COMPARISON OF GRAPHICAL APPROACHES TO DESCRIPTION OF CONDITIONAL INDEPENDENCE STRUCTURES

M. Studený

Two traditional classes of graphs used to describe probabilistic conditional independence structures are undirected graphs and directed acyclic graphs. In recent years several wider classes of graphs have been used in literature for this purpose: general directed graphs, chain graphs, reciprocal graphs and annotated graphs. The aim of this contribution is to give except an overview a rough comparison of achieved results. That means, several questions of general interest are raised and it is mentioned which questions are already answered.

INTRODUCTION

Graphs whose nodes correspond to random variables are traditional tools for description of structures of multidimensional probability distributions. One can distinguish two classic graphical approaches to description of conditional independence structure of a probability distribution: either using *undirected graphs* (we will use the abbreviation UG) or using *directed acyclic graphs* (we will use the abbreviation DAG). However, in recent years more general classes of graphs have been proposed for description of conditional independence structures (of probability distributions).

This paper tries to give a rough overview of these new graphical approaches. Each mentioned approach is recorded by an informal description of its characteristic features and by corresponding reference. Several relevant results concerning the respective approach are sometimes recalled or paraphrased (this is the case of classic approaches). The ambition of the paper is not a profound complete survey of all graphical approaches to description of probabilistic conditional independence structure. It is simply only an overview of those approaches which the author came across and its aim is to inform other participants of WUPES'97 about recent development in this area. I would like to make a preventive apology to all authors of referenced papers for possible misinterpretation.

The structure of the paper is the following one. In the first section we recall a few basic concepts. To give certain unifying point of view on all graphical approaches we try to give in the second section a list of general questions of theoretical significance which can be raised and studied for each mentioned graphical approach. Such

a (subjective) list makes it possible to classify in subsequent sections the results concerning respective graphical approaches.

Then, we recall the results concerning the classic approaches. The third section deals with UGs, the fourth section with DAGs. Similar attention is devoted to the class of *chain graphs* (we use the abbreviation CG) in the fifth section. The class of CGs involves both UGs and DAGs. Further three classes of graphs which can be (in certain sense) considered as subclasses of the class of CGs will be mentioned in the sixth section.

Remainig four sections describe graphical approaches which somehow go beyond the framework of CGs. Each section is devoted to another direction, that is to another way of generalization. The seventh section deals with *annotated graphs*, that is undirected graphs whose edges (in fact not only edges) are annotated by a subset of remaining nodes. The eighth section is devoted to graphs in which directed cycles are allowed. We will mention both general directed graphs and a very general class of *reciprocal graphs* which involves the class of CGs. Another direction to generalization is to consider possibility of a *hidden variable*, that is to consider a model obtained by a restriction of a graphical model to a subset of variables (see the ninth section). Last section deals with approaches which do not generalize graphs but ascribe the corresponding model of conditional independence structure in a different way. Shortly, these approaches give alternative interpretation to the graphs. These approaches involve so-called *covariance graphs* and generalized CGs.

1. BASIC CONCEPTS

Supposing N is a nonempty finite set of variables let us denote by $\mathcal{T}(N)$ the class of triplets $\langle X, Y|Z\rangle$ of disjoint subsets of N whose first two components X and Y are nonempty. These triplets will be used to describe particular conditional independency statements. A *independency model* over a N is a subset of $\mathcal{T}(N)$. The class of all independency models over N can be denoted by $\mathcal{I}(N)$.

If $\mathcal{M} \subset \mathcal{T}(N)$ is an independency model over N, then the *restriction* of $\mathcal{M}$ to a subset of variables $\emptyset \neq T \subset N$ is $\mathcal{M} \cap \mathcal{T}(T)$.

A (discrete) *probability distribution over* N is specified by a collection of nonempty finite sets $\{\mathbf{X}_i; i \in N\}$ (indexed by N) and by a function $P : \prod_{i\in N} \mathbf{X}_i \to [0,1]$ with $\sum\{P(\mathbf{x});\ \mathbf{x} \in \prod_{i\in N} \mathbf{X}_i\} = 1$. If $P(\mathbf{x}) > 0$ for all $\mathbf{x} \in \prod_{i\in N} \mathbf{X}_i$, then P is called *strictly positive*. Whenever $\emptyset \neq T \subset N$ and P is a probability distribution over N its *marginal distribution* for T is a probability distribution P^T (over T) defined as follows ($P^N \equiv P$) :

$P^T(\mathbf{x}) = \sum\{P(\mathbf{x},\mathbf{y});\ \mathbf{y} \in \prod_{i\in N\setminus T} \mathbf{X}_i\}$ for $\mathbf{x} \in \prod_{i\in T} \mathbf{X}_i$.

Having $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ and a probability distribution P over N we say that X is *conditionally independent of* Y *given* Z *with respect to* P and write $X \perp\!\!\!\perp Y \,|\, Z\ (P)$ if

$\forall\ \mathbf{x} \in \prod_{i\in X} \mathbf{X}_i \quad \mathbf{y} \in \prod_{i\in Y} \mathbf{X}_i \quad \mathbf{z} \in \prod_{i\in Z} \mathbf{X}_i$

$$P^{X\cup Y\cup Z}(\mathbf{x},\mathbf{y},\mathbf{z}) \cdot P^Z(\mathbf{z}) = P^{X\cup Z}(\mathbf{x},\mathbf{z}) \cdot P^{Y\cup Z}(\mathbf{y},\mathbf{z}) \ ,$$

where we accept the convention $P^\emptyset(-) \equiv 1$.

The independency model *induced by* a probability distribution P over N is the collection of all triplets $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ such that $X \perp\!\!\!\perp Y \mid Z\ (P)$.

(Discrete) *probabilistic independency model* over N is an independency model induced by a (discrete) probability distribution over N. The class of all such probabilistic independency models over N can be denoted by $\mathcal{P}(N)$. Evidently, restriction of a probabilistic independency model to a set $\emptyset \neq T \subset N$ is a probabilistic independency model over T here.

Note that in our overview also independency models induced by continous (more exactly nondegenerate Gaussian) distributions are mentioned although we decided to omit the corresponding definitions.

