Unifying DAGs and UGs

Jose M. Peña JOSE.M.PENA@LIU.SE
STIMA, IDA, Linköping University, Sweden

Abstract

We introduce a new class of graphical models that generalizes Lauritzen-Wermuth-Frydenberg chain graphs by relaxing the semi-directed acyclity constraint so that only directed cycles are forbidden. Moreover, up to two edges are allowed between any pair of nodes. Specifically, we present local, pairwise and global Markov properties for the new graphical models and prove their equivalence. We also present an equivalent factorization property.

Keywords: Directed acyclic graphs, undirected graphs, chain graphs, Markov properties.

1. Introduction

Lauritzen-Wermuth-Frydenberg chain graphs (LWF CGs) are usually described as unifying directed acyclic graphs (DAGs) and undirected graphs (UGs) (Lauritzen, 1996, p. 53). However, this is arguable because the only constraint that DAGs and UGs jointly impose is the absence of directed cycles, whereas LWF CGs forbid semi-directed cycles which is a stronger constraint. Moreover, LWF CGs do not allow more than one edge between any pair of nodes. In this work, we consider graphs with directed and undirected edges but without directed cycles. The graphs can have up to two different edges between any pair of nodes. Therefore, our graphs truly unify DAGs and UGs. Hence, we call them UDAGs.

As we will see, UDAGs generalize LWF CGs. Three other such generalizations that can be found in the literature are reciprocal graphs (RGs) (Koster, 1996), acyclic graphs (AGs) (Lauritzen and Sadeghi, 2017) and segregated graphs (SGs) (Shpitser, 2015). The main differences between UDAGs and these three classes of graphical models are the following. UDAGs are not a subclass of RGs or SGs because RGs only allow certain semi-directed cycles, and SGs do not allow semi-directed cycles at all. UDAGs are a subclass of AGs. However, Lauritzen and Sadeghi define a global Markov property for AGs but no local or pairwise Markov property. We define the three properties for UDAGs. Lauritzen and Sadeghi do define though a pairwise Markov property for a subclass of AGs called chain mixed graphs (CMGs), but no local Markov property. Moreover, UDAGs are not a subclass of CMGs because CMGs cannot have semi-directed cycles. In addition to the local, pairwise and global Markov properties, we also define a factorization property for UDAGs. Such a property exists for RGs and SGs but not yet for AGs or CMGs. We also note that the algorithm developed by Sonntag et al. (2015) for learning LWF CGs from data can easily be adapted to learn UDAGs (see Appendix A). To our knowledge, there is no algorithm for learning RGs, SGs, AGs or CMGs. Finally, it is worth mentioning that our work complements that by Richardson (2003), where DAGs and covariance (bidirected) graphs are unified.

The rest of the paper is organized as follows. Section 2 introduces some notation and definitions. Sections 3 and 4 present the global, local and pairwise Markov properties for UDAGs and prove their equivalence. Section 5 does the same for the factorization property. Section 6 closes the paper with some discussion.

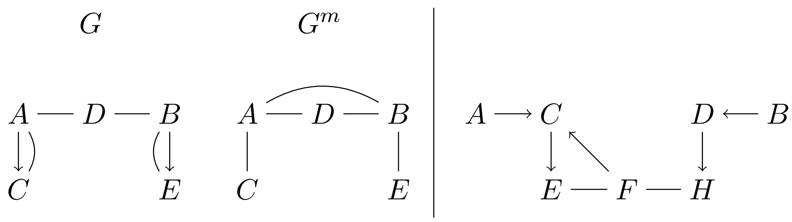


Figure 1: Left: Example of moralization. Right: Example where non-adjacency does not imply separation.

2. Preliminaries

In this section, we introduce some concepts about graphical models. Unless otherwise stated, all the graphs and probability distributions in this paper are defined over a finite set of random variables V. The elements of V are not distinguished from singletons. An UDAG G is a graph with possibly directed and undirected edges but without directed cycles, i.e. $A \to \dots \to A$ is forbidden. There may be up to two different edges between any pair of nodes. Edges between a node and itself are not allowed. We denote by $A \multimap B$ that the edge $A \to B$ or $A - B$ or both are in G. Given an UDAG G, the parents of a set $X \subseteq V$ are $pa(X) = \{B | B \to A$ is in G with $A \in X\}$. The children of X are $ch(X) = \{B | A \to B$ is in G with $A \in X\}$. The neighbors of X are $ne(X) = \{B | A - B$ is in G with $A \in X\}$. The ancestors of X are $an(X) = X \cup \{B | B \multimap \dots \multimap A$ is in G with $A \in X\}$. Moreover, X is called ancestral set if $X = an(X)$. The descendants of X are $de(X) = \{B | A \multimap \dots \multimap B$ is in G with $A \in X\}$. The sets just defined are defined with respect to G. When they are defined with respect to another UDAG, this is indicated with a subscript.

