EVERY LWF AND AMP CHAIN GRAPH ORIGINATES FROM A SET OF CAUSAL MODELS

JOSE M. PEÑA
ADIT, IDA, LINKÖPING UNIVERSITY, SE-58183 LINKÖPING, SWEDEN
JOSE.M.PENA@LIU.SE

Abstract. This paper aims at justifying LWF and AMP chain graphs by showing that they do not represent arbitrary independence models. Specifically, we show that every chain graph is inclusion optimal wrt the intersection of the independence models represented by a set of directed and acyclic graphs under conditioning. This implies that the independence model represented by the chain graph can be accounted for by a set of causal models that are subject to selection bias. The set of causal models can be accounted for by a system that switches between different regimes or configurations.

1. Introduction

Chain graphs (CGs) are graphs with possibly directed and undirected edges, and no semidirected cycle. They have been extensively studied as a formalism to represent independence models. CGs extend Bayesian networks (BNs), i.e. directed and acyclic graphs (DAGs), and Markov networks, i.e. undirected graphs. Therefore, they can model symmetric and asymmetric relationships between the random variables of interest. This was actually one of the main reasons for developing them. However, unlike Bayesian and Markov networks whose interpretation is unique, there are three main interpretations of CGs as independence models: The Lauritzen-Wermuth-Frydenberg (LWF) interpretation (Frydenberg, 1990; Lauritzen and Wermuth, 1989), the multivariate regression (MVR) interpretation (Cox and Wermuth, 1993, 1996), and the Andersson-Madigan-Perlman (AMP) interpretation (Andersson et al., 2001; Levitz et al., 2001). A fourth interpretation has been proposed by Drton (2009) but it has not been studied sufficiently and, thus, it will not be discussed in this paper. It should be mentioned that any of the three main interpretations can represent independence models that cannot be represented by the other two interpretations (Sonntag and Peña, 2013).

Along with other reasons, DAGs can convincingly be justified by the fact that each of them represents a causal model. Whether this is an ontological model is still debated. However, it is widely accepted that the causal model is at least epistemological and thus worth studying (Pearl, 2000). Of the three main interpretations of CGs, however, only MVR CGs have a convincing justification: Since MVR CGs are maximal ancestral graphs without undirected edges, every MVR CG represents the independence model represented by a DAG under marginalization (Richardson and Spirtes, 2002, Theorem 6.4). That is, every MVR CG can be accounted for by a causal model that is partially observed. Unfortunately, LWF and AMP CGs cannot be justified in the same manner because (i) LWF and AMP CGs can represent independence models that cannot be represented by maximal ancestral graphs (Richardson and Spirtes, 2002, Section 9.4), and (ii) maximal ancestral graphs can represent all the independence models represented by DAGs under marginalization and conditioning (Richardson and Spirtes, 2002, Theorem 4.18). In other words, LWF and AMP CGs can represent independence models that cannot be represented by any DAG under marginalization and conditioning. Of course, LWF and AMP CGs can be justified by the fact that they improve the expressivity of DAGs, i.e. they can represent more independence models...
than DAGs (Peña, 2007). However, this is a weak justification unless those independence models are not arbitrary but induced by some class of knowledge representatives within some uncertainty calculus of artificial intelligence, e.g., the class of probability distributions (Studený, 2005, Section 1.1). This is exactly what Levitz et al. (2001), Peña (2009, 2011a), and Studený and Bouckaert (1998) do by showing that every LWF and AMP CG is faithful to some probability distribution. However, this does not strengthen much the justification unless these probability distributions are not arbitrary but they represent meaningful systems or phenomena. This is exactly what Lauritzen and Richardson (2002) do. In particular, the authors show that every LWF CG includes the independence model induced by the equilibrium probability distribution of a dynamic model with feed-back. The downside of this justification is that the equilibrium distribution may not be reached in finite time and, thus, it may not coincide with the distribution that represents the behaviour of the dynamic model at any finite time point. Therefore, there is no guarantee that the CG includes the independence model induced by the latter, which is the goal. The authors are aware of this and state that their justification should better be understood as an approximated one. Another work in the same vein is that of Ferrandiz et al. (2005), who show that some LWF CGs are inclusion minimal wrt the result of temporal aggregation in a DAG representing a spatio-temporal process. Unfortunately, the authors do not show whether their result holds for every LWF CG. Yet another work along the same lines is that of Peña (2014), who shows that every AMP CG is faithful to the independence model represented by a DAG under marginalization and conditioning. It is worth noting that the DAG contains deterministic nodes, because the result does not hold otherwise (Richardson, 1998). Finally, Studený (1998) presents a justification of a LWF CG as representing a sequence of areas of expertise, one for each connectivity component of the CG, where each area is described by an expert in terms of independencies holding in it. The directed edges in the CG indicate the sequence of the areas and the undirected edges indicate lack of independencies in the corresponding area. Note that whereas this justification may offer a causal explanation for the directed edges in the CG, it does not do it for the undirected edges, which are simply explained as lack of independencies. We would like to give a causal explanation even to the undirected edges in the CG. In other words, we would like to have a justification of LWF CGs that builds solely on causal models.