Several authors independently accentuated some basic properties of probabilistic independency models [9, 36]. Pearl and Paz [28] highlighted the importance of those properties in probabilistic reasoning and introduced the concept of *semigraphoid* as an independency model closed under the 'inference rules' below. The 'rules' should be understood as follows: if an independeny model contains the triplets before the long arrow, then it contains also the triplet after the long arrow.

$\langle A, B|C\rangle \longrightarrow \langle B, A|C\rangle$ symmetry
$\langle A, B \cup C|D\rangle \longrightarrow \langle A, C|D\rangle$ decomposition
$\langle A, B \cup C|D\rangle \longrightarrow \langle A, B|C \cup D\rangle$ weak union
$[\langle A, B|C \cup D\rangle \ \&\ \langle A, C|D\rangle] \longrightarrow \langle A, B \cup C|D\rangle$ contraction.

It is called a *graphoid* if it moreover satisfies:

$[\langle A, B|C \cup D\rangle \ \&\ \langle A, C|B \cup D\rangle] \longrightarrow \langle A, B \cup C|D\rangle$ intersection.

It is known [9] that every independency model induced by a strictly positive distribution is a graphoid. A *graphoid closure* of $\mathcal{M} \subset \mathcal{T}(N)$ consists of those triplets in $\mathcal{T}(N)$ which are derivable from $\mathcal{M}$ by consecutive application of graphoid inference rules. Evidently, it is a graphoid.

In general, a graph over a nonempty finite set of variables N will have the set N as the set of *nodes* and will be given usually by a set of 'edges'. What is meant by an 'edge' in a graph will be specified later separately for each considered class of graphs. In this section we introduce certain quite general class of graphs which involves the majority of treated classes of graphs (but not all of them). We distinguish here two basic types of edges over N: undirected edges, called lines and directed edges, called arrows. A *line* over N is an unordered couple $\{u, v\}$ where $u, v \in N$, $u \neq v$ (that is a two-element subset of N). An *arrow* over N is an ordered couple (u, v) where $u, v \in N$, $u \neq v$.

A *graph with mixed edges* over (a set of nodes N) is specified by a set of lines $\mathcal{L}$ over N and by a set of arrows $\mathcal{A}$ over N. Supposing $G = (N, \mathcal{L}, \mathcal{A})$ is such a graph in case $\{u, v\} \in \mathcal{L}$ we say that there exists a line between u and v in G and write '$u - v$ in G'. Similarly, in case $(u, v) \in \mathcal{A}$ we say that there exists an arrow from u to v in G and write '$u \to v$ in G' or '$v \leftarrow u$ in G'. If either $u - v$ in G, or $u \to v$ in G, or $u \leftarrow v$ in G, then we say that $[u, v]$ is an *edge* in G. Note explicitly that such a definition allows (for a couple of distinct nodes $u, v \in N$) that each of $u - v$, $u \to v$ and $u \leftarrow v$ are simultaneously edges in G! A *hybrid graph* over N is such a graph with mixed edges G that for a couple of distinct nodes $u, v \in N$ at most one of those possibilities occurs in G.

If $\emptyset \neq T \subset N$, then the *induced subgraph* of G for T is a graph $G_T = (T, \mathcal{L}_T, \mathcal{A}_T)$ over T where $\mathcal{L}_T$ (resp. $\mathcal{A}_T$) is the set of those lines (resp. arrows) over T which are also in $\mathcal{L}$ (resp. in $\mathcal{A}$).

A *route* from a node u to a node v (or between nodes u and v) in such a graph G is a sequence of nodes $w_1, \ldots, w_n \in N$, $n \geq 1$ together with a seguence of edges $\epsilon_1, \ldots, \epsilon_{n-1} \in \mathcal{L} \cup \mathcal{A}$ (possibly empty in case $n = 1$) such that $u = w_1$, $v = w_n$ and ϵ_i is either $w_i - w_{i+1}$, or $w_i \to w_{i+1}$, or $w_i \leftarrow w_{i+1}$ for $i = 1, \ldots, n-1$. A route is called *descending* if ϵ_i is either $w_i - w_{i+1}$, or $w_i \to w_{i+1}$ for $i = 1, \ldots, n-1$, and *undirected* if ϵ_i is $w_i - w_{i+1}$ for $i = 1, \ldots, n-1$. Especially, every undirected route is a descending route.

A *path* is a route in which all nodes $w_1, \ldots, w_n$ are distinct, a *cycle* is a route where $n \geq 2$, $w_1 = w_n$ and $w_1, \ldots, w_{n-1}$ are distinct. A *directed cycle* is a cycle which is a descending route and where ϵ_i is $w_i \to w_{i+1}$ at least once.

We say that a node u is a *parent* of a node v in G if $u \to v$ in G, and u is an *ancestor* of v in G if there exists a descending route (equivalently a descending path) from u to v in G, and u is *connected* to v in G if there exists an undirected route (equivalently an undirected path) between u and v. Supposing $A \subset N$ the symbol $an_G(A)$ denoted the set of ancestors of the nodes of A in G. Clearly, the relation 'be connected' is an equivalence relation which decomposes N into equivalence classes, named *connectivity components*.

An *undirected graph* (UG) is a graph containing only lines (that is $\mathcal{A} = \emptyset$), a *directed graph* is a graph containing only arrows (that is $\mathcal{L} = \emptyset$). The *underlying graph* H of a graph with mixed edges $G = (N, \mathcal{L}, \mathcal{A})$ is an undirected graph H over N such that $u - v$ in H iff $[u, v]$ is an edge in G.

However, one of the later mentioned approaches even allows two different types of lines (so-called 'solid' lines and 'dashed' lines) and analogously two different types of arrows. In another approach, a graph is not specified by a set of edges only but also by further mathematical entities (annotation).

2. QUESTIONS RELATED TO GRAPHICAL APPROACHES

In this section we formulate several questions of theoretical interest which arise in connection with graphical methods of description of probabilistic independency models. If one decides to use certain class of graphs $\mathcal{G}$ for description of those models, then one should clarify undoubtedly the following two basic issues.