A route between two nodes V_1 and V_n of an UDAG G is a sequence of (not necessarily distinct) edges $E_1, \dots, E_{n-1}$ in G such that E_i links the nodes V_i and V_{i+1}. A route is called a path if the nodes in the route are all different. An undirected route is a route whose edges are all undirected. A section of a route ρ is a maximal undirected subroute of ρ. A section $V_2 - \dots - V_{n-1}$ of ρ is called collider section if $V_1 \to V_2 - \dots - V_{n-1} \leftarrow V_n$ is a subroute of ρ. Given a set $Z \subseteq V$, ρ is said to be Z-active if (i) every collider section of ρ has a node in Z, and (ii) every non-collider section of ρ has no node in Z. Given an UDAG G, the moral graph of G is the UG G^m such that $A - B$ is in G^m if and only if the edge(s) $A \multimap B$ or the route $A \to V_1 - \dots - V_n \leftarrow B$ are in G. In the latter, note that A and/or B may occur in $V_1, \dots, V_n$. See Figure 1 (left) for an example. Given a set $W \subseteq V$, we let G_W denote the subgraph of G induced by W. Given an UG H, we let H^W denote the marginal subgraph of H over W, i.e. the edge $A - B$ is in H^W if and only if the edge $A - B$ is in H or the route $A - V_1 - \dots - V_n - B$ is H with $V_1, \dots, V_n \notin W$. A set of nodes of H is complete if there exists an undirected edge between every pair of nodes in the set. A clique of H is a maximal complete set of nodes. The cliques of H are denoted as $cl(H)$.

Let X, Y, W and Z be disjoint subsets of V. We represent by $X \perp_p Y | Z$ that X and Y are conditionally independent given Z in a probability distribution p. Every probability distribution p satisfies the following four properties: Symmetry $X \perp_p Y | Z \Rightarrow Y \perp_p X | Z$, decomposition $X \perp_p Y \cup W | Z \Rightarrow X \perp_p Y | Z$, weak union $X \perp_p Y \cup W | Z \Rightarrow X \perp_p Y | Z \cup W$, and contraction $X \perp_p Y | Z \cup W \wedge X \perp_p W | Z \Rightarrow X \perp_p Y \cup W | Z$. If p is strictly positive, then it also satisfies the intersection property $X \perp_p Y | Z \cup W \wedge X \perp_p W | Z \cup Y \Rightarrow X \perp_p Y \cup W | Z$.

3. Global Markov Property

Given three disjoint sets $X, Y, Z \subseteq V$ where $X, Y \neq \emptyset$ and Z may be empty, we say that X is separated from Y given Z in an UDAG G, denoted as $X \perp Y | Z$, if every path in $(G_{an(X \cup Y \cup Z)})^m$ between a node in X and a node in Y has a node in Z. As the theorem below proves, this is equivalent to saying that there is no route in G between a node of X and a node of Y that is Z-active. Note that these separation criteria generalize those developed by Lauritzen (1996) and Studený (1998) for LWF CGs. Appendix A shows how to implement the second criterion efficiently.

Theorem 1 *The two separation criteria for UDAGs in the paragraph above are equivalent.*

Proof Assume that there is a Z-active route ρ in G between $A \in X$ and $B \in Y$. Clearly, every node in a collider section is in $an(Z)$. Moreover, every node in a non-collider section is ancestor of A, B or a node in a collider section, which implies that it is in $an(A \cup B \cup Z)$. Therefore, there is a route between A and B in $(G_{an(X \cup Y \cup Z)})^m$. Moreover, the route can be modified into a route ϱ that circumvents Z by noting that there is an edge $V_1 - V_n$ in $(G_{an(X \cup Y \cup Z)})^m$ whenever $V_1 \to V_2 - \cdots - V_{n-1} \leftarrow V_n$ is a subroute of ρ. The route ϱ can be converted into a path by removing loops.

Conversely, assume that there is a path ρ in $(G_{an(X \cup Y \cup Z)})^m$ between $A \in X$ and $B \in Y$ that circumvents Z. Note that ρ can be converted into a route ϱ in G as follows: If the edge $V_1 - V_n$ in ρ was added to $(G_{an(X \cup Y \cup Z)})^m$ because the edge(s) $V_1 \multimap V_n$ or $V_1 \leftarrow V_n$ or the route $V_1 \to V_2 - \cdots - V_{n-1} \leftarrow V_n$ were in $G_{an(X \cup Y \cup Z)}$, then replace $V_1 - V_n$ with $V_1 \multimap V_n$, $V_1 \leftarrow V_n$ or $V_1 \to V_2 - \cdots - V_{n-1} \leftarrow V_n$, respectively. Note that the non-collider sections of ϱ have no node in Z for ρ to circumvent Z, whereas the collider sections of ϱ have all their nodes in $an(X \cup Y \cup Z)$ by definition of $(G_{an(X \cup Y \cup Z)})^m$.

Note that we can assume without loss of generality that all the collider sections of ϱ have some node in $an(Z)$ because, otherwise, if there is a collider section with no node in $an(Z)$ but with some node C in $an(X)$ then there is a route $A' \circ\!- \cdots \circ\!- C$ with $A' \in X$ which can replace the subroute of ϱ between A and C. Likewise for $an(Y)$ and some $B' \in Y$.

