Chain graphs: semantics and expressiveness*

Remco R. Bouckaert[1] and Milan Studený[2]

[1] Utrecht University, Department of Computer Science, P.O.Box 80.089, 3508 TB Utrecht, The Netherlands, remco@cs.ruu.nl
[2] Inst. of Inform. Theory & Autom., Czech Academy of Sciences, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic, studeny@utia.cas.cz

Abstract. A *chain graph* (CG) is a graph admitting both directed and undirected edges with forbidden directed cycles. It generalizes both the concept of undirected graph (UG) and the concept of directed acyclic graph (DAG). CGs can be used efficiently to store *graphoids*, that is, independency knowledge of the form "X is independent of Y given Z" obeying a set of five properties (axioms).
Two equivalent criteria for reading independencies from a CG are formulated, namely the *moralization criterion* and the *separation criterion*. These criteria give exactly the graphoid closure of the input list for the CG. Moreover, a construction of a CG from a graphoid (through an input list), which produces a minimal I-map of that graphoid, is given.

1 Introduction

Using graphs to describe independency structure arising among variables has a long and rich tradition. One can distinguish two classic approaches (for details see the book [11]): using *undirected graphs* (UGs), called also Markov networks, or using *directed acyclic graphs* (DAGs), named also Bayesian networks, recursive models or influence diagrams. The aim is to describe efficiently independency models in the form of lists of statements "X is independent of Y given Z", where X, Y, Z are disjoint sets of variables. Such structures can arise in several calculi for dealing with uncertainty in artificial intelligence: in probabilistic reasoning, in theory of natural conditional functions known also as kappa-calculus, in possibility theory or Dempster-Shafer theory of evidence (for overview see [12]) but also in the theory of relational databases. Of course, different calculi produce different independency models, but in case of non-extreme knowledge representation they share five properties which define the class of *graphoids*.

Graphoids can be sometimes described graphically. Thus, every UG defines by means of *separation criterion* an independency model which is a graphoid. The use of UGs in probabilistic reasoning justified by the result from [5], where every such UG-model is shown to be a probabilistic independency model. Nevertheless, a lot of graphoids (even probabilistic models) have no UG representation (that is, are not UG-models). Therefore Pearl [11] proposed to approximate graphoids

* This work was partially supported by the grants: GA AVČR no. 275105 and CEC no. CIPA3511CT930053.

by their contained UG-models (I-maps) and showed that for every graphoid M there exist a unique maximal UG-model contained in M, called *minimal I-map of* M.

Evolution of DAG-models was more complicated. Originally, DAGs were used to describe recursive factorizations of probability distributions. But such a factorization is equivalent to the requirement that the considered distribution complies with a set of independencies called often *causal input list*. Nevertheless, the distribution usually complies with many other independencies outside the input list. A lot of effort was exerted to achieve a graphical criterion which makes it possible to read from the DAG all independencies in the factorizable distribution.

In fact, two equivalent criteria were found. Lauritzen *et. al.* [10] generalized an incomplete criterion from [6] and formulated a *moralization criterion* where testing consist of 3 steps: restriction of the DAG to certain set of nodes, transforming it properly to an UG (called *moral graph*), and using the separation criterion for UGs with respect to the moral graph. The group around Pearl developed a direct *separation criterion* [3], for this purpose they introduced the concept of *d-separation* (*d-* stands for directional) for paths in DAGs. It was shown that the criteria are equivalent [10] and that they give exactly the graphoid closure of the input list [13]. Finally, the criteria were shown to be complete for probabilistic reasoning by showing that every independency model defined by the separation criterion is a probabilistic model [4]. Thus, DAG-models were established and their use in probabilistic reasoning was justified. Like in case of UGs Pearl [11] considered the problem of inner approximation of graphoids by DAG-models. In contrast to the case of UGs several maximal DAG-models contained in a graphoid (*minimal I-maps*) may exist. In fact, any ordering of variables can generate such a minimal I-map, the corresponding construction is given in [13].