A *Establishment of the class of graphs:*
How exactly the class of respective graphs $\mathcal{G}(N)$ over N is specified for each nonempty finite set of variables N?

B *Ascription of the independency model:*
How exactly an independency model over N is ascribed to every graph from $\mathcal{G}(N)$?

Thus, the first question is aswered if one gives the definition of the class $\mathcal{G}(N)$ of considered graphs over N, and the second question is answered if one defines (for

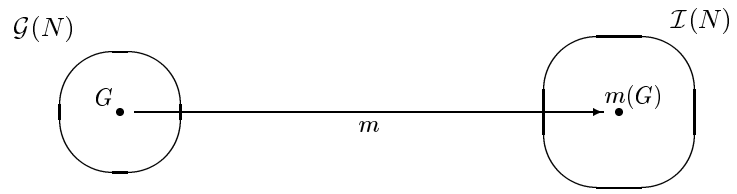

Figure 1: An independency model over N is ascribed to every graph over N.

each N) an interpretation mapping m from $\mathcal{G}(N)$ to the class $\mathcal{I}(N)$ of independency models over N (figure 1 gives an illustrative picture).

Remark Although we did not mention it explicitly (since it is almost impossible to formulate exactly) we have in mind a 'consistent' way of introducing of a graphical framework. That means introducing of $\mathcal{G}(N)$ and the definition of the mapping m for each N has a common intuitive 'source' behind which does not depend on N. Bluntly said, we are not interested in 'inconsistent' graphical frameworks where for example $\mathcal{G}(N)$ is in case $\mathrm{card}\, N = 3$ the class of undirected graphs over N but in case $\mathrm{card}\, N = 4$ it is the class of directed graphs over N. We hope that the reader understands our intuitive aim: it will become more clear from examples in subsequent sections. To be true, our 'intuitive' assumption has as a result many elegant properties which we consider as self-evident. For example, a permutation π on a set N induces naturally a permutation π_* of $\mathcal{G}(N)$ and a permutation π^* of $\mathcal{I}(N)$ and the interpretation mapping m should commute with these permutations, that one should have $m(\pi_*(G)) = \pi^*(m(G))$ for every $G \in \mathcal{G}(N)$.

As concerns the question (B) let us note that its formulation above 'simplifies' the real historical development. In fact, for both classic graphical approaches (i.e. UGs and DAGs) at first the class of distributions having the structure described by a graph was somehow introduced. For example, several structural conditions on distributions with respect to a graph were introduced (i.e. the Gibbs condition, the pairwise Markov condition, the local Markov condition, the global Markov condition etc.), their relation was studied (namely their equivalence for strictly positive distributions was shown) and thus, the class of Markovian distributions was assigned to every graph from the considered class of graphs. Then the researchers solved the question whether the assigned class of distributions can be equivalently characterized

as the class of distributions inducing certain independency model. So, the independency model $m(G)$ ascribed to a graph $G \in \mathcal{G}(N)$ has meaning of the independency structure described by the graph.

Instead of describing the conditional independence structure of a probability distribution P over N by means of the independency model $\mathcal{M} \in \mathcal{I}(N)$ induced by P we wish to describe the structure by a graph $G \in \mathcal{G}(N)$ such that $\mathcal{M} = m(G)$. Thus, the the first natural question is the following one.

C *Correctness:*
Is the model $m(G)$ ascribed to a graph G over N indeed a probabilistic independency model over N (for every N and every $G \in \mathcal{G}(N)$ of course)?

In mathematical language we ask whether $m(\mathcal{G}(N)) \subset \mathcal{P}(N)$ for every N. Note that some authors call positive results of this type the strong completeness results. By the completeness result they understand a weaker result saying that for every $G \in \mathcal{G}(N)$ and for every $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ outside $m(G)$ there exists a probability distribution P over N such that its induced independency model contains $m(G)$ but $\neg[\, X \perp\!\!\!\perp Y|Z\,(P)\,]$. As the reader will see later, this basic question is still open for some of mentioned new approaches. However, another relevant question was studied.

D *Characterization of ascribed independecy models:*
Is it possible to characterize the ascribed independency models in terms of properties of those independency models?

More exactly, we wish to characterize the range of the mapping m, that is $m(\mathcal{G}(N))$, without a reference to $\mathcal{G}(N)$ (for every N). The aim is to obtain certain type of 'axiomatic' characterization of ascribed independency models, that is to characterize them as independency models closed under a finite number of inference rules of semigraphoid type. A special task within the framework of this question, which was often a topic of research, is the task whether every ascribed independency model is a graphoid.

The interpretation mapping m induces on $\mathcal{G}(N)$ a natural equivalence relation: we say that two graphs G and H from $\mathcal{G}(N)$ are *equivalent* if $m(G) = m(H)$. The equivalence relation characterizes the situation when G and H describe the same independency model. This leads to a couple of questions.

E *Characterization of equivalent graphs:*
Is it possible to characterize the equivalence relation of graphs in graphical terms?

F *Representation of the class of equivalent graphs:*
Is there any (natural) representative of every class of equivalent graphs?

As concerns the previous question, an ideal representative of the equivalence class is surely a naturally distinguished graph from the equivalence class, but one can consider also represetation by more general mathematical objects (for example by more general graphs) which somehow encode features shared by all graphs within the equivalence class.

Further question of our interest concerns restriction of independency models.

G *Behaviour with respect to restriction:*
Could be the restriction of every independency model ascribed to a graph $G \in \mathcal{G}(N)$ to every subset of $\emptyset \neq T \subset N$ realized as the independency model ascribed to a graph from $\mathcal{G}(T)$?

In mathematical language we ask whether it holds $r_T(m(\mathcal{G}(N))) \subset m_T(\mathcal{G}(T))$ for every $\emptyset \neq T \subset N$ where r_T denotes the operation of restriction of independency models to T and m_T the interpretation mapping for the set T. The significance of a positive answer to this question is evident: restriction of graphical independency model to a set of variables can be performed directly in graphical terms.

The previous questions were mainly of theoretical interest. However, most of the papers in the field of probabilistic reasoning are concerned with the following, more specific question.