Finally, note that every collider section $V_1 \to V_2 - \cdots - V_{n-1} \leftarrow V_n$ of ϱ that has no node in Z must have a node V_i in $an(Z) \setminus Z$ with $2 \leq i \leq n-1$, which implies that there is a route $V_i \multimap \cdots \multimap C$ where C is the only node of the route that is in Z. Therefore, we can replace the collider section with $V_1 \to V_2 - \cdots - V_i \multimap \cdots \multimap C \circ\!- \cdots \circ\!- V_i - \cdots - V_{n-1} \leftarrow V_n$. Repeating this step results in a Z-active route between a node in X and a node in Y. ∎

We say that a probability distribution p satisfies the global Markov property with respect to an UDAG G if $X \perp_p Y | Z$ for all disjoint sets $X, Y, Z \subseteq V$ such that $X \perp Y | Z$. Note that two non-adjacent nodes in G are not necessarily separated. For example, $A \perp B | Z$ does not hold for any $Z \subseteq \{C, D, E\}$ in the UDAG in Figure 1 (left). Likewise, $C \perp D | Z$ does not hold for any $Z \subseteq \{A, B, E, F, H\}$ in the UDAG in Figure 1 (right). Since AGs include UDAGs, the former also have this problem.[1] Although it cannot be solved for general AGs (Lauritzen and Sadeghi, 2017, Figure 6), the problem can be solved for the subclass of CMGs by adding edges without altering the separations represented (Lauritzen and Sadeghi, 2017, Corollary 3.1). Unfortunately, a similar solution does not exist for UDAGs. For example, adding the edge $C \multimap D$ to the UDAG in Figure 1

1. Although we call it a problem, there is nothing wrong with it per se. It is just a counterintuitive feature.

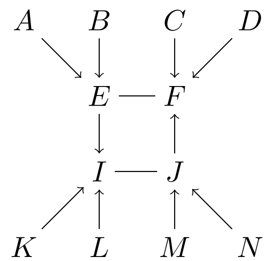


Figure 2: Example of UDAG without Markov equivalent LWF CG.

(right) makes $A \perp B|D$ cease holding, whereas adding the edge $C \leftarrow D$ makes $A \perp B|C \cup F$ cease holding. Adding two edges between C and D does not help either, since one of them must be $C - D$. The following lemma characterizes the problematic pairs of nodes.

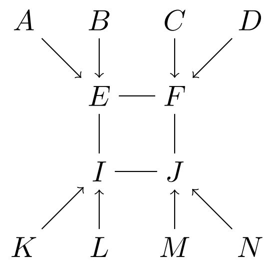
Lemma 2 *Given two non-adjacent nodes V_1 and V_n in an UDAG G, $V_1 \perp V_n|Z$ does not hold for any $Z \subseteq V \setminus (V_1 \cup V_n)$ if and only if the route $V_1 \rightarrow V_2 - \cdots - V_{n-1} \leftarrow V_n$ is in G, and $V_i \in an(V_1 \cup V_n)$ for some $1 < i < n$.*[2]

Proof To prove the if part, assume without loss of generality that $V_i \in an(V_1)$. This together with the route in the lemma imply that G has a route ρ of the form $V_1 \multimapinv \cdots \multimapinv V_i - \cdots - V_{n-1} \leftarrow V_n$. If no node in Z is in ρ, then $V_1 \perp V_n|Z$ does not hold due to ρ. If a node $C \in Z$ is in the subroute $V_i - \cdots - V_{n-1} \leftarrow V_n$ of ρ, then $V_1 \perp V_n|Z$ does not hold due to the route in the lemma. Finally, if a node $C \in Z$ is in the subroute $V_1 \multimapinv \cdots \multimapinv V_i$ of ρ, then $V_1 \perp V_n|Z$ does not hold due to the route $V_1 \rightarrow V_2 - \cdots - V_i \multimap \cdots \multimap C \multimapinv \cdots \multimapinv V_i - \cdots - V_{n-1} \leftarrow V_n$.

To prove the only if part, simply consider $Z = \emptyset$ and note that V_1 and V_n are adjacent in $(G_{an(V_1 \cup V_n)})^m$ only if G has a subgraph of the form described in the lemma. ■

Finally, we show that the independence models representable with UDAGs are a proper superset of those representable with LWF CGs. In particular, we show that there is no LWF CG that is Markov equivalent to the UDAG in Figure 2, i.e. there is no LWF CG that represents exactly the independence model represented by the UDAG.[3] Assume to the contrary that there is a LWF CG H that is Markov equivalent to the UDAG G in the figure. First, note that $A \perp B|\emptyset$ and $A \not\perp B|E$ imply that H must have an induced subgraph $A \rightarrow E \leftarrow B$. Likewise, H must have induced subgraphs $C \rightarrow F \leftarrow D$, $K \rightarrow I \leftarrow L$, and $M \rightarrow J \leftarrow N$. Next, note that $A \perp I|\{E, J\}$ implies that H cannot have an edge $E \leftarrow I$. Likewise, H cannot have an edge $F \rightarrow J$. Note also that $A \perp F|\{B, C, D, E, J\}$ and $D \perp E|\{A, B, C, F, J\}$ imply that H cannot have an edge $E \leftarrow F$ or $E \rightarrow F$. Likewise, H cannot have an edge $I \leftarrow J$ or $I \rightarrow J$. Consequently, H must have a subgraph of the form

2. In the terminology of Lauritzen and Sadeghi (2017), this route is a primitive inducing walk.
3. We do not mean that this is the simplest such UDAG. We choose this because it allows us to prove our point concisely.



and thus $A \perp N | \{B, C, D, E, F, I, J, K, L, M\}$ holds in G but not in H, which is a contradiction.

The following lemma shows that the existence of a semi-directed cycle is not sufficient to declare an UDAG non-equivalent to any LWF CG. Instead, the semi-directed cycle must occur in a particular configuration, e.g. as in Figure 2. For instance, the lemma implies that the UDAG $A \to B \leftarrow C - B$ is Markov equivalent to the LWF CG $A - B - C - A$.