In this work, we show that every LWF and AMP CG $G$ is inclusion optimal wrt the intersection of the independence models represented by a set of DAGs under conditioning. In other words, we show that (i) the independencies represented by $G$ are a subset of the intersection, and (ii) the property (i) is not satisfied by any CG that represents a proper superset of the independencies represented by $G$. Note that if there exists a CG that is faithful to the intersection, then that CG is inclusion optimal. In general, several inclusion optimal CGs exist and they do not necessarily represent the same independence model. Therefore, in principle, one prefers the inclusion optimal CGs that represent the largest number of independencies. However, finding any such CG seems extremely difficult, probably NP-complete in the light of the work by Peña (2011b). Thus, one is typically content with finding any inclusion optimal CG. An example of this are the algorithms for learning inclusion optimal BNs (Chickering and Meek, 2002; Nielsen et al., 2003) and LWF CGs (Peña et al., 2014). This is also why we are content with showing in this paper that every LWF and AMP CG is inclusion optimal wrt the intersection of the independence models represented by a set of DAGs under conditioning. The intersection can be thought of as a consensus independence model, in the sense that it contains all and only the independencies upon which all the DAGs under conditioning agree. We elaborate further on the term consensus in the paragraph below. The fact that every LWF and AMP CG originates from a set of DAGs under conditioning implies that the independence model represented by the former can be accounted for by a set of causal models that are subject to selection bias. The set of causal models can be accounted for by a system that switches between different regimes or configurations. Two
examples of such a system are the progression of a disease through different stages, and the behaviour of a broker alternating between looking for buying and selling opportunities. We have recently introduced a new family of graphical models aiming at modeling such systems (Bendtsen and Peña, 2013, 2014). In summary, we provide an alternative justification of LWF and AMP CGs that builds solely on causal models and does not involve equilibrium distributions or deterministic nodes, which may seem odd to some readers. Our hope is that this strengthens the case of LWF and AMP CGs as a useful representation of the independence models entailed by causal models.

Before we proceed further, it is worth discussing the relationship between our justification of LWF and AMP CGs and belief aggregation. First, recall that a BN is an efficient representation of a probability distribution. Specifically, a BN consists of structure and parameter values. The structure is a DAG representing an independence model. The parameter values specify the conditional probability distribution of each node given its parents in the BN structure. The BN represents the probability distribution that results from the product of these conditional probability distributions. Moreover, the probability distribution satisfies the independence model represented by the BN structure. Belief aggregation consists in obtaining a group consensus probability distribution from the probability distributions specified by the individual members of the group. Probably, the two most commonly used consensus functions are the weighted arithmetic and geometric averages. Pennock and Wellman (2005) show that belief aggregation is problematic when the consensus and the individual probability distributions are represented as BNs. Specifically, they show that even if the group members agree on the BN structure, there is no sensible consensus function that always returns a probability distribution that can be represented as a BN whose structure represents the same independence model as the agreed BN structure (Pennock and Wellman, 2005, Proposition 2). The only exception to this negative result is when the individual BN structures are decomposable and the consensus function is the weighted geometric average (Pennock and Wellman, 2005, Sections 3.3-3.4). However, the authors also point out that this negative result does not invalidate the arguments of those who advocate preserving the agreed independencies, e.g. Laddaga (1977) and Raiffa (1968, Section 8.12). It simply indicates that a different approach to belief aggregation is needed in this case. They actually mention one such approach that consists in performing the aggregation in two steps: First, find a consensus BN structure that preserves as many of the agreed independencies as possible and, second, find consensus parameter values for the consensus BN structure. The first step has received significant attention in the literature (del Sagrado and Moral, 2003; Matzkevich and Abramson, 1992a,b; Nielsen and Parsons, 2007). A work that studies both steps is that of Bonduelle (1987). We have also studied both steps (Etminani et al., 2013; Peña, 2011b). The two step approach described above is also suitable when some of the group members are able to contribute with a BN structure but not with parameter values. This scenario is not unlikely given that people typically find easier to gather qualitative than quantitative knowledge.

Our justification of LWF and AMP CGs implicitly advocates preserving the agreed independencies, because the DAGs in the justification are combined through the intersection of the independence models that they represent and, thus, the agreed independencies are kept. As shown above, this is a sensible advocation. Therefore, in this paper we make use of it to propose a sensible justification of LWF and AMP CGs. The DAGs in our justification are hand-picked to ensure that the combination thereof produces the desired result. This raises the question of how to combine a set of arbitrary DAGs under marginalization and conditioning into a LWF or AMP CG. In this paper, we also investigate this question. Ideally, we would like to find a LWF or AMP CG that is inclusion optimal wrt the intersection of the independence models represented by the DAGs under marginalization and conditioning.

\[1\] Unfortunately, we could not get access to this work. So, we trust the description of it made by Pennock and Wellman (2005, Section 3.5).
Unfortunately, this problem seems extremely hard. So, we actually study a simpler version of it. Note that this problem corresponds to the first step of the approach to belief aggregation described above. The second step, i.e. combining the parameter values associated to the DAGs, is beyond the scope of this paper.