H *Estimation of graphs from data:*
Suppose that statistical data are 'generated from' a distribution over N inducing an independency model $m(G)$ for a $G \in \mathcal{G}(N)$. How to obtain (an estimate of) such a graph on basis of the data?

Note that we omit this important question in our overview below for limited scope of a conference paper. Nevertheless, we will pay little attention to the following task.

I *Comparison:*
Compare different classes of graphical independency models, either qualitatively or quantitatively.

As concerns qualitative comparison we can ask whether for every N the class of independency models ascribed to one class of graphs $\mathcal{G}_1(N)$ (by a mapping m_1) contains the class of independency models ascribed to another class of graphs $\mathcal{G}_2(N)$ (by a mapping m_2). By quantitative comparison we understand to compare the number of independency models ascribed to graphs from $\mathcal{G}_1(N)$ with the number of independency models ascribed to graphs from $\mathcal{G}_2(N)$.

3. UNDIRECTED GRAPHS

The class of UGs was already defined in the first section. Note that in [29] UGs were named also *Markov networks*.

A triplet $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ is *represented* in an UG G over N if every path in G between a node in X and a node in Y contains a node in Z (such a path is necessarily an undirected path). One can ascribe to G the independency model over N consisting of the triplets over N represented in G.

It was shown in [15] that every independency model ascribed to an UG is a probabilistic independency model induced by a strictly positive distribution. Especially, every UG-model, that is an independency model ascribed to an UG, is a graphoid. See also [11] for the completeness result with respect to the class of nondegenerate Gaussian distributions. Thus, the correctness of UG-models is ensured.

In [29, 28] one can find also an answer to the question (D), that is the independency models ascribed to UGs are characterized in terms of properties analogous to the semigraphoid properties. The properties are more complex, they allow disjunction of triplets as a consequent of an inference rule unlike one triplet in case of properties of semigraphoid type. It follows from that construction that two UGs over N are equivalent iff they are equal. Thus, the questions (E) and (F) are solved in an ideal way for UGs.

The answer to the question (G) is also positive in the framework of UGs. The reader can be surprised that the corresponding graph over a subset $T \subset N$ is not the induced subgraph for T in general. The corresponding *restricted* graph G^T is defined as follows: $u - v$ in G^T iff there exists a path in G between u and v consisting of nodes of $\{u, v\} \cup (N \setminus T)$. The proof of the corresponding claim can be found in [27].

4. DIRECTED ACYCLIC GRAPHS

A *directed acyclic graph* (DAG) over N is a directed graph over N without directed cycles. Note that a more precise name is 'acyclic directed graph' but unfortunately the most of authors became accustomed to the abbreviation DAG or use another names like *Bayesian network* [29] or influence diagram. A DAG can be equivalently introduced as a directed graph G whose (all) nodes can be ordered in a sequence $u_1, \ldots, u_k$, $k \geq 1$ such that if $[u_i, u_j]$ is an edge in G for $i < j$, then $u_i \to u_j$ in G.

There are two equivalent criteria to decide whether a triplet $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ is represented in an DAG G over N. Lauritzen *et. al.* [25] used so-called *moralization criterion* while the group around J. Pearl [13] used a direct *d-separation criterion* in which one tests whether paths in G from X to Y are 'blocked' by Z (the definition of blocking is special, it depends on directions of arrows of the path).

Let us describe the moralization criterion here. It has several stages. At first, one takes the set $T = an_G(X \cup Y \cup Z)$ and considers the induced subgraph G_T. Then G_T is changed into its *moral graph* H, that is the underlying graph of the graph K (with mixed egdes) over T which is obtained from the graph G_T by adding lines $u - v$ in K whenever there exists $w \in T$ having both u and v as parents in G_T. The name 'moral graph' was motivated by the fact that the parents of every node are 'married'. The last step is to decide whether $\langle X, Y|Z\rangle$ is represented in the moral graph H over T - this determines whether $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ is represented in the DAG G. The ascribed independency model over N then consists of triplets (over N) represented in G.

Also in the framework of DAGs the answer to the question (C) is positive. In [14] a construction of a desired probability distribution is given (which need not be strictly positive).

As far as I know, the DAG-models were not so far satisfactory characterized in terms of properties of independency models. In [29] several properties of models ascribed to DAGs are mentioned which imply that DAG-models are always graphoids. However, these properties of semigraphoid type do not characterize DAG-models. In [12] it is proved that DAG-models cannot be characterized by means of a finite

number of properties of semigraphoid type. On the other hand it is mentioned in the paper [12] that Verma in a prepared (unfinished ?) research report found very complex characterization of DAG-model which cannot be considered as an 'axiomatic' characterization. Note that a special class of singly connected DAG-models was characterized in terms of properties analogous to the semigraphoid properties in [5]. So, the question whether there exists certain reasonable 'axiomatic' characterization of DAG-models seems to remain open.

The answer to the question (E) was given at first in [40], but this result can be found also in other publications, for example in [1, 34]. Let's call *immorality* in an DAG G over N every induced subgraph of G for a set $T = \{u, v, w\}$ such that $u \to w$ in G, $v \to w$ in G and $[u, v]$ is not an edge in G. It was shown that two DAGs over N are equivalent iff they have the same underlying graph and the same occurence of immoralities.

On the other hand the question of representation of equivalence classes of DAGs has not so elegant solution. There is no natural representative of a class of equivalent DAGs within the class. Thus, hybrid graphs were used in literature [40, 1] to represent uniquely the equivalence classes of DAGs. Let's introduce the *essential graph* of a class of equivalent DAGs over N is such a hybrid graph over N which has the same underlying graph as every DAG from the considered equivalence class and whose only arrows are those arrows which occur in every DAG from the equivalence class (with the same orientation). The remaining edges of the essential graph are lines. The essential graphs were characterized in graphical terms in [1].

The answer to the question (G) is negative in case of DAGs. This led to an effort to study the models which are restrictions of DAG-models, see the ninth section.

Note that the problem of estimation of DAGs from data, more exactly estimation of the essential graph on basis of the induced independency model (which can be obtained as a result of statistical test based on data) was treated in [41] and [26]. A related question of choice of a suitable essential graph on basis of statistical data was treated in [7].