Lemma 3 *Let G denote an UDAG. If (i) W is an ancestral set of nodes in G of size greater than one, and (ii) W is minimal with respect to the property (i), then replacing G_W with $(G_W)^m$ in G results in an UDAG H that is Markov equivalent to G.*

Proof

First, note that H is an UDAG because no directed cycle can be created by replacing G_W with $(G_W)^m$ in G. Now, consider checking whether a separation $X \perp Y | Z$ holds in G and H. Consider the following cases.

(1.) Assume that $an(X \cup Y \cup Z)$ in G includes no node in W. Then, $(G_{an(X \cup Y \cup Z)})^m = (H_{an(X \cup Y \cup Z)})^m$ and thus $X \perp Y | Z$ holds in both G and H or in none.

(2.) Assume that $an(X \cup Y \cup Z)$ in G includes exactly one of the nodes in W, here denoted by A. Then, $an(X \cup Y \cup Z)$ in H includes all the nodes in W because W is connected in G since, otherwise, it is not minimal which is a contradiction. Moreover, note that A is the only node shared by $(G_{an(X \cup Y \cup Z)})^m$ and $(G_W)^m$ because, otherwise, there must be a second node in W that is in $an(X \cup Y \cup Z)$ in G, which is a contradiction. Then, $(H_{an(X \cup Y \cup Z)})^m = (G_{an(X \cup Y \cup Z)})^m \cup (G_W)^m$ and thus $X \perp Y | Z$ holds in both G and H or in none.

(3.) Assume that $an(X \cup Y \cup Z)$ in G includes more than one of the nodes in W. Then, $(G_{an(X \cup Y \cup Z)})^m$ includes all the nodes in W because, otherwise, W is not minimal which is a contradiction. Then, $(G_{an(X \cup Y \cup Z)})^m = (H_{an(X \cup Y \cup Z)})^m$ and thus $X \perp Y | Z$ holds in both G and H or in none. To see it, note that by construction $(G_{an(X \cup Y \cup Z)})^m$ and $(H_{an(X \cup Y \cup Z)})^m$ differ only if the former has an edge $V_1 - V_n$ that the latter does not have. This occurs only if $G_{an(X \cup Y \cup Z)}$ has a route $V_1 \to V_2 - \cdots - V_{n-1} \leftarrow V_n$, whereas $H_{an(X \cup Y \cup Z)}$ has a route $V_1 - V_2 - \cdots - V_{n-1} \leftarrow V_n$ or $V_1 - V_2 - \cdots - V_{n-1} - V_n$. The former case implies that $V_1, \ldots, V_{n-1}$ are in W whereas V_n is not. This contradicts the fact that W is an ancestral set. The latter case implies that $V_1, \ldots, V_n$ are in W, which implies that $V_1 - V_n$ is in H, which is a contradiction.

■

Note that the condition in the lemma above is sufficient but not necessary. For instance, the UDAGs $G = \{A \to B \to C\}$ and $H = \{A \to B - C\}$ are Markov equivalent, although $W = \{B, C\}$ is not ancestral. Or the UDAGs $G = \{A \to B \leftarrow C - A\}$ and $H = \{A - B - C - A\}$ are Markov equivalent, although $W = \{A, B, C\}$ is ancestral but not minimal.

4. Local and Pairwise Markov Properties

We say that a probability distribution p satisfies the local Markov property with respect to an UDAG G if for any ancestral set W,

$$A \perp_p W \setminus (A \cup ne_{(G_W)^m}(A)) | ne_{(G_W)^m}(A)$$

for any $A \in W$. Similarly, we say that a probability distribution p satisfies the pairwise Markov property with respect to G if for any ancestral set W,

$$A \perp_p B | W \setminus (A \cup B)$$

for any $A, B \in W$ such that $B \notin ne_{(G_W)^m}(A)$.

Theorem 4 *Given a probability distribution p satisfying the intersection property, p satisfies the local Markov property with respect to an UDAG G if and only if it satisfies the pairwise Markov property with respect to G.*

Proof The if part follows by repeated application of the intersection property. The only if part follows by the weak union property. ■

Theorem 5 *Given a probability distribution p satisfying the intersection property, p satisfies the pairwise Markov property with respect to an UDAG G if and only if it satisfies the global Markov property with respect to G.*

Proof The if part is trivial. To prove the only if part, let $W = an(X \cup Y \cup Z)$ and note that the pairwise and global Markov properties are equivalent for UGs (Lauritzen, 1996, Theorem 3.7). ■

Note that the local Markov property for LWF CGs specifies a single independence for each node (Lauritzen, 1996, p. 55). However, the local Markov property for UDAGs specifies many more independences, specifically an independence for any node and ancestral set containing the node. All in all, our local Markov property serves its purpose, namely to identify a subset of the independences specified by the global Markov property that implies the rest. In the next section, we show how to reduce this subset.