The rest of the paper is organized as follows. In Section 2 we introduce some preliminaries and notation. In Section 3 we present our justification of LWF and AMP CGs. In Section 4 we discuss how to combine arbitrary DAGs into a LWF or AMP CG. We close with some discussion in Section 5.

2. Preliminaries

In this section, we review some concepts from graphical models that are used later in this paper. Unless otherwise stated, all the graphs in this paper are defined over a finite set $V$. Moreover, they are all simple, i.e. they contain at most one edge between any pair of nodes. The elements of $V$ are not distinguished from singletons. The set operators union, intersection and difference are given equal precedence in the expressions. The term maximal is always wrt set inclusion.

If a graph $G$ contains an undirected or directed edge between two nodes $V_i$ and $V_2$, then we write that $V_i - V_2$ or $V_i \rightarrow V_2$ is in $G$. The parents of a set of nodes $X$ of $G$ is the set $pa_G(X) = \{V_i|V_i \rightarrow V_2 \text{ is in } G, V_i \notin X \text{ and } V_2 \in X\}$. The children of $X$ is the set $ch_G(X) = \{V_i|V_i \leftarrow V_2 \text{ is in } G, V_i \notin X \text{ and } V_2 \in X\}$. The neighbors of $X$ is the set $ne_G(X) = \{V_i|V_1 \rightarrow V_2 \text{ is in } G, V_i \notin X \text{ and } V_2 \notin X\}$. The boundary of $X$ is the set $bd_G(X) = ne_G(X) \cup pa_G(X)$. The adjacents of $X$ is the set $ad_G(X) = ne_G(X) \cup pa_G(X) \cup ch_G(X)$. A route between a node $V_i$ and a node $V_n$ in $G$ is a sequence of (not necessarily distinct) nodes $V_i, \ldots, V_n$ st $V_i \in ad_G(V_{i+1})$ for all $1 \leq i < n$. If the nodes in the route are all distinct, then the route is called a path. A route is called undirected if $V_i - V_{i+1}$ is in $G$ for all $1 \leq i < n$. A route is called descending if $V_i \rightarrow V_{i+1}$ or $V_i - V_{i+1}$ is in $G$ for all $1 \leq i < n$. A route is called strictly descending if $V_i \rightarrow V_{i+1}$ is in $G$ for all $1 \leq i < n$. A chain graph (CG) is a graph whose every edge is directed or undirected st it has no semidirected cycles. Note that a CG with only directed edges is a directed and acyclic graph (DAG), and a CG with only undirected edges is an undirected graph (UG). A set of nodes of a CG is connected if there exists an undirected path in the CG between every pair of nodes in the set. A connectivity component of a CG is a maximal connected set. We denote by $co_G(X)$ the connectivity component of the CG $G$ to which a node $X$ belongs. A chain $\alpha$ is a partition of $V$ into ordered subsets, which we call blocks. We say that a CG $G$ and a chain $\alpha$ are consistent when (i) for every edge $X \rightarrow Y$ in $G$, the block containing $X$ precedes the block containing $Y$ in $\alpha$, and (ii) for every edge $X - Y$ in $G$, $X$ and $Y$ are in the same block of $\alpha$. Note that the blocks of $\alpha$ and the connectivity components of $G$ may not coincide, but each of the latter must be included in one of the former.

Let $X, Y, Z$ and $W$ denote four disjoint subsets of $V$. An independence model $M$ is a set of statements of the form $X \perp_M Y|Z$, meaning that $X$ is independent of $Y$ given $Z$. Moreover, $M$ is called graphoid if it satisfies the following properties: Symmetry $X \perp_M Y|Z \Rightarrow Y \perp_M X|Z$, decomposition $X \perp_M Y \cup W|Z \Rightarrow X \perp_M Y|Z$, weak union $X \perp_M Y \cup W|Z \Rightarrow X \perp_M Y|Z \cup W$, contraction $X \perp_M Y|Z \cup W \wedge X \perp_M W|Z \Rightarrow X \perp_M Y|Z \cup W$, and intersection $X \perp_M Y|Z \cup W \wedge X \perp_M W|Z \cup Y \Rightarrow X \perp_M Y|Z \cup W$. Moreover, $M$ is called compositional graphoid if it is a graphoid that also satisfies the composition property $X \perp_M Y|Z \wedge X \perp_M W|Z \Rightarrow X \perp_M Y \cup W|Z$. By convention, $X \perp_M \emptyset|Z$ and $\emptyset \perp_M Y|Z$. 


We now recall the semantics of LWF and AMP CGs. A section of a route $\rho$ in a LWF CG is a maximal undirected subroute of $\rho$. A section $V_2-\ldots-V_{n-1}$ of $\rho$ is a collider section of $\rho$ if $V_1 \rightarrow V_2-\ldots-V_{n-1} \leftarrow V_n$ is a subroute of $\rho$. Moreover, $\rho$ is said to be Z-open with $Z \subseteq V$ when (i) every collider section of $\rho$ has a node in $Z$, and (ii) no non-collider section of $\rho$ has a node in $Z$.