The classes of UG-models and DAG-models are incomparable from qualitative point of view. Examples can be found in [39]. As concerns quantitative comparison, the number of DAG-models is higher than the number of UG-models for every N with at least 3 nodes.

5. CHAIN GRAPHS

A *chain graph* (CG) over N is a hybrid graph over N without directed cycles. The class of chain graphs was introduced by Lauritzen and Wermuth in middle eighties [21]. Let us recall here the original equivalent definition which explains the terminology. A *chain* for a hybrid graph G is a partition of N into ordered disjoint (nonempty) subsets $B_1, \ldots, B_n$, $n \geq 1$ called *blocks* such that, if $[u, v]$ is an edge in G with $u, v \in B_i$ then $u - v$, and if $[u, v]$ is an edge in G with $u \in B_i, v \in B_j, i < j$ then $u \to v$. A CG is then a hybrid graph which admits a chain.

Lauritzen [23], followed by Frydenberg [10], introduced the *moralization criterion for CGs*. The only difference from the moralization criterion described in the preceding section is in a more general definition of the moral graph. Supposing G_T is a hybrid graph over $\emptyset \neq T \subset N$ one defines a graph K with mixed edges over T by adding lines $u - v$ in K whenever there exist $w, t \in T$ belonging to the same connectivity component of G_T (possibly $w = t$) such that $u \to w$ in G_T and $v \to t$ in G_T. The *moral graph* H of G_T is the underlying graph of K. Well, the independency model over N ascribed to a CG G consists of triplets $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ represented in G according to the moralization criterion for CGs.

An equivalent *c-separation criterion* which generalizes the d-separation criterion for DAGs was introduced in [4]. In case of the separation criterion the difference from the case of DAGs is more visible: if one verifies using c-separation criterion whether $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ is represented in a CG G over N, then one has to test a more wider class of routes in a CG G from X to Y whether they are 'blocked' by Z (it is not sufficient to test only paths between X and Y, but on the other hand one need not to test all routes).

The c-separation criterion made it possible to prove in [37] that every independency model ascribed to a CG is a probabilistic independency model induced by a strictly positive probability distribution. Thus, the correctness of CG-models is ensured and every CG-model is a graphoid.

As far as I know, the question (D) for case of CGs was not studied so far.

However, equivalent CGs were characterized in graphical terms by Frydenberg [10]. Let's call a *complex* in a CG G over N every induced subgraph of G for a set $T = \{w_1, \ldots, w_k\}$, $k \geq 3$ such that $w_1 \to w_2$, $w_i - w_{i+1}$ for $i = 2, \ldots, k-2$, $w_{k-1} \leftarrow w_k$ in G, and no additional edge between (distinct) nodes of $\{w_1, \ldots, w_k\}$ exist in G. It was shown that two CGs over N are equivalent iff they have the same underlying graph and the same occurence of complexes. Evidently, the concept of a complex generalizes the concept of an immorality from the preceding section (take $k = 3$) and immoralities are the only complexes in a DAG. Thus, the characterization of equivalent CGs generalizes the above mentioned characterization of equivalent DAGs.

Moreover, unlike the case of DAGs, an elegant answer to the question (F) for CGs was found: every class of equivalent CGs contains a naturally distinguished CG! We say for two CGs G and H over N having the same underlying graph that G is *larger* than H if every arrow in G is an arrow in H with the same orientation. Frydenberg [10] showed that every class of equivalent CGs contains a CG which is larger than every other CG within the class (i.e. it has the greatest number of lines). This distinguished CG is named the *largest chain graph* of that equivalence class.

Certain graphical characterization of largest CGs is given in the paper [38] where also the problem of estimation of the largest CG on basis of the induced independency model is treated. However, perhaps a more elegant graphical characterization of largest CGs was lately found by M. Volf (my diploma student).

Although I did not checked it carefully, I think that the answer to the question (G) in the framework of CGs is also negative: one can use the same example as in case of DAGs.

As concerns qualitative comparison, it is evident from the definitions above that the class od CG-models involves both the class of UG-models and the class of DAG-models. Thus, CGs give certain unifying point of view on Markov and Bayesian networks. Note for information that in case $\mathrm{card}\, N = 3$ one has 8 UG-models and 11 DAG-models (= CG-models), in case $\mathrm{card}\, N = 4$ there exist 64 UG-models, 185 DAG-models and 200 CG-models, and in case $\mathrm{card}\, N = 5$ one has 1024 UG-models, 8782 DAG-models (see [1]) and 11519 CG-models (a calculation of M. Volf).

6. MODELS IN THE FRAMEWORK OF CHAIN GRAPHS

There are three interesting classes of graphs used for description of conditional independence structures which can be somehow considered as subclasses of the class of CGs.

Quite important class is the class of chordal (= triangulated) graphs. An UG G is called *chordal* if every cycle $u_1, \ldots, u_n$, $n \geq 5$ in G has a 'chord', that is a line between nodes of $\{u_1, \ldots, u_{n-1}\}$ different from the lines of the cycle. Note that chordal graphs induce so-called *decomposable models* [29] which have very pleasant properties from the computational point of view and form mathematical basis of the well-known method of local computation [24].

What is worthwhile to mention is that the class of independency models which are simutaneously UG-models and DAG-models is exactly the class independency models ascribed to chordal graphs. An 'axiomatic' characterization of those independency models in terms of properties analogous to the semigraphoid properties was found in [6]. The reader may be interested in the fact that the answer to the question (G) is positive for chordal graphs. One can easily verify that the restriction G^T of a chordal graph G is again a chordal graph.

In the paper [18] the class of recursive causal graphs is studied. A *recursive causal graph* G over N can be equivalently defined as a CG which admits a chain such that the only lines of G are lines within the first block. Thus, UGs and DAGs are special cases of recursive graphs. Note that the way of ascribing the independency model to recursive graphs is consonant with the way used in case of CGs. Thus, also from the point of view of description of conditional independence structure CGs strictly generalize recursive graphs: one can find a CG whose CG-model is not ascribed to a recursive graph.