4.1 Reduction

The number of independences specified by the local Markov property for UDAGs can be reduced by considering only maximal ancestral sets for any node A, i.e. those ancestral sets W' such that $A \in W'$ and $ne_{(G_{W'})^m}(A) \subset ne_{(G_{W''})^m}(A)$ for any ancestral set W'' such that $W' \subset W''$. Note that

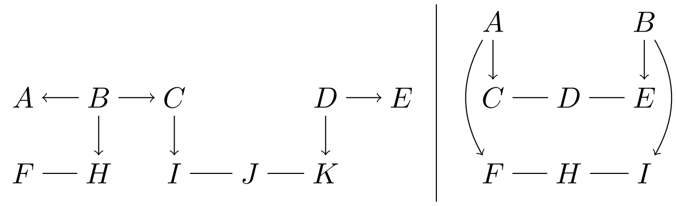


Figure 3: Left: Example where the local Markov property can be improved by considering only maximal ancestral sets. Right: Example where the factorization property can be improved by considering only maximal ancestral sets.

there may be several maximal ancestral sets W' for A, each for a different set $ne_{(G_{W'})^m}(A)$ as will be shown. The independences for the non-maximal ancestral sets follow from the independences for the maximal ancestral sets by the decomposition property. In other words, for any non-maximal ancestral set W and $A \in W$,

$$A \perp_p W \setminus (A \cup ne_{(G_W)^m}(A)) | ne_{(G_W)^m}(A)$$

follows from

$$A \perp_p W' \setminus (A \cup ne_{(G_{W'})^m}(A)) | ne_{(G_{W'})^m}(A)$$

where W' is the maximal ancestral set for A such that $ne_{(G_W)^m}(A) = ne_{(G_{W'})^m}(A)$. In the UDAG in Figure 3 (left), for instance, $W_1 = \{A, B, C, D\}$, $W_2 = \{A, B, C, D, E, I, J, K\}$, and $W_3 = \{A, B, C, D, E, F, H, I, J, K\}$ are three ancestral sets that contain the node B. However, only W_2 and W_3 are maximal for B: W_1 is not maximal because $W_1 \subset W_2$ but $ne_{(G_{W_1})^m}(B) = ne_{(G_{W_2})^m}(B)$, and W_2 is maximal because $W_2 \subset W_3$ and $ne_{(G_{W_2})^m}(B) \subset ne_{(G_{W_3})^m}(B)$. Note that W_1 specifies $B \perp_p D | \{A, C\}$, and W_2 specifies $B \perp_p \{D, E, I, J, K\} | \{A, C\}$. Clearly, the latter independence implies the former by the decomposition property. Therefore, there is no need to specify both independences, as the local Markov property does. It suffices to specify just the second.

A more convenient characterization of maximal ancestral sets is the following. An ancestral set W' is maximal for $A \in W'$ if and only if $W' = V \setminus [(ch(A) \cup de(ch(A))) \setminus W']$. To see it, note that $B \in ne_{(G_{W'})^m}(A)$ if and only if the edge(s) $A \multimap B$ or $A \to B$ or the route $A \to V_1 - \cdots - V_n \leftarrow B$ are in $G_{W'}$. Note that all the parents and neighbors of A are in W', because W' is ancestral. However, if there is some child B of A that is not in W', then any ancestral set W'' that contains W' and B or any node that is a descendant of B will be such that $ne_{(G_{W'})^m}(A) \subset ne_{(G_{W''})^m}(A)$.

The number of independences specified by the pairwise Markov property can also be reduced by considering only maximal ancestral sets. This can be proven in the same way as Theorem 4.

5. Factorization Property

Theorem 6 *Given a probability distribution p satisfying the intersection property, p satisfies the pairwise Markov property with respect to an UDAG G if and only if for any ancestral set W,*

$$p(W) = \prod_{K \in cl((G_W)^m)} \varphi(K)$$

where $\varphi(K)$ is a non-negative function.

Proof It suffices to recall the equivalence between the pairwise Markov property and the factorization property for UGs (Lauritzen, 1996, Theorem 3.9). ■

The following theorem presents a necessary but not sufficient factorization property for UDAGs. However, compared to that in Theorem 6, it is simpler and resembles the factorization property for LWF CGs. Given an UDAG G, let $C_1, \dots, C_t$ denote all the sets of nodes such that (i) every node in C_i is an ancestor of the rest of the nodes in C_i, and (ii) C_i is maximal with respect to the property (i). Assume that the sets are sorted such that all the edges between a node in C_i and a node in C_j with $i < j$ are directed from C_i to C_j. Let $bd(C_i) = pa(C_i) \setminus C_i$. Moreover, let $(G_{C_i \cup bd(C_i)})^*$ be the result of adding undirected edges to $(G_{C_i \cup bd(C_i)})^m$ until $bd(C_i)$ is a complete set. Note that for LWF CGs, the sets C_i correspond to the chain components, $bd(C_i) = pa(C_i)$ and $(G_{C_i \cup bd(C_i)})^* = (G_{C_i \cup bd(C_i)})^m$. For instance, in the UDAG in Figure 1 (right) we have that $C_1 = \{A\}$, $C_2 = \{B\}$, $C_3 = \{D\}$ and $C_4 = \{C, E, F, H\}$, and $bd(C_1) = \emptyset$, $bd(C_2) = \emptyset$, $bd(C_3) = \{B\}$ and $bd(C_4) = \{A, D\}$.

Theorem 7 *Let p be a probability distribution satisfying the intersection property. If p satisfies the pairwise Markov property with respect to an UDAG G, then*

$$p(V) = \prod_i p(C_i | bd(C_i)) = \prod_i \prod_{K \in cl((G_{C_i \cup bd(C_i)})^*)} \varphi(K)$$

where $\varphi(K)$ is a non-negative function.