A node $B$ in a route $\rho$ in an AMP CG $G$ is called a triplex node in $\rho$ if $A \rightarrow B \leftarrow C$, $A \rightarrow B \leftarrow C$, or $A \rightarrow B \leftarrow C$ is a subroute of $\rho$. Note that maybe $A = C$ in the first case. Note also that $B$ may be both a triplex and a non-triplex node in $\rho$. Moreover, $\rho$ is said to be Z-open with $Z \subseteq V$ when (i) every triplex node in $\rho$ is in $Z$, and (ii) every non-triplex node in $\rho$ is outside $Z$.

Let $X$, $Y$, and $Z$ denote three disjoint subsets of $V$. When there is no Z-open route in a LWF or AMP CG $G$ between a node in $X$ and a node in $Y$, we say that $X$ is separated from $Y$ given $Z$ in $G$ and denote it as $X \perp_G Y|Z$. The independence model represented by $G$, denoted as $I(G)$, is the set of separations $X \perp_G Y|Z$. In general, $I(G)$ is different depending on whether $G$ is interpreted as a LWF or AMP CG. However, if $G$ is a DAG or UG, then $I(G)$ is the same under the two interpretations. Given a CG $G$ and two disjoint subsets $L$ and $S$ of $V$, we denote by $[I(G)]_L^S$ the independence model represented by $G$ under marginalization of the nodes in $L$ and conditioning on the nodes in $S$. Specifically, $X \perp_G Y|Z$ is in $[I(G)]_L^S$ iff $X \perp_{G \setminus L} Y|Z \cup S$ is in $I(G)$ and $X,Y,Z \subseteq V \setminus L \setminus S$.

We say that a CG $G$ includes an independence model $M$ if $I(G) \subseteq M$. Moreover, we say that $G$ is inclusion minimal wrt $M$ if removing any edge from $G$ makes it cease to include $M$. We say that a CG $G_{\alpha}$ is inclusion minimal wrt an independence model $M$ and a chain $\alpha$ if $G_{\alpha}$ is inclusion minimal wrt $M$ and $G_{\alpha}$ is consistent with $\alpha$. We also say that a CG $G$ is inclusion optimal wrt an independence model $M$ if $I(G) \subseteq M$ and there exists no other CG $H$ st $I(G) \subseteq I(H) \subseteq M$.

Finally, a subgraph of a CG $G$ is a CG whose nodes and edges are all in $G$. The subgraph of a CG $G$ induced by a set of its nodes $X$ is the CG over $X$ that has all and only the edges in $G$ whose both ends are in $X$. A complex in a LWF CG is an induced subgraph of it of the form $V_1 \rightarrow V_2-\ldots-V_{n-1} \leftarrow V_n$. A triplex in an AMP CG is an induced subgraph of it of the form $A \rightarrow B \leftarrow C$, $A \rightarrow B \leftarrow C$, or $A \rightarrow B \leftarrow C$.

### 3. Justification of LWF and AMP CGs

The theorem below shows that every LWF or AMP CG $G$ is inclusion optimal wrt the intersection of the independence models represented by some DAGs under conditioning. The DAGs are obtained as follows. First, we decompose $G$ into a DAG $G_D$ and an UG $G_U$, i.e. $G_D$ contains all and only the directed edges in $G$, and $G_U$ contains all and only the undirected edges in $G$. Then, we construct a DAG $G_S$ from $G_U$ by replacing every edge $X \rightarrow Y$ in $G_U$ with $X \rightarrow S_{XY} \leftarrow Y$. The nodes $S_{XY}$ are called selection nodes. Let $S$ denote all the selection nodes in $G_S$. Note that $G_D$ and $G_U$ are defined over the nodes $V$, but $G_S$ is defined over the nodes $V \cup S$.

**Theorem 1.** The LWF or AMP CG $G$ is inclusion optimal wrt $I(G_D) \cap [I(G_S)]_G$.  

**Proof.** First, assume that $G$ is a LWF CG. Assume to the contrary that there exists a LWF CG $H \subseteq I(G) \cap I(H) \subseteq I(G_D) \cap [I(G_S)]_G$. Note that $G$ and $H$ must have the same adjacencies because, otherwise, there are two nodes $X,Y \in V$ that are adjacent in $H$ but not in $G$, or vice versa. The first case implies that $X \perp_G Y|Z$ holds but $X \perp_{\Gamma} Y|Z$ does not hold for some $Z \subseteq V \setminus X \setminus Y$, which contradicts that $I(G) \subseteq I(H)$. The second case implies that $X \perp_{G_D} Y|Z$ or $X \perp_{G_S} Y|Z \cup S$ does not hold for any $Z \subseteq V \setminus X \setminus Y$. Then, $X \perp Y$ is in $I(H)$ but not in $I(G_D) \cap [I(G_S)]_G$, which contradicts that $I(H) \subseteq I(G_D) \cap [I(G_S)]_G$. Moreover, if $G$...
and $H$ have the same adjacencies, then they must also have the same complexes because, otherwise, there are two nodes $X, Y \in V$ st $X \perp_G Y | Z$ holds but $X \perp_H Y | Z$ does not hold for some $Z \subseteq V \setminus X \setminus Y$, which contradicts that $I(G) \subset I(H)$. However, that $G$ and $H$ have the same adjacencies and complexes contradicts that $I(G) \subset I(H)$ (Frydenberg, 1990, Theorem 5.6).