Shafer in his recent book [33] deals also with bubble graphs. A *bubble graph* over N is specified by an ordered decomposition $B_1, \ldots, B_n$, $n \geq 1$ of N into nonempty subsets, named *bubbles*, and by a collection of 'arrows' which point to bubbles although they originate from single nodes taken from the preceding bubbles. Every such a graph describes the class of probability distributions over N which satisfy certain factorization formula.

A bubble graph B over N is not a graph in standard sense, but one can associate with it a CG G over N made by joining nodes in each bubble of B by lines and by replacing any 'arrow' from a node $u \in N$ to a bubble $B \subset N$ by the collection of ordinary arrows from u to every node of B. Then one can derive easily using

the result about the factorization formula associated with a CG from [10] that a probability distribution over N satisfies the factorization formula corresponding to the bubble graph B iff it induces the independency model ascribed to G. Thus, bubble graphs can be naturally embedded into the class of CGs.

7. ANNOTATED GRAPHS

Paz and Geva introduced a few years ago so-called annotated graphs which can describe very wide class of independency models. Let's call an *element* over N any couple $[\,\{u,v\}\,|\,S\,]$ where $u,v \in N$, $u \neq v$ and $S \subset N$ is a set with $\{u,v\} \cap S = \emptyset$. An *annotated graph* over N is a UG $H = (N, \mathcal{L})$ supplemented by a collection $\mathcal{E}$ of elements over N. Intended interpretation of an element $[\,\{u,v\}\,|\,S\,] \in \mathcal{E}$ is that a (possible) line u — v in H is 'annotated' by a set of nodes S. However the definition above is too wide. The class of *regular* annotated graphs is defined in [27] as the class of those annotated graphs which satisfy three regularity conditions. We will not repeat these technical conditions here.

In the mentioned paper a so-called 'membership algorithm' is introduced. This algorithm decides whether a triplet $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ is represented in a regular annotated graph over N. Loosely said, the 'membership algorithm' consists in successive removal of the elements from $\mathcal{E}$, by (corresponding) restriction of H and $\mathcal{E}$ and by removal of certain edges of H. After those changes (when all elements from $\mathcal{E}$ are removed) one tests whether $\langle X, Y|Z\rangle$ is represented in the resulting UG. In certain sense the 'membership algorithm' is analogous to the moralization criterion.

Thus, one can ascribe an independency model over N to every (regular) annotated graph over N. It is shown in [27] that the independency model ascribed to a regular annotated graph is always a graphoid.

The aim of introducing of annotated graphs was to have a condensed 'graphical record' for description of graphoid closures of unions of UG-models. It is shown in [27] that for every sequence of UGs $G_i = (N_i, \mathcal{L}_i)$, $i = 1, \ldots, k$, $k \geq 1$ such that $N_i \subset N_{i+1}$ and $\mathcal{L}_i \subset \mathcal{L}_{i+1}$ for $i = 1, \ldots, k-1$ one can construct a regular annotated graph over $N = N_k$ such that the independency model over N ascribed to that annotated graph is exactly the graphoid closure of $\bigcup\{m(G_i); i = 1, \ldots, k\}$ where $m(G_i) \subset \mathcal{T}(N_i) \subset \mathcal{T}(N)$ denotes the independency model ascribed to the UG G_i for every $i = 1, \ldots, k$. As every CG-model can be obtained as such a graphoid closure the annotated graphs generalize CGs from the point of view of description of independency models.

On the other hand the question of correctness of annotated graphs is so far unsolved. Similarly, the other questions mentioned in the second section are open.

Note for information that the idea to annnotate edges of a graph by sets of remaining nodes appeared also in the paper [3], this time for DAGs and by a couple of disjoint subsets of the set of remaining modes. The class of IDAGs (that is 'annotated' DAGs) introduced there allows to describe every graphoid. Therefore, the answer to the question (C) for IDAGs is negative. Moreover, the question whether every independency model ascribed to an IDAG was not mentioned in [3].

8. GRAPHS ALLOWING DIRECTED CYCLES

A natural way of generalization of DAGs and CGs is to allow directed cycles. We have mentioned two such approaches in literature.

Spirtes *et. al.* in their book [34] alredy mentioned possible use of general directed graphs for desription of models allowing also feedback. Note that they used partially misleading terminology 'directed cyclic graphs', although the considered class of graphs involves DAGs.

In fact, one can ascribe an independency model over N to every directed graph over N using the moralization criterion described in the fourth section for DAGs. It was mentioned in [35] that one can also generalize directly d-separation criterion and that both criteria are equivalent also in case of general directed graphs. Moreover, Spirtes [35] showed that the completeness result with respect to the class of non-degenerate Gaussian distributions. Thus, (Gaussian) correctness of those models is almost ensured.

Attention was devoted also to the questions (E) and (F) from the second section. Richardson [31] gave a graphical characterization of equivalent (general) directed graphs. However, the mentioned characterization is quite complicated (unlike the case of DAGs): it involves 6 special independent conditions. In another paper Richardson [32] proposed to use special representatives of classes of equivalent directed graphs, named 'PAGs'. These representatives are much more complicated mathematical objects, namely graphs over the N whose edges have 3 possible endings for both end-nodes and where the endings of different edges near a common end-node may be connected by two possible 'connections'. Each 'mark' in a PAG express certain graphical property shared by all graphs within the equivalence class. On the other hand it is claimed that PAGs allow also to describe restrictions of (general) directed graph models (Richardson's oral communication). Note for explanation that the answer to the question (G) is negative also in case of general directed graphs.

Koster [19] introduced lately a very general class of reciprocal graphs. A *reciprocal graph* G over N is a graph with mixed edges over N such that there is no arrow in G between two nodes belonging to the same connectivity component of G. Thus, every CG is a reciprocal graph and every (general) directed graph is a reciprocal graph. Koster ascribed an independency model over N to every reciprocal graph by means of the moralization criterion described in the fifth section for the case of CGs. Note that in case of directed graphs it collapses to the moralization criterion treated by Spirtes *et. al.*

Thus, the independency models ascribed to reciprocal graphs involve both CG-models and the models ascribed to (general) cyclic graphs. On the other hand, the question of correctness of reciprocal graphs is solved only partially so far. Koster established connection of so-called 'simultaneous equation systems' (LISREL models) [20] with certain subclass of the class of reciprocal graphs (not involving CGs). Furthemore, for this special subclass a completeness result was achieved with respect to the class of Gaussian distributions.