Proof The first equality follows from the fact that $C_i \perp (\cup_{j<i} C_j \setminus bd(C_i)) | bd(C_i)$ and the fact that p satisfies the global Markov property with respect to G by Theorem 5. To prove the second equality for $i = t$, note that p satisfies the pairwise Markov property with respect to G^m, because V is an ancestral set. Then, p satisfies the global Markov property with respect to G^m by Theorem 5. Now, add undirected edges to G^m until $bd(C_t)$ is a complete set, and call the resulting undirected graph H. Note that p satisfies the global Markov property with respect to H, because H is a supergraph of G^m. Then, $p(C_t, bd(C_t))$ satisfies the global Markov property with respect to $H_{C_t \cup bd(C_t)}$ (Frydenberg, 1990b, Proposition 2.2). This implies the second equality in the theorem because (i) $(G_{C_t \cup bd(C_t)})^* = H_{C_t \cup bd(C_t)}$, (ii) $p(C_t, bd(C_t)) = \prod_{K \in cl((G_{C_i \cup bd(C_i)})^*)} \phi(K)$ with the normalization constant being absorbed in one of the functions in the product (Lauritzen, 1996, Theorem 3.9), (iii) $p(C_t | bd(C_t)) = p(C_t, bd(C_t)) / p(bd(C_t))$, and (iv) $bd(C_t)$ is a complete set in $(G_{C_t \cup bd(C_t)})^*$. Finally, note that $V \setminus C_t$ is an ancestral set and, thus, $p(V \setminus C_t)$ satisfies the pairwise Markov property with respect to $G_{V \setminus C_t}$. Then, repeating the reasoning above for $p(V \setminus C_t)$ and $G_{V \setminus C_t}$ proves the second equality in the theorem for all i. ■

For instance, in the UDAG in Figure 1 (right) we have that

$$\begin{aligned} p(V) =& p(A)p(B)p(D|B)p(CEFH|AD) \\ =& \varphi(A)\varphi(B)\varphi(DB)\varphi(CAD)\varphi(CFA)\varphi(CEF)\varphi(FH)\varphi(HD). \end{aligned}$$

5.1 Reduction

The number of factorizations specified by the factorization property for UDAGs in Theorem 6 can be reduced by considering only maximal ancestral sets, i.e. those ancestral sets W' such that $(G_{W'})^m$

is a proper subgraph of $((G_{W''})^m)^{W'}$ for any ancestral set W'' such that $W' \subset W''$. These maximal ancestral sets do not necessarily coincide with the ones defined in Section 4.1. The factorizations for the non-maximal ancestral sets follow from the factorizations for the maximal ancestral sets. To see it, note that for any non-maximal ancestral set W, the probability distribution $p(W)$ can be computed by marginalization from $p(W')$ where W' is any maximal ancestral set such that $((G_{W'})^m)^W = (G_W)^m$. Note also that $p(W)$ factorizes according to $((G_{W'})^m)^W$ and thus according to $(G_W)^m$, by (Studený, 1997, Lemma 3.1) and (Lauritzen, 1996, Theorems 3.7 and 3.9). In the UDAG in Figure 3 (right), for instance, $W_1 = \{A, B\}$, $W_2 = \{A, B, C, D, E\}$, and $W_3 = \{A, B, C, D, E, F, H, I\}$ are three ancestral sets. However, only W_1 and W_3 are maximal: W_2 is not maximal because $W_2 \subset W_3$ but $((G_{W_3})^m)^{W_2} = (G_{W_2})^m$, and W_1 is maximal because $W_1 \subset W_3$ and $(G_{W_1})^m$ is a proper subgraph of $((G_{W_3})^m)^{W_1}$. Note that W_3 specifies

$$p(W_3) = \varphi_3(A,B)\varphi_3(A,C)\varphi_3(B,E)\varphi_3(C,D)\varphi_3(D,E)\varphi_3(A,F)\varphi_3(B,I)\varphi_3(F,H)\varphi_3(H,I)$$

and W_2 specifies

$$p(W_2) = \varphi_2(A,B)\varphi_2(A,C)\varphi_2(B,E)\varphi_2(C,D)\varphi_2(D,E).$$

Clearly, the former factorization implies the latter by taking

$$\varphi_2(A,B) = \varphi_3(A,B) \sum_{F,H,I} \varphi_3(A,F)\varphi_3(B,I)\varphi_3(F,H)\varphi_3(H,I)$$

$$\varphi_2(A,C) = \varphi_3(A,C); \varphi_2(B,E) = \varphi_3(B,E); \varphi_2(C,D) = \varphi_3(C,D); \varphi_2(D,E) = \varphi_3(D,E).$$

Therefore, there is no need to specify both factorizations, as the factorization property does. It suffices to specify just the first.