This paper deals with *chain graphs* (CGs) which allow both directed and undirected edges. This class of graphs, introduced by Lauritzen and Wermuth [7], generalizes both UGs and DAGs. To establish semantics of CGs one should associate an independency model to every CG. Some steps were already made. Lauritzen and Wermuth [8] intended to use CGs to describe independency models for positive distributions and introduced the concept of *chain Markov property* which is an analogy of the concept of causal input list for DAGs. Frydenberg [2] generalized the concept of moral graph and introduced a *moralization criterion* for reading independencies from a CG.

In this paper some of above mentioned results concerning UGs and DAGs are extended to the case of CGs. We introduce the concept of *c-separation* (chain separation) for trails in CGs, which generalizes both separation in UGs and d-separation in DAGs. This gives a direct *separation criterion* for reading independencies from a CG. The main result of the contribution says that an independency statement belongs to the graphoid closure of the *input list* for a CG iff it is derived by the moralization criterion, which is equivalent to the separation criterion. Moreover, the construction of a minimal I-map from [13] is generalized to the case of CGs.

2 Independency models

Throughout the paper, in apposite situations, we will use a reduced notation: juxtaposition XY instead of $X \cup Y$, u instead of $\{u\}$ and $X - Y - u$ instead of $X \setminus (Y \cup \{u\})$.

Supposing N is a nonempty finite set the symbol $T(N)$ will denote the class of all triplets $\langle X, Y|Z\rangle$ of disjoint subsets of N whose first two components X and Y are nonempty. An *independency model* over N is a subset of $T(N)$. It is called *graphoid* iff it satisfies the following properties:

$\langle X, Y|Z\rangle \ \rightarrow \ \langle Y, X|Z\rangle$ symmetry
$\langle X, YW|Z\rangle \ \rightarrow \ \langle X, W|Z\rangle$ decomposition
$\langle X, YW|Z\rangle \ \rightarrow \ \langle X, Y|WZ\rangle$ weak union
$[\langle X, Y|WZ\rangle \ \& \ \langle X, W|Z\rangle] \ \rightarrow \ \langle X, YW|Z\rangle$ contraction
$[\langle X, Y|WZ\rangle \ \& \ \langle X, W|YZ\rangle] \ \rightarrow \ \langle X, YW|Z\rangle$ intersection

Having a set $L \subset T(N)$ its *graphoid closure*, denoted by $gr(L)$ consists of all triplets in $T(N)$ derivable from L by means of consecutive application of graphoid properties.

3 Graphs

A graph is a couple (N, E) where N is a nonempty finite set of *nodes* and E is a set of *edges*, that is, two-element subsets of N. In this paper we consider several types of edges (every edge belongs exclusively to one of possible types) and this gives several types of graphs. An *undirected graph* (UG) admits only undirected edges, called *links*. We will write $u - v$ to denote that there exists a link between a node u and a node v. A *directed acyclic graph* (DAG) is a graph having only directed edges called *arcs* (we will write $u \rightarrow v$ to denote that there exists an *arc* from a node u to a node v) such that there exists no directed cycle in the graph (that is, sequence of distinct nodes $v_1, \ldots, v_k$, $k \geq 2$ with $v_i \rightarrow v_{i+1}$, $i = 1, \ldots, k-1$ and $v_k \rightarrow v_1$).

A *chain graph* (CG) admits both links and arcs. It is required that the set of nodes can be partitioned into ordered disjoint (nonempty) subsets $B_1, \ldots, B_n$, $n \geq 1$ called *blocks* in such a way that the types of edges are determined as follows:

(i) if $\{u, v\}$ is an edge with $u, v \in B_i$ then $u - v$,
(ii) if $\{u, v\}$ is an edge with $u \in B_i, v \in B_j, i < j$ then $u \rightarrow v$.

Note that CGs were characterized in [2] as graphs not having any directed cycles, but the definition above is more suitable for our purposes. It is evident that CGs involve both UGs and DAGs. Every ordered partitioning satisfying (i)-(ii) will be called a *chain* for the CG. Of course, a CG admits several chains.