However, as far as I know, the other questions of our interest were not solved in the framework of reciprocal graphs.

9. MODELS WITH HIDDEN VARIABLES

It was mentioned in the fourth section that the restriction of a DAG-model may not be a DAG-model. This led to an idea to describe the restrictions of DAG-models by means of graphical diagrams. Geiger, Paz and Pearl [16] developed the ideas from [30] and introduced so-called 'embedded Bayesian networks' for this purpose. An *embedded Bayesian network* over N is graph over N allowing both directed and bidirected edges (at most one edge is allowed between two different nodes) such that purely directed cycles (i.e. directed cycles made of arrows only, without bidirected edges) are not present in the graph.

In [16] an independency model over N is ascribed to every such a special graph G over N (using a generalized d-separation criterion) and it is mentioned in text that one can always find a DAG H over a superset $M \supset N$ such that the restriction of the independency model $m(H)$ over M (i.e. of the model ascribed to the DAG H) to N is exactly the independency model over N ascribed to G. Especially, owing to correctness of DAG-models and the fact the restriction of a probabilistic model is a probabilistic model, the independency model over N ascribed to G is a probabilistic independency model. So, the answer to the question (C) is positive for embedded Bayesian networks.

Moreover, according to Pearl's oral communication, Verma showed that every restriction of a DAG-model can be ascribed in such a way to certian embedded Bayesian network - probably in [30].

Note that in recent literature (not referenced here) the restrictions of DAG-models are often named the *models with hidded variables*. The reason is evident: one can imagine that N contains 'observed' variables but the actual DAG-model is over its superset $M \supset N$ and therefore $M \setminus N$ is the set of 'unobserved' or hidded variables.

10. MODELS WITH ALTERNATIVE INTERPRETATION

Some authors interpret the above mentioned graphs in another way, that is they define the interpretation mapping m mentioned in the second section in another way. These approaches were developed especially by Wermuth and were intended mainly for description of conditional independence structures induced by nondegenerate Gaussian distributions. To distinguish the 'classic' and the alternative interpretation Cox and Wermuth use in their book [8] in pictures dashed lines respectively dashes arrows instead of 'classic' solid lines and solid arrows if they have in mind an alternative way of interpretation.

Cox and Wermuth [8] consider a class of generalized CGs. A *joint-response chain graph* G is a CG in which every arrow is either a *solid arrow* or a *dashed arrow* and every line is either a *solid line* or a *dashed line* and where for every connectivity component C the following two conditions hold:

- all lines within C are either solid or dashed,
- all arrows directed to nodes of C are either solid or dashed.

Nevertheless, Cox and Wermuth do not ascribe directly the independency models to their generalized CGs. They rather describe in words (in one special chapter) what they understand by the class of (non-degenerate Gaussian) distributions which are Markovian with respect to a joint-response CG. So, the question whether the assigned class of distributions can be characterized as the class of distribution inducing certain independency model, that is the question (B), is open in general.

However, the mentioned question was anwered in a few special cases. Kauermann [17] answered it for the subclass of *covariance graphs* that is UGs made of dashed lines. The corresponding independency model which should be ascribed to such a graph G over N consists of those triplets $\langle X, Y|Z\rangle \in \mathcal{T}(N)$ such that every path in G from a node in X to a node in Y contains a node in $N \setminus Z$. He also showed the completeness result with respect to the class of nondegenerate Gaussian distributions.

A similar attitude occurs in the paper [2] where another independency model is assigned to every CG. Andersson, Madigan and Perlman introduced a criterion to decide whether a triplet over N is represented in a CG. This criterion is analogous to the moralization criterion described in the fifth section but it is different from it. Instead of the moral graph they define so-called 'augmented graph', otherwise everything is analogous. The alternative CGs from [2] then probably correspond to the generalized CGs with solid lines and dashed lines from [8] (according to Madigan's oral communication). Moreover (also according to an oral communication), one can introduce an equivalent 'separation' criterion which allows to verify that the alternatively ascribed independency model is indeed a probabilistic independency model.

ACKNOWLEDGEMENT

The paper was partially supported by the grant MŠMT n. VS96008.

REFERENCES

[1] S. A. Andersson, D. Madigan and M. D. Perlman: A characterization of Markov equivalence classes for acyclic digraphs. Tech. rep. 287, University of Washington, Seattle 1995, submitted to Ann. Statist.

[2] S. A. Andersson, D. Madigan and M. D. Perlman: An alternative Markov property for chain graphs. In Uncertainty in Artificial Intelligence 12 (E. Horvitz and F. Jensen eds.), Morgan Kaufmann 1996, 40–48.

[3] R. R. Bouckaert: IDAGs: a perfect map for any distribution. In Symbolic and Quantitative Approaches to Reasoning and Uncertainty (M. Clarke, R. Kruse and S. Moral eds.), Lecture Notes in Computer Science 747, Springer-Verlag 1993, 49–56.

[4] R. R. Bouckaert and M. Studený: Chain graphs: semantics and expressiveness. In Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Ch. Froidevaux and J. Kohlas eds.), Lecture Notes in Artificial Intelligence 946, Springer-Verlag 1995, 67–76.