Now, assume that $G$ is an AMP CG. Assume to the contrary that there exists an AMP CG $H$ st $I(G) \subset I(H) \subseteq I(G_D) \cap [I(G_S)]^S_{G}$. Note that $G$ and $H$ must have the same adjacencies, by a reasoning similar to the one used above for LWF CGs. Then, they must also have the same triplexes because, otherwise, there are two nodes $X, Y \in V$ st $X \perp_G Y | Z$ holds but $X \perp_H Y | Z$ does not hold for some $Z \subseteq V \setminus X \setminus Y$, which contradicts that $I(G) \subset I(H)$. However, that $G$ and $H$ have the same adjacencies and triplexes contradicts that $I(G) \subset I(H)$ as shown by Andersson et al. (2001, Theorem 5) and Levitz et al. (2001, Theorem 6.1). \hfill \square

Unfortunately, the LWF or AMP CG $G$ may not be faithful to $I(G_D) \cap [I(G_S)]^S_{G}$. To see it, let $G$ be $A \rightarrow B \rightarrow C \leftarrow D$. Then, $A \perp D | B \cup C$ is in $I(G_D) \cap [I(G_S)]^S_{G}$ but not in $I(G)$. We doubt that one can prove (and so strengthen our justification) that every LWF or AMP CG is faithful to the intersection of the independence models represented by some DAGs under conditioning. However, it is true that the decomposition of $G$ into $G_D$ and $G_U$ is not the only one that allows us to prove that $G$ is inclusion optimal wrt to the intersection of the independence models represented by some DAGs under conditioning. For instance, we can also prove this result if $G$ is decomposed into a set of DAGs and UGs st none of them has more than one edge, or if $G$ is decomposed into a set of CGs st none of them has a subgraph of the form $A \rightarrow B \rightarrow C$. We omit the proofs. In any case, this does not change the main message of this work, namely that LWF and AMP CGs can be justified on the sole basis of causal models. Having said this, we prefer the original decomposition because it is not completely arbitrary: $G_D$ represents the relationships in $G$ that are causal, and $G_U$ those that are non-causal and need to be explained through conditioning.

Finally, note that the LWF or AMP CG $G$ may not be the only inclusion optimal CG wrt $I(G_D) \cap [I(G_S)]^S_{G}$. To see it, let $G$ be $A \rightarrow B \rightarrow C \leftarrow D$. Then, any LWF or AMP CG that has the same adjacencies as $G$ is inclusion optimal wrt $I(G_D) \cap [I(G_S)]^S_{G}$. Some of these other inclusion optimal CGs may even be preferred instead of $G$ according to some criteria (e.g. number of independencies represented, or number of directed and/or undirected edges). However, $G$ is preferred according to an important criterion: It is the only one that has all and only the strictly ascendant relationships (i.e. direct and indirect causal relationships) between two nodes in $V$ that exist in $G_D$ and $G_S$.

4. Combination of arbitrary DAGs into a LWF or AMP CG

Let $G_1, \ldots, G_r$ denote $r$ arbitrary DAGs, where any $G_i$ is defined over the nodes $V \cup L_i \cup S_i$ and it is subject to marginalization of the nodes in $L_i$ and conditioning on the nodes in $S_i$. We would like to find a LWF or AMP CG that is inclusion optimal wrt $\bigcap_{i=1}^r [I(G_i)]^S_{L_i}$. However, this seems to be an extremely hard problem. So, we study a simpler version of it in which we are only interested in those CGs that are consistent with a chain $\alpha$. Then, our goal becomes to find an inclusion minimal LWF or AMP CG wrt $\bigcap_{i=1}^r [I(G_i)]^S_{L_i}$ and $\alpha$. The prior knowledge of $\alpha$ represents our a priori knowledge on which nodes may be causally related and which nodes may be non-causally related. The latter determine the blocks of $\alpha$, and the former the ordering of the blocks in $\alpha$. The theorems below solve our problem. Specifically, they give a constructive characterization of the unique LWF (respectively AMP) CG that is inclusion minimal wrt a graph (respectively compositional graph) and a chain. Note that any $I(G_\alpha)$ is a compositional graphoid (Sadeghi and Lauritzen, 2012, Theorem 1). Moreover, it is easy to verify that any $[I(G_i)]^S_{L_i}$ is also a compositional graphoid and, thus, $\bigcap_{i=1}^r [I(G_i)]^S_{L_i}$ is also a compositional graphoid. Thus, the theorems below apply to our problem.
Theorem 2. Let $M$ denote an independence model, and $\alpha$ a chain with blocks $b_1, \ldots, b_n$. If $M$ is a graphoid, then there exists a unique LWF CG $G_\alpha$ that is inclusion minimal wrt $M$ and $\alpha$. Specifically, for each node $X$ of each block $b_i$ of $\alpha$, $bd_{G_\alpha}(X)$ is the smallest subset of $\bigcup_{i=1}^n b_j \setminus X$ st $X \perp b_1 \setminus X \setminus ne_{G_\alpha}(X) \cup b_{j=1}^n b_j \cup ne_{G_\alpha}(X)$.

