allows us to know that $sp_G(L)$ only contains V_1. If constraint 1 fails when performing a split with these sets it is clear from case 1 that a contradiction occurs. If constraint 2 or 3 fails we know there exist two new nodes V_2 and P_2 s.t. $P_2 \circ\!\!\rightarrow V_1 \leftrightarrow V_2$ exists in G but where $P_2 \notin ad_G(V_2)$. Note that V_2 or P_2 cannot be P_1 since $P_1 \notin ad_G(P_1)$. We now get that V_2 cannot be Y or an induced subgraph like that in Figure 11 occurs. $V_2 \in ad(Y)$ and $P_2 \in ad(Y)$ must also hold or the induced subgraph V_2 (resp.

$P_2) \circ\!\!\rightarrow V_1 \leftrightarrow Y \leftarrow\!\!\circ P_1$ occurs. By replacing V_1 with V_2, setting the proper U and L as described above it we can now repeat this procedure iteratively. Moreover, for every repetition i we must have that V_i and P_i must be adjacent of every $V_j (j < i)$ as well as Y or a contradiction occurs. This means that any nodes V_i and P_i already used in a previous repetition cannot be used in a later one, or both $P_i \in ad(V_i)$ and $P_i \notin ad(V_i)$ would have to hold. This in turn means that the number of repetitions is bounded since $|C|$ is finite and hence we have a contradiction that condition 2 or 3 can fail.

This means that all three conditions in Definition 7 must hold and hence a split must be feasible if the induced subgraph shown in Figure 11 does not occur. □

Lemma 23. *If a CG G in the MVR interpretation contains an induced subgraph of the form shown in Figure 11 then G cannot be translated into a CG H in the LWF interpretation.*

Proof. Assume to the contrary that there exists a CG H, in the LWF interpretation, with the same independence model as G while G contains an induced subgraph of the form shown in Figure 11. Clearly H and G must contain the same nodes and adjacencies or some separations or non-separations must exist in G but not in H.

From Figure 11 we can read that $A \perp\!\!\!\perp_G D | pa_G(D)$ and $C \perp\!\!\!\perp_G B | pa_G(C)$ hold. For $A \perp\!\!\!\perp_G D | pa_G(D)$ to hold in H C must be a collider between A and D and hence H must contain the induced subgraph $A \rightarrow C \leftarrow D$. Similarly $C \perp\!\!\!\perp_G B | pa_G(C)$ gives that H must contain the induced subgraph $C \rightarrow D \leftarrow B$ and hence we have a contradiction. □

Translation of LWF CGs to MVR CGs

Theorem 24. *Given a LWF CG G there exists a CG H s.t. $I_{LWF}(G) = I_{MVR}(H)$ iff $(G_{cl(K)})^m$ is chordal for all $K \in cc(G)$.*

Proof. To prove the "if" part, note that if $(G_{cl(K)})^m$ is chordal for all $K \in cc(G)$, then there is a DAG D s.t. $I_{LWF}(G) = I_{BN}(D)$ [1, Proposition 4.2] and, thus, it suffices to take $H = D$.

To prove the "only if" part, assume to the contrary that $V_1 - \ldots - V_n$ is a chordless undirected cycle in $(G_{cl(K)})^m$ for some $K \in cc(G)$. Note that H has the same adjacencies as G. Therefore, $V_{i-1} \leftarrow V_i$ and/or $V_i \rightarrow V_{i+1}$ must be in H because, otherwise, $V_{i-1} \perp\!\!\!\perp_G V_{i+1} | Z \in I_{LWF}(G)$ for some Z s.t. $V_i \in Z$ whereas

$V_{i-1} \perp\!\!\!\perp_H V_{i+1} | Z \notin I_{MVR}(H)$, which contradicts that $I_{LWF}(G) = I_{MVR}(H)$. Assume without loss of generality that $V_i \rightarrow V_{i+1}$ is in H. Then, $V_{i+1} \rightarrow V_{i+2}$ must be in H too, by an argument similar to the previous one. Repeated application of this reasoning implies that H has a semi-directed cycle, which contradicts the definition of CG. □

Bibliography

[1] S. A. Andersson, D. Madigan, and M. D. Perlman. A Characterization of Markov Equivalence Classes for Acyclic Digraphs. *The Annals of Statistics*, 25:505–541, 1997.