A more convenient characterization of maximal ancestral sets is the following. An ancestral set W' is maximal if and only if $pa(A \cup (an(A) \setminus W')) \cap W'$ is not a complete set in $(G_{W'})^m$ for any node $A \in V \setminus W'$.[4] To see it, note that any ancestral set W'' that contains $W' \cup A$ will also contain $an(A) \setminus W'$. Note also that no node $B \in A \cup (an(A) \setminus W')$ has a neighbor or child in W' because, otherwise, $B \in W'$ which is a contradiction. So, any such node B can only have parents in W'. Moreover, since $pa(A \cup (an(A) \setminus W')) \cap W'$ is not a complete set in $(G_{W'})^m$, there must be two nodes in $pa(A \cup (an(A) \setminus W')) \cap W'$ that are not adjacent in $(G_{W'})^m$. However, there is a path between these two nodes in $(G_{W''})^m$ through A, which implies that $(G_{W'})^m$ is a proper subgraph of $((G_{W''})^m)^{W'}$.

6. Discussion

We have introduced UDAGs, a new class of graphical models that unifies DAGs and UGs since it just forbids directed cycles and it allows up to two edges between any pair of nodes. We have presented local, pairwise and global Markov properties for UDAGs and proved their equivalence for positive probability distributions. We have also presented an equivalent factorization property.

Some natural question to tackle in the future follow. We have shown that UDAGs can represent independence models that LWF CGs cannot. It would be interesting to know how much more expressive UDAGs are. It would also be interesting to know if these independence models are

4. In the terminology of Frydenberg (1990a), $A \cup an(A) \setminus W'$ is a non-simplicial set in $(G_{W'})^m$.

probabilistic, i.e. if for any UDAG there is a probability distribution that is faithful to it. We would also like to characterize when two UDAGs are Markov equivalent. Finally, we are also interested in studying methods for parameterizing the factorization for UDAGs proposed.

Appendix A. Separation and Learning Algorithms

In this appendix, we describe an algorithm to check in a finite number of steps whether $X \perp Y | Z$ holds in G. Note that the UDAG G may have infinite routes. The algorithm is essentially the same as the one developed by Studený (1998) for LWF CGs, which was later slightly improved by Sonntag et al. (2015). The algorithm basically consists in repeatedly executing some rules to build the sets $U_1, U_2, U_3 \subseteq V$, which can be described as follows.

- $B \in U_1$ if and only if there exists a Z-active route between $A \in X$ and B in G which ends with the subroute $V_i \rightarrow V_{i+1} - \cdots - V_{i+k} = B$ with $k \geq 1$.
- $B \in U_2$ if and only if there exists a Z-active route between $A \in X$ and B in G which does not end with the subroute $V_i \rightarrow V_{i+1} - \cdots - V_{i+k} = B$ with $k \geq 1$.
- $B \in U_3$ if and only if there exists a node $C \in U_1 \cup U_2$ and a route $C = V_1 \rightarrow V_2 - \cdots - V_k = B$ in G with $k \geq 2$ such that $\{V_2, \ldots, V_k\} \cap Z \neq \emptyset$.

The algorithm starts with $U_1 = U_3 = \emptyset$ and $U_2 = X$. The algorithm executes the following rules until U_1, U_2 and U_3 cannot be further enlarged.

- $C \in U_2$, $C \multimap D$ is in G, and $D \notin Z \Rightarrow D \in U_2$.
- $C \in U_1 \cup U_2$, $C \rightarrow D$ is in G, and $D \notin Z \Rightarrow D \in U_1$.
- $C \in U_1$, $C - D$ is in G, and $D \notin Z \Rightarrow D \in U_1$.
- $C \in U_1 \cup U_2$, $C \rightarrow D$ is in G, and $D \in Z \Rightarrow D \in U_3$.
- $C \in U_1$, $C - D$ is in G, and $D \in Z \Rightarrow D \in U_3$.
- $C \in U_3$, and $C - D$ is in $G \Rightarrow D \in U_3$.
- $C \in U_3$, $C \leftarrow D$ is in G, and $D \notin Z \Rightarrow D \in U_2$.

One can prove that, when the algorithm halts, there is a Z-active route between each node in $U_1 \cup U_2$ and some node in X. The proof is identical to the one for LWF CGs by (Studený, 1998, Lemma 5.2) and (Sonntag et al., 2015, Proposition 1). Therefore, $X \perp Y | Z$ if and only if $Y \subseteq V \setminus (U_1 \cup U_2)$.

We now use the result above to develop an algorithm for learning UDAGs from data via answer set programming (ASP). The algorithm is essentially the same as the one developed by Sonntag et al. (2015) for learning LWF CGs. ASP represents constraints in terms of first-order logical rules. Each rule in the constraint declaration is of the form `head :- body`. The head contains an atom, i.e. a fact. The body may contain several literals, i.e. negated and non-negated atoms. Intuitively, the rule is a justification to derive the head if the body is true. The body is true if its non-negated atoms can be derived, and its negated atoms cannot. A rule with only the head is an atom. A rule without the head is a hard-constraint, meaning that satisfying the body results in a contradiction.