Proof. The theorem has been proven by Peña et al. (2014, Lemma 1).

Theorem 3. Let $M$ denote an independence model, and $\alpha$ a chain with blocks $b_1, \ldots, b_n$. If $M$ is a compositional graphoid, then there exists a unique AMP CG $G_\alpha$ that is inclusion minimal wrt $M$ and $\alpha$. Specifically, consider the blocks in $\alpha$ in reverse order and perform the following two steps for each of them. First, for each node $X$ of the block $b_i$, $ne_{G_\alpha}(X)$ is the smallest subset of $b_i \setminus X \setminus \bigcup_{j=1}^n b_j \cup ne_{G_\alpha}(X)$ such that $X \perp b_1 \setminus X \setminus ne_{G_\alpha}(X) \cup b_{j=1}^n b_j \cup ne_{G_\alpha}(X)$. Second, for each node $X$ of the block $b_i$, $pa_{G_\alpha}(X)$ is the smallest subset of $\bigcup_{j=1}^n b_j \setminus X \perp b_1 \setminus X \setminus de_{G_\alpha}(X) \cup b_{j=1}^n b_j \cup pa_{G_\alpha}(X) | pa_{G_\alpha}(X)$.

Proof. Consider any $X \in V$. Assume that $X \in b_i$. By construction, we have that $Y \perp Y \setminus de_{G_\alpha}(Y) | pa_{G_\alpha}(Y) | pa_{G_\alpha}(Y)$ for any $Y \in X \cup ne_{G_\alpha}(X)$ Then, $Y \perp Y \setminus de_{G_\alpha}(Y) | pa_{G_\alpha}(X \cup ne_{G_\alpha}(X)) | pa_{G_\alpha}(X \cup ne_{G_\alpha}(X))$ for any $Y \in X \cup ne_{G_\alpha}(X)$ by weak union. Then, $X \perp X \setminus ne_{G_\alpha}(X) | de_{G_\alpha}(X) \setminus pa_{G_\alpha}(X \cup ne_{G_\alpha}(X)) | pa_{G_\alpha}(X \cup ne_{G_\alpha}(X))$ by repeated application of symmetry and composition. Then, $X \perp X \setminus de_{G_\alpha}(X) \setminus pa_{G_\alpha}(X \cup ne_{G_\alpha}(X)) | pa_{G_\alpha}(X \cup ne_{G_\alpha}(X)) \cup ne_{G_\alpha}(X)$ by symmetry and weak union. Then, $X \perp X \setminus \bigcup_{j=1}^n b_j \setminus X \setminus ne_{G_\alpha}(X) \cup b_{j=1}^n b_j \cup ne_{G_\alpha}(X)$ by decomposition. This together with $X \perp X \setminus ne_{G_\alpha}(X) | de_{G_\alpha}(X) \perp X \setminus ne_{G_\alpha}(X) \cup b_{j=1}^n b_j \cup ne_{G_\alpha}(X)$ which follows by construction, imply that $X \perp X \setminus ne_{G_\alpha}(X) \cup b_{j=1}^n b_j \cup ne_{G_\alpha}(X)$ by contraction and decomposition. This together with $X \perp X \setminus de_{G_\alpha}(X) \setminus pa_{G_\alpha}(X | pa_{G_\alpha}(X)$ which follows by construction, imply by decomposition that $X \perp X \setminus ne_{G_\alpha}(X) \cup b_{j=1}^n b_j \cup ne_{G_\alpha}(X)$ for any $Y \in X \setminus ne_{G_\alpha}(X)$, and $X \perp X \setminus pa_{G_\alpha}(X)$ for any $Y \in X \setminus de_{G_\alpha}(X) \setminus pa_{G_\alpha}(X)$. This implies that $G_\alpha$ includes $M$. In fact, $G_\alpha$ is inclusion minimal wrt $M$ and $\alpha$ by construction of $ne_{G_\alpha}(X)$ and $pa_{G_\alpha}(X)$.