[2] S. A. Andersson, D. Madigan, and M. D. Perlman. An Alternative Markov Property for Chain Graphs. *Scandinavian Journal of Statistics*, 28:33–85, 2001.

[3] S. A. Andersson and M. D. Perlman. Characterizing Markov Equivalence Classes For AMP Chain Graph Models. *The Annals of Statistics*, 34:939–972, 2006.

[4] D. M. Chickering. Optimal Structure Identification with Greedy Search. *Journal of Machine Learning Research*, 3:507–554, 2002.

[5] D. M. Chickering and C. Meek. Finding Optimal Bayesian Networks. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pages 94–102, 2002.

[6] D. R. Cox and N. Wermuth. Linear Dependencies Represented by Chain Graphs. *Statistical Science*, 8:204–218, 1993.

[7] D. R. Cox and N. Wermuth. *Multivariate Dependencies: Models, Analysis and Interpretation.* Chapman and Hall, 1996.

[8] M. Drton. Discrete Chain Graph Models. *Bernoulli*, 15:736–753, 2009.

[9] M. Frydenberg. The Chain Graph Markov Property. *Scandinavian Journal of Statistics*, 17:333–353, 1990.

[10] D. Koller and N. Friedman. *Probabilistic Graphcal Models.* MIT Press, 2009.

[11] S. L. Lauritzen and N. Wermuth. Graphical Models for Association Between Variables, Some of Which are Qualitative and Some Quantitative. *The Annals of Statistics*, 17:31–57, 1989.

[12] B. Liu, Z. Zheng, and H. Zhao. An Efficient Algorithm for Finding the Largest Chain Graph According to a Given Chain Graph. *Science in China Series A: Mathematics*, 48:1517–1530, 2005.

[13] J. D. Nielsen, T. Kočka, and J. M. Peña. On Local Optima in Learning Bayesian Networks. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, pages 435–442, 2003.

[14] J. M. Peña. Learning AMP Chain Graphs under Faithfulness. In *Proceedings of the 6th European Workshop on Probabilistic Graphical Models*, pages 251–258, 2012.

[15] J. M. Peña. Learning AMP Chain Graphs and some Marginal Models Thereof under Faithfulness. *International Journal of Approximate Reasoning*, 55:1011–1022, 2014.

[16] J. M. Peña, D. Sonntag, and J. Nielsen. An Inclusion Optimal Algorithm for Chain Graph Structure Learning. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, pages 778–786, 2014.

[17] A. Roverato and M. Studený. A Graphical Representation of Equivalence Classes of AMP Chain Graphs. *Journal of Machine Learning Research*, 7:1045–1078, 2006.

[18] D. Sonntag and J. M. Peña. Learning Multivariate Regression Chain Graphs under Faithfulness. In *Proceedings of the 6th European Workshop on Probabilistic Graphical Models*, pages 299–306, 2012.

[19] D. Sonntag, J. M. Peña, and Manuel Gómez-Olmedo. Approximate Counting of Graphical Models Via MCMC Revisited. *International Journal of Intelligent Systems*, 2014. Accepted.

[20] M. Studený, A. Roverato, and Š. Štěpánová. Two Operations of Merging and Splitting Components in a Chain Graph. *Kybernetika*, 45:208–248, 2009.

[21] M. Volf and M. Studený. A Graphical Characterization of the Largest Chain Graphs. *International Journal of Approximate Reasoning*, 20:209–236, 1999.

[22] N. Wermuth and K. Sadeghi. Sequences of Regression and Their Independences. arXiv:1103.2523 [stat.ME], 2012.

[23] N. Wermuth, M. Wiedenbeck, and D. R. Cox. Partial Inversion for Linear Systems and Partial Closure of Independence Graphs. *BIT Numerical Mathematics*, 46:883–901, 2006.