Table 1: Algorithm for learning UDAGs.

```
% input predicate: ind(A,B,Z), the nodes A and B are independent given
%                  the set of nodes Z

#const n=7.
node(1..n).
set(0..(2**n)-1).

% edges
{ line(A,B) } :- node(A), node(B), A != B.                              % rule 4
{ arrow(A,B) } :- node(A), node(B), A != B.
line(A,B) :- line(B,A).                                                 % rule 6
:- arrow(A,B), arrow(B,A).

% directed acyclity
ancestor(A,B) :- arrow(A,B).                                            % rule 8
ancestor(A,B) :- ancestor(A,C), ancestor(C,B).
:- ancestor(A,B), arrow(B,A).

% set membership
in(A,Z) :- node(A), set(Z), 2**(A-1) & Z != 0.                          % rule 11
out(A,Z) :- node(A), set(Z), 2**(A-1) & Z = 0.

% rules
inU2(A,A,Z) :- node(A), set(Z), out(A,Z).                               % rule 13
inU2(A,D,Z) :- inU2(A,C,Z), arrow(D,C), out(D,Z).
inU2(A,D,Z) :- inU2(A,C,Z), line(D,C), out(D,Z).
inU1(A,D,Z) :- inU1(A,C,Z), arrow(C,D), out(D,Z).
inU1(A,D,Z) :- inU2(A,C,Z), arrow(C,D), out(D,Z).
inU1(A,D,Z) :- inU1(A,C,Z), line(C,D), out(D,Z).
inU3(A,D,Z) :- inU1(A,C,Z), arrow(C,D), in(D,Z).
inU3(A,D,Z) :- inU2(A,C,Z), arrow(C,D), in(D,Z).
inU3(A,D,Z) :- inU1(A,C,Z), line(C,D), in(D,Z).
inU3(A,D,Z) :- inU3(A,C,Z), line(C,D).
inU2(A,D,Z) :- inU3(A,C,Z), arrow(D,C), out(D,Z).

% active routes
act(A,B,Z) :- inU1(A,B,Z), A != B.                                      % rule 24
act(A,B,Z) :- inU2(A,B,Z), A != B.

% satisfy all the dependences
:- not ind(A,B,Z), not act(A,B,Z), node(A), node(B), set(Z), A != B,
   out(A,Z), out(B,Z).                                                  % rule 26

% minimize the number of lines/arrows
:~ line(A,B), A < B. [1,A,B,1]                                          % rule 27
:~ arrow(A,B). [1,A,B,2]

% show results
#show.
#show arrow(A,B) : arrow(A,B).
#show line(A,B) : line(A,B), A < B.
```

Soft-constraints are encoded as rules of the form `:~ body. [W]`, meaning that satisfying the body results in a penalty of W units. The ASP solver returns the solutions that meet the hard-constraints and minimize the total penalty due to the soft-constraints. A popular ASP solver is `clingo` (Gebser et al., 2011).

Table 1 shows the ASP encoding of our learning algorithm. The input to the algorithm is the set of independences in the probability distribution at hand, e.g. as determined from some available data. These are represented as a set of predicates `ind(A,B,Z)` indicating that the nodes A and B are independent given the set of nodes Z. It is known that these pairwise independences (also called elementary triplets) uniquely identify the rest of independences in the distribution, or in a semi-graphoid for that matter (Studený, 2005, Lemma 2.2). The predicates `node(A)` and `set(Z)` represent that A is the index of a node and Z is the index of a set of nodes. The predicates `line(A,B)` and `arrow(A,B)` represent that there is an undirected and directed edge from the node A to the node B. The rules 4 and 5 encode a non-deterministic guess of the edges, which means that the ASP solver will implicitly consider all possible UDAGs during the search, hence the exactness of the search. The rules 6 and 7 enforce the fact that undirected edges are symmetric and there is at most one directed edge between two nodes. The predicate `ancestor(A,B)` represents that the node A is an ancestor of the node B. The rules 8-10 enforce that there are no directed cycles. The predicates in the rules 11 and 12 represent whether a node A is or is not in a set of nodes Z. The rules 13-23 encode the separation criterion for UDAGs as it was described above. Specifically, the predicate `inU1(A,D,Z)` represents that there is a Z-active route from the node A to the node D that warrants the inclusion of D in the set U_1. Similarly for the predicates `inU2(A,D,Z)` and `inU3(A,D,Z)`. The predicate `act(A,B,Z)` in the rules 24 and 25 represents that there is a Z-active route between the node A and the node B. The rule 26 enforces that each dependence in the input must correspond to an active route. The rules 27 and 28 represent a penalty of one unit per edge. Other penalty rules can be added similarly. By calling the ASP solver, the solver will essentially perform an exhaustive search over the space of UDAGs and return the sparsest minimal independence map.

References

M. Frydenberg. Marginalization and Collapsibility in Graphical Interaction Models. *Annals of Statistics*, 18:790–805, 1990a.

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

M. Gebser, B. Kaufmann, R. Kaminski, M. Ostrowski, T. Schaub, and M. Schneider. Potassco: The Potsdam Answer Set Solving Collection. *AI Communications*, 24:107–124, 2011.

J. T. A. Koster. Markov Properties of Nonrecursive Causal Models. *Annals of Statistics*, 24:2148–2177, 1996.

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

S. L. Lauritzen and K. Sadeghi. Unifying Markov Properties for Graphical Models. *Annals of Statistics*, 2017.

T. Richardson. Markov Properties for Acyclic Directed Mixed Graphs. *Scandinavian Journal of Statistics*, 30:145–157, 2003.

I. Shpitser. Segregated Graphs and Marginals of Chain Graph Models. In *Advances in Neural Information Processing Systems*, pages 1720–1728, 2015.

D. Sonntag, M. Järvisalo, J. M. Peña, and A. Hyttinen. Learning Optimal Chain Graphs with Answer Set Programming. In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence*, pages 822–831, 2015.

M. Studený. On Marginalization, Collapsibility and Precollapsibility. In *Distributions with Given Marginals and Moment Problems*, pages 191–198, 1997.

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

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