Assume to the contrary that there exists another AMP CG $H_\alpha$ that is inclusion minimal wrt $M$ and $\alpha$. Let $X \in V$ denote any node st $ne_{G_\alpha}(X) \neq ne_{H_\alpha}(X)$. Assume that $X \in b_i$. Then, $X \perp b_1 \setminus X \setminus ne_{G_\alpha}(X) | \bigcup_{j=1}^n b_j \cup ne_{G_\alpha}(X)$ and $X \perp b_1 \setminus X \setminus ne_{H_\alpha}(X) | \bigcup_{j=1}^n b_j \cup ne_{H_\alpha}(X)$ because $G_\alpha$ and $H_\alpha$ include $M$. Then, $X \perp b_1 \setminus X \setminus ne_{G_\alpha}(X) \cap ne_{H_\alpha}(X) | \bigcup_{j=1}^n b_j \cup [ne_{G_\alpha}(X) \cap ne_{H_\alpha}(X)]$ and $X \perp b_1 \setminus X \setminus [ne_{G_\alpha}(X) \cap ne_{H_\alpha}(X)] | \bigcup_{j=1}^n b_j \cup [ne_{G_\alpha}(X) \cap ne_{H_\alpha}(X)]$

Note that $de_{G_\alpha}(X)$ for any $X \in b_i$ is known when the second step for $b_i$ starts, because $ne_{G_\alpha}(X)$ for any $X \in \bigcup_{j=1}^n b_j$ and $pa_{G_\alpha}(X)$ for any $X \in \bigcup_{j=1}^n b_j$ have already been identified.

Theorems 5 and 6 in the work of Peña (2014) are stated for so-called marginal AMP CGs. However, they also apply to AMP CGs because these are marginal AMP CGs without bidirected edges.

This result may also be derived by adapting to general independence models the results reported by Andersson et al. (2001, Section 4) for probability distributions.
by intersection. However, this contradicts the definition of \( ne_{G_\alpha}(X) \), because \( ne_{G_\alpha}(X) \cap ne_{H_\alpha}(X) \) is smaller than \( ne_{G_\alpha}(X) \). Consequently, \( ne_{G_\alpha}(X) = ne_{H_\alpha}(X) \) for any \( X \in V \).

Let \( i \) denote the largest block index st there is some \( X \in b_i \) st \( pa_{G_\alpha}(X) \neq pa_{H_\alpha}(X) \). Note that \( de_{G_\alpha}(X) = de_{H_\alpha}(X) \), because \( pa_{G_\alpha}(Y) = pa_{H_\alpha}(Y) \) for any \( Y \in \bigcup_{j=1}^n b_j \) and, as proven above, \( ne_{G_\alpha}(Y) = ne_{H_\alpha}(Y) \) for any \( Y \in V \). Then, \( X \perp M V \setminus X \setminus de_{G_\alpha}(X) \setminus pa_{G_\alpha}(X) | pa_{G_\alpha}(X) \)

and \( X \perp M V \setminus X \setminus de_{H_\alpha}(X) \setminus pa_{H_\alpha}(X) | pa_{H_\alpha}(X) \)

because \( G_\alpha \) and \( H_\alpha \) include \( M \). Then, \( X \perp M V \setminus X \setminus de_{G_\alpha}(X) \setminus [pa_{G_\alpha}(X) \cap pa_{H_\alpha}(X)] | [pa_{G_\alpha}(X) \cap pa_{H_\alpha}(X)] \)

by intersection. However, this contradicts the definition of \( pa_{G_\alpha}(X) \), because \( pa_{G_\alpha}(X) \cap pa_{H_\alpha}(X) \) is smaller than \( pa_{G_\alpha}(X) \). Consequently, \( pa_{G_\alpha}(X) = pa_{H_\alpha}(X) \) for any \( X \in V \). Therefore, \( G_\alpha \) and \( H_\alpha \) have the same edges, which is a contradiction.

\[ \square \]

5. Discussion

The purpose of this paper has been to justify LWF and AMP CGs by showing that they do not represent arbitrary independence models. Unlike previous justifications, ours builds solely on causal models and does not involve equilibrium distributions or deterministic nodes, which may seem odd to some readers. Specifically, for any given LWF or AMP CG, we have imagined a system that switches between different regimes or configurations, and we have shown that the given CG represents the different regimes jointly. To do so, we have assumed that each of the regimes can be represented by a causal model. We have also assumed that the causal models may be subject to selection bias. In other words, we have assumed that each of the regimes can be represented by a DAG under conditioning.

In this paper, we have also studied how to combine a set of arbitrary DAGs under marginalization and conditioning into a consensus LWF or AMP CG. We have shown how to do it optimally when the consensus CG must be consistent with a given chain. The chain may represent our prior knowledge about the causal and non-causal relationships in the domain at hand. In the future, we would like to drop this requirement. We would also like to find parameter values for the consensus CG by combining the parameter values associated to the given DAGs.

Acknowledgments

We thank the Reviewers for their thorough comments, which helped us to improve the manuscript significantly. We specially thank Reviewer 2 for pointing out a flaw in our original manuscript. This work is funded by the Center for Industrial Information Technology (CENIIT) and a so-called career contract at Linköping University, and by the Swedish Research Council (ref. 2010-4808).

References

Andersson, S. A., Madigan, D. and Perlman, M. D. Alternative Markov Properties for Chain Graphs. Scandinavian Journal of Statistics, 28:33-85, 2001.

Bendtsen, M. and Peña, J. M. Gated Bayesian Networks. In Proceedings of the 12th Scandinavian Conference on Artificial Intelligence, 35-44, 2013.