A *subgraph* of a graph $G = (N, E)$ is a graph $H = (V, F)$ with $V \subset N$ and $F \subset E$; its *restriction* to a nonempty set $T \subset N$ is the subgraph $G_T = (T, E_T)$, where $E_T = \{\, \{u, v\} \in E \,;\, u, v \in T \,\}$. Of course, the types respectively orientations of edges remain unchanged. Let us mention that a restriction of a CG is again a CG.

A *path* in a CG is a sequence of its distinct nodes $v_1, \dots, v_k$, $k \geq 1$ such that $\forall i = 1, \dots, k-1$ $\{v_i, v_{i+1}\}$ is an edge. We will say that it is a path *from* a node u *to* a node w iff $v_1 = u$ and $v_k = w$. We will say that a path $v_1, \dots, v_k$, $k \geq 1$ *meets* a set of nodes Z iff $\{v_1, \dots v_k\} \cap Z \neq \emptyset$. The path is *undirected* iff $\forall i = 1, \dots, k-1$ $v_i - v_{i+1}$. The path is *descending* iff $\forall i = 1, \dots, k-1$ either $v_i \to v_{i+1}$ or $v_i - v_{i+1}$. If there exists a descending path from a node u to a node v, then v is a *descendant* of u, or dually u is an *ancestor* of v. The symbol $ds_G(u)$ will denote the set of descendants of u; $ds_G(X)$ is the union of $ds_G(u)$'s for $u \in X$ (X is a set of nodes). It is worthwhile to realize the following simple fact.

FACT 1 If there exists an undirected path from u to v, then $ds_G(u) = ds_G(v)$.

Similarly, $an_G(X)$ denotes the set of ancestors of nodes from X. We will omit the symbol of the graph G if it will be clear from the context. A set of nodes X is *ancestral* (in G) iff it contains ancestors of its nodes, that is, $an_G(X) \subset X$.

4 Moralization criterion

The moralization criterion for CGs is based on the classic separation criterion for UGs. Thus, we recall that a triplet $\langle X, Y|Z \rangle \in T(N)$ is *represented* in an UG $H = (N, E)$, denoted by $\langle X, Y|Z \rangle_H$, iff every path in H from a node of X to a node of Y in G meets Z.

Given a CG $G = (N, E)$ its *moral graph*, denoted by G^{mor}, is an UG having the same set of nodes as G, but the set of links established as follows: $u - v$ in G^{mor} iff $u - v$ in G or $u \to v$ in G or $u \leftarrow v$ in G or there exists a path $v_1, \dots, v_k$, $k \geq 3$ from u to v in G such that $v_1 \to v_2, \forall i = 2, \dots, k-2 \quad v_i - v_{i+1}, v_{k-1} \leftarrow v_k$.

Let $G = (N, E)$ be a CG, $\langle X, Y|Z \rangle \in T(N)$ and H be the moral graph of $G_{an(XYZ)}$. We will say that $\langle X, Y|Z \rangle$ is *represented* in G according to the *moralization criterion* and write $\langle X, Y|Z \rangle_G^{mor}$ iff $\langle X, Y|Z \rangle_H$. Let us mention that the moral graph H depends on $\langle X, Y|Z \rangle$. The reader can verify that this moralization criterion specified to DAGs gives exactly the criterion from [10].

5 Separation criterion

To formulate the separation criterion for CGs we have to introduce some special graphical concepts. Given a CG, a *slide* from a node u to a node w is a path $v_1, \dots, v_k$, $k \geq 2$ such that $u = v_1 \to v_2$, $\forall i = 2, \dots, k-1$ $v_i - v_{i+1}$ and $v_k = w$. A *trail* in a CG is a sequence of its nodes $v_1, \dots, v_k$, $k \geq 1$ such that
(i) $\forall i = 1, \dots, k-1$ $\{v_i, v_{i+1}\}$ is an edge of G,
(ii) $\forall i = 2, \dots, k-1$ the nodes v_{i-1}, v_i, v_{i+1} are distinct,
(iii) every its undirected subsequence $v_j - v_{j+1} - \dots - v_{j+t}$, $1 \leq j \leq k$, $0 \leq t \leq k-j$ consists of distinct nodes.
The concept of trail is more general than the concept of path since a node can occur several times in a trail.

In contrast to d-separation in DAGs we will not define blocking for nodes of a trail, but for its *sections*, that is, maximal undirected subpaths. Evidently, every trail can be decomposed uniquely into sections. Moreover, sections of a trail can be classified according to types (or existence) of edges of the trail entering the section. Namely, just one of the following three possibilities can occur for the first terminal node v_j of a section $S : v_j, \ldots, v_{j+t}$, $1 \leq j \leq k$, $0 \leq t \leq k-j$. If $j > 1$ & $v_{j-1} \rightarrow v_j$, then v_j is a *head-terminal* node of S; if $j > 1$ & $v_{j-1} \leftarrow v_j$, then v_j is a *tail-terminal* node of S; if $j = 1$, then v_j is an *end-terminal* node of S. An analogous classification holds for the second terminal node v_{j+t}. Thus, according to the type of terminal nodes[3] one can classify sections of a trail into the following 6 classes. A section of a trail is called a *head-to-head* section iff it has two head-terminal nodes, or a *head-to-tail* section iff it has one head-terminal node and one tail-terminal node. Analogously are defined *head-to-end*, *tail-to-tail*, *tail-to-end* and *end-to-end* sections.

Let $G = (N, E)$ be a CG, $Z \subset N$ and S be a section of a trail in G. The definition of *blocking* of S by Z depends on the type of the section S:
- if S is a head-to-head section, then S is blocked by Z iff $ds(S) \cap Z = \emptyset$[4],
- if S is a head-to-tail (respectively head-to-end) section, then S is blocked by Z iff $S \cap Z \neq \emptyset$ & every slide to the tail-terminal (respectively end-terminal) node of S meets Z,
- if S is a tail-to-tail or tail-to-end or end-to-end section, then S is blocked by Z iff $S \cap Z \neq \emptyset$ & every slide to any of the terminal nodes of S meets Z.

A trail in a CG is *c-separated* (chain separated) by Z iff there exists a section of the trail which is blocked by Z.

Let $G = (N, E)$ be a CG and $\langle X, Y|Z \rangle \in T(N)$. We will say that $\langle X, Y|Z \rangle$ is *represented* in G according to the *separation criterion* and write $\langle X, Y|Z \rangle_G^{sep}$ iff every trail from X to Y in G is c-separated by Z. We left to the reader to verify that c-separation specified to the case of DAGs gives exactly d-separation from [11]. Note that in case of c-separation we have to consider trails, the requirement of blocking paths only is indeed weaker.

Lemma 1. *Let $G = (N, E)$ be a CG, $\langle X, Y|Z \rangle \in T(N)$. Then $\langle X, Y|Z \rangle_G^{sep}$ iff $\langle X, Y|Z \rangle_G^{mor}$.*

The proof of this lemma is beyond the scope of a conference contribution and can be found in [1]. To prove the lemma the concept of moral graph is formally modified: edges of the original graph keep their type (that is, links or arcs) and the added edges are consider of a third type, say, virtual edges called *virts*. We can extend the concept of blocking for head-to-virt, tail-to-virt, virt-to-virt and end-to-virt sections. Then we show that for every $\langle X, Y|Z \rangle \in T(N)$, there exists a path from X to Y outside Z in the moral graph of $G_{an(XYZ)}$ iff there exists a trail from X to Y in G which is not blocked by Z. Both implications can be

[3] If $t = 0$, then the terminal nodes v_j and v_{j+t} coincide. In this case the node $v_j = v_{j+t}$ is considered as a double terminal node, that is, it can be for example both head- and tail-terminal, or for example twice head-terminal node and so on.