Bendtsen, M. and Peña, J. M. Learning Gated Bayesian Networks for Algorithmic Trading. In Proceedings of the 7th European Workshop on Probabilistic Graphical Models, accepted, 2014.

Bonduelle, Y. Aggregating Expert Opinions by Resolving Sources of Disagreement. PhD Thesis, Stanford University, 1987.
Chickering, D. M. and Meek, C. Finding Optimal Bayesian Networks. In Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence, 94-102, 2002.

Cox, D. R. and Wermuth, N. Linear Dependencies Represented by Chain Graphs. Statistical Science, 8:204-218, 1993.

Cox, D. R. and Wermuth, N. Multivariate Dependencies - Models, Analysis and Interpretation. Chapman & Hall, 1996.

del Sagrado, J. and Moral, S. Qualitative Combination of Bayesian Networks. International Journal of Intelligent Systems, 18:237-249, 2003.

Drton, M. Discrete Chain Graph Models. Bernoulli, 15:736-753, 2009.

Etminani, K., Naghibzadeh, M. and Peña, J. M. DemocraticOP: A Democratic Way of Aggregating Bayesian Network Parameters. International Journal of Approximate Reasoning, 54:602-614, 2013.

Ferrandiz, J., Castillo, E. and Sanmartín, P. Temporal Aggregation in Chain Graph Models. Journal of Statistical Planning and Inference, 133:69-93, 2005.

Frydenberg, M. The Chain Graph Markov Property. Scandinavian Journal of Statistics, 17:333-353 1990.

Laddaga, R. Lehrer and the Consensus Proposal. Synthese, 36:473-477, 1977.

Lauritzen, S. L. Graphical Models. Oxford University Press, 1996.

Lauritzen, S. L. and Richardson, T. S. Chain Graph Models and Their Causal Interpretations. Journal of the Royal Statistical Society B, 64:321-361, 2002.

Lauritzen, S. L. and Wermuth, N. Graphical Models for Associations between Variables, some of which are Qualitative and some Quantitative. Annual of Statistics, 17:31-57, 1989.

Levitz, M., Perlman M. D. and Madigan, D. Separation and Completeness Properties for AMP Chain Graph Markov Models. The Annals of Statistics, 29:1751-1784, 2001.

Matzkevich, I. and Abramson, B. The Topological Fusion of Bayes Nets. In Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence, 191-198, 1992.

Matzkevich, I. and Abramson, B. Some Complexity Considerations in the Combination of Believe Networks. In Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence, 152-158, 1993a.

Matzkevich, I. and Abramson, B. Deriving a Minimal I-Map of a Belief Network Relative to a Target Ordering of its Nodes. In Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence, 159-165, 1993b.

Maynard-Reid II, P. and Chajewska, U. Agregating Learned Probabilistic Beliefs. In Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, 354-361, 2001.

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

Nielsen, S. H. and Parsons, S. An Application of Formal Argumentation: Fusing Bayesian Networks in Multi-Agent Systems. Artificial Intelligence 171:754-775, 2007.

Pearl, J. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2000.

Pennock, D. M. and Wellman, M. P. Graphical Models for Groups: Belief Aggregation and Risk Sharing. Decision Analysis, 2:148-164, 2005.

Peña, J. M. Approximate Counting of Graphical Models Via MCMC. In Proceedings of the 11th International Conference on Artificial Intelligence and Statistics, 352-359, 2007.

Peña, J. M. Faithfulness in Chain Graphs: The Discrete Case. International Journal of Approximate Reasoning, 50:1306-1313, 2009.

Peña, J. M. Faithfulness in Chain Graphs: The Gaussian Case. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, 588-599, 2011a.

Peña, J. M. Finding Consensus Bayesian Network Structures. Journal of Artificial Intelligence Research, 42:661-687, 2011b.
Peña, J. M. Learning AMP Chain Graphs under Faithfulness. In *Proceedings of the 6th European Workshop on Probabilistic Graphical Models*, 251-258, 2012.

Peña, J. M. Marginal AMP Chain Graphs. *International Journal of Approximate Reasoning*, 55:1185-1206, 2014.

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

Raiffa, H. *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Addison-Wesley, 1968.

Richardson, T. S. Chain Graphs and Symmetric Associations. In *Learning in Graphical Models*, 231-260, 1998.

Richardson, T. and Spirtes, P. Ancestral Graph Markov Models. *The Annals of Statistics*, 30:962-1030, 2002.

Sadeghi, K. and Lauritzen, S. L. Markov Properties for Mixed Graphs. *Bernoulli*, 20:676-696, 2014.

Sonntag, D. and Peña, J. M. Chain Graph Interpretations and their Relations. In *Proceedings of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty*, 510-521, 2013.

Studený, M. Bayesian Networks from the Point of View of Chain Graphs. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 496-503, 1998.

Studený, M. *Probabilistic Conditional Independence Structures*. Springer, 2005.

Studený, M. and Bouckaert, R. R. On Chain Graph Models for Description of Conditional Independence Structures. *The Annals of Statistics*, 26:1434-1495, 1998.