[4] It follows from Fact 1 that $ds(S) = ds(u)$ for any $u \in S$.

verified by consecutive transformation of the considered trail (respectively path) – by replacing sections meeting Z (respectively virts) by a 'detour'.

6 Input list

Let $G = (N, E)$ be a CG and $\mathcal{B} : B_1, \ldots, B_n$ a chain for G. The *domain* of a node u, written $dom^{\mathcal{B}}(u)$, is the union of blocks $B_1, \ldots, B_k$, where B_k is the block containing u. The set *adjacents* of u, written by $ad_G(u)$, is $\{v \in N;\ v - u \text{ in } G\}$, the *neighborhood* of u, written $nb_G(u)$ is $\{v \in N;\ v \rightarrow u \text{ or } v - u \text{ in } G\}$. Note that for every chain $\mathcal{B}$ for G and $u \in N$ it holds $nb_G(u) \subset dom^{\mathcal{B}}(u)$.

The *input list* associated with G and a chain $\mathcal{B}$ for G is the set of triplets:

$$L_G^{\mathcal{B}} = \{\langle\, u, dom^{\mathcal{B}}(u) - nb_G(u) - u | nb_G(u)\,\rangle;\ u \in N\}.$$

Note, that it generalizes the concept of causal input list for a DAG. Input lists have the following properties.

Lemma 2. *Every triplet from the input list is represented in G according to the moralization criterion.*

Proof. Consider the triplet corresponding to $u \in N$. The corresponding ancestral set is $dom(u)$, and moreover $ad_H(u) = nb_G(u)$, where H is the corresponding moral graph. Hence, $nb_G(u)$ separates u from the rest of $dom(u)$ in H.

Lemma 3. *The independency model given by the moralization criterion is a graphoid.*

A proof can be found in [1]. The lemma can be shown by checking for each graphoid axiom that if the moralization criterion holds for the triplets on the left-hand side of the axiom, then it implies that the moralization criterion holds for the triplet on the right-hand side.

Lemma 4. *Let $G = (N, E)$ be a CG, $\mathcal{B}$ a chain for G, $\langle X, Y|Z\rangle \in T(N)$. Then $\langle X, Y|Z\rangle_G^{mor}$ implies $\langle X, Y|Z\rangle \in gr(L_G^{\mathcal{B}})$.*

A proof can be found in [1]. The lemma states that every triplet for which the moralization criterion holds in a CG is in the graphoid closure of the input list of the CG.

We can summarize Lemmas 1, 2, 3 and 4 as follows:

Theorem 5. *Supposing $G = (N, E)$ be a CG and $\mathcal{B}$ be a chain for G, the following conditions are equivalent for a triplet t from $T(N)$:*
(i) t is represented in G according to the moralization criterion,
(ii) t is represented in G according to the separation criterion,
(iii) t belongs to the graphoid closure of the input list associated with G and $\mathcal{B}$.

It follows from the theorem that the graphoid closure of the input list does not depend on the choice of the chain.

7 Minimal I-map

In this section we generalize the construction of a minimal I-map (see [11]) to the case of CGs. Let $M \subset T(N)$ be a graphoid and $\mathcal{B} : B_1, \ldots, B_n$, $n \geq 1$ an ordered partition of N (into nonempty sets). Then for every $u \in N$ there exists the least set $X \subset dom^{\mathcal{B}}(u) - u$ for which $\langle u, dom^{\mathcal{B}}(u) - X - u | X \rangle \in M$.[5] Its existence and uniqueness follows from the assumption that M is a graphoid. Let us denote it by X_u. Our aim is to establish a CG with such a prescribed input list.

Lemma 6. *There exists a CG G having the given ordered partition $\mathcal{B}$ as its chain and the list $\{ \langle u, dom^{\mathcal{B}}(u) - X_u - u | X_u \rangle \,;\, u \in N \}$ as its input list $L_G^{\mathcal{B}}$. This CG is moreover a minimal I-map of M.*

A proof can be found in [1].

8 Conclusions

In this paper we have introduced a causal input list for chain graphs whose graphoid closure is shown in [1] to be exactly the set of triplets for which the moralization criterion holds. This implies that chain graphs are indeed a generalization of both DAGs and UGs as formalisms for representing independency relations. So, the concept of *chain graph* (CG) makes it possible to describe a wider class of independency models involving both UG-models and DAG-models. This raises expressiveness of graphical models. The presented results give certain unifying point of view on graphical models and establishes semantics for CGs.

Further, we have presented a separation criterion which is shown in [1] to be equivalent with the moralization criterion. The new separation criterion, based on the concept *c-separation* has its own significance. For example, it easily implies that every CG-model satisfies *composition* property [11][6] which may be complicated to verify using the moralization criterion. Nevertheless, its main profit is expected in future. In [4] it is shown that for every DAG there exists a probability distribution in which exactly those conditional independency statements hold that are represented in the graph. We hope that analogously to this result the concept of c-separation will help to prove a similar result for CGs. In fact, in [9] this is claimed to be an open question, and in [2] even a wish to have a proper separation criterion for this purpose is expressed. Such a result would justify completely the use of CG in probabilistic reasoning. We expect analogous results also in other calculi for dealing with uncertainty in artificial intelligence.

---

[5] We keep the notation $dom^{\mathcal{B}}(u)$ from the preceding section, by convention $\langle u, \emptyset | dom^{\mathcal{B}}(u) - u \rangle \in M$.

[6] The composition property: $[\langle X, Y|Z \rangle \;\&\; \langle X, W|Z \rangle] \;\rightarrow\; \langle X, YW|Z \rangle$.

References

1. R. Bouckaert and M. Studený: Chain graphs: semantics and expressiveness - extended version. Res. rep. Institute of Information Theory and Automation, Prague 1995 (in preparation).
2. M. Frydenberg: The chain graph Markov property. Scand. J. Statist. 17 (1990), 333-353.
3. D. Geiger, T. Verma, J. Pearl: Identifying independence in Bayesian networks. Networks 20 (1990), 507-534.
4. D. Geiger, J. Pearl: On the logic of causal models. In Uncertainty in Artificial Intelligence 4 (R. D. Shachter, T. S. Lewitt, L. N. Kanal, J. F. Lemmer eds.), North-Holland 1990, 3-14.
5. D. Geiger, J. Pearl: Logical and algorithmic properties of conditional independence and graphical models. Ann. Statist. 21 (1993), 2001-2021.
6. H. Kiiveri, T. P. Speed, J. B. Carlin: Recursive causal models. J. Aust. Math. Soc. A 36 (1984), 30-52.
7. S. L. Lauritzen, N. Wermuth: Mixed interaction models. Res. rep. R-84-8, Inst. Elec. Sys., University of Aalborg 1984 (the report was later changed and published as an journal paper here referenced as [8]).
8. S. L. Lauritzen, N. Wermuth: Graphical models for associations between variables, some of which are qualitative and some quantitative. Ann. Statist. 17 (1989), 31-57.
9. S. L. Lauritzen: Mixed graphical association models. Scand. J. Statist. 16 (1989), 273-306.
10. S. L. Lauritzen, A. P. Dawid, B. N. Larsen, H.-G. Leimer: Independence properties of directed Markov fields. Networks 20 (1990), 491-505.
11. J. Pearl: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA 1988.
12. M. Studený: Formal properties of conditional independence in different calculi if AI. In Symbolic and Quantitative Approaches to Reasoning and Uncetainty (M. Clarke, R. Kruse, S. Moral eds.), Springer-Verlag, Berlin Heidelberg 1993, 341-348.
13. T. Verma, J. Pearl: Causal netwoks: semantics and expressiveness. In Uncertainty in Artificial Intelligence 4 (R. D. Shachter, T. S. Lewitt, L. N. Kanal, J. F. Lemmer eds.), North-Holland 1990, 69-76.

This article was processed using the LaTeX macro package with LLNCS style