[5] L. M. de Campos: Independency relationships in singly connected networks. DESCAI tech. rep. n. 960204, University of Granada 1996. Submitted to Artif. Intelligence.
[6] L. M. de Campos: Characterization of decomposable dependency models. DESCAI tech. rep. n. 960225, University of Granada, June 1996.
[7] D. M. Chickering: A transformational characterization of equivalent Bayesian network structures. In Uncertainty in Artificial Intelligence 11 (P. Besnard and S. Hanks eds.), Morgan Kaufmann, 1995, 87–98.
[8] D. R. Cox and N. Wermuth: Multivariate Dependencies - Models, Analysis and Interpretation. Chapman & Hall, 1996.
[9] A. P. Dawid: Conditional independence in statistical theory. J. Roy. Statist. Soc. ser. B, *41* (1979), 1–31.
[10] M. Frydenberg: The chain graph Markov property. Scand. J. Statist., *17* (1990), 333–353.
[11] M. Frydenberg: Marginalization and collapsibility in graphical interaction models. Ann. Statist., *18* (1990), 790–805.
[12] D. Geiger: Towards the formalization of informational dependencies. Tech. report CSD-850053, R-102, UCLA Los Angeles, December 1987.
[13] D. Geiger, T. Verma and J. Pearl: Identifying independence in Bayesian networks. Networks, *20* (1990), 507–534.
[14] D. Geiger and J. Pearl: On the logic of causal models. In Uncertainty in Artificial Intelligence 4 (R. D. Shachter, T. S. Lewitt, L. N. Kanal and J. F. Lemmer eds.), North-Holland 1990, 3–14.
[15] D. Geiger and J. Pearl: Logical and algorithmic properties of conditional independence and graphical models. Ann. Statist., *21* (1993), 2001–2021.
[16] D. Geiger, A. Paz and J. Pearl: On testing whether an embedded Bayesian network represents a probability model. In Uncertainty in Artificial Intelligence 10 (R. L. de Mantaras and D. Poole eds.), Morgan Kaufmann 1994, 244–252.
[17] G. Kauermann: On a dualization of graphical Gaussian models. Scand. J. Statist., *23* (1996), 105–116.
[18] H. Kiiveri, T. P. Speed and J. B. Carlin: Recursive causal models. J. Aust. Math. Soc. ser. A, *36* (1984), 30–52.
[19] J. T. A. Koster: Markov properties of non-recursive causal models. To appear in Ann. Statist.
[20] K. G. Jöreskog and D. Sörbom: LISREL 7. A Guide to the Program and Application. SPSS Inc., 1989.
[21] S. L. Lauritzen and N. Wermuth: Mixed interaction models. Res. rep. R-84-8, Inst. Elec. Sys., University of Aalborg 1984 (the research report was later modified and became a basis of the journal paper [22]).
[22] S. L. Lauritzen and N. Wermuth: Graphical models for associations between variables, some of which are qualitative and some quantitative. Ann. Statist., *17* (1989), 31–57.
[23] S. L. Lauritzen: Mixed graphical association models. Scand. J. Statist., *16* (1989), 273–306.
[24] S. L. Lauritzen and D. Spiegelhalter: Local computation with probabilities on graphical structures and their application to expert systems. J. Roy. Stat. Soc. ser. B, *50* (1988), 157–224.
[25] S. L. Lauritzen, A. P. Dawid, B. N. Larsen and H.-G. Leimer: Independence properties of directed Markov fields. Networks, *20* (1990), 491–505.
[26] Ch. Meek: Causal inference and causal explanation with background knowlewdge. In Uncertainty in Artificial Intelligence 11 (P. Besnard and S. Hanks eds.), Morgan Kaufmann 1995, 403–410.
[27] A. Paz, R. Geva and M. Studený: Representation of irrelevance relations by annotated graphs. In preparation for Fundamenta Informatica.

[28] J. Pearl and A. Paz: Graphoids: a graph based logic for reasoning about relevance relations. In Advances in Artificial Intelligence II (B. Du Boulay, D. Hogg and L. Stells eds.), North-Holland, 1987, 357–363.
[29] J. Pearl: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988.
[30] J. Pearl and T. Verma: A theory of inferred causation. In Principles of Knowledge Representation and Reasoning: proceedings of the 2nd conference (J. A. Allen, R. Fikes and E. Sandewal eds.), Morgan Kaufmann, 1991, 441–452.
[31] T. Richardson: A polynomial-time algorithm for deciding Markov equivalence of directed cyclic graphical models. In Uncertainty in Artificial Intelligence 12 (E. Horvitz and F. Jensen eds.), Morgan Kaufmann 1996, 462–469.
[32] T. Richardson: A discovery algorithm for directed cyclic graphs. In Uncertainty in Artificial Intelligence 12 (E. Horvitz and F. Jensen eds.), Morgan Kaufmann 1996, 454–461.
[33] G. Shafer: Probabilistic Expert Systems. CBMS-NSF Regional Conference Series in Applies Mathematics 67, SIAM, 1996.
[34] P. Spirtes, C. Glymour and R. Scheines: Causation, Prediction and Search. Lecture Notes in Statistics 81, Springer-Verlag, 1993.
[35] P. Spirtes: Directed cyclic graphical representations of feedback models. In Uncertainty in Artificial Intelligence 11 (P. Besnard and S. Hanks eds.), Morgan Kaufmann 1995, 491–498.
[36] W. Spohn: Stochastic independence, causal independence and shieldability. J. Philos. Logic, *9* (1980), 73–99.
[37] M. Studený and R. R. Bouckaert: On chain graph models for description of conditional independence structures. Submitted to Ann. Statist.
[38] M. Studený: On recovery algorithm for chain graphs. Res. rep. n. 1874, Institute of Information Theory and Automation, Prague, July 1996. Submitted to Int. J. Approx. Reasoning.
[39] S. Ur and A. Paz: The representation power of probabilistic knowledge by undirected graphs and directed acyclic graphs. a
comparison. Int. J. Gen. Syst., *22* (1994), 219–231.
[40] T. Verma and J. Pearl: Equivalence and synthesis of causal models. In Uncertainty in Artificial Intelligence 6 (P. P. Bonissone, M. Henrion, L. N. Kanal and J. F. Lemmer eds.), Elsevier 1991, 220–227.
[41] T. Verma and J. Pearl: An algorithm for deciding if a set of observed independencies has a causal explanation. In Uncertainty in Artificial Intelligence 8 (D. Dubois, M. P. Wellman, B. D'Ambrosio and P. Smets eds.), Morgan Kaufmann 1992, 323–330.

M. Studený, Institute of Information Theory and Automation, Pod vodárenskou věží 4, 18208 Prague 8, Czech Republic. E-mail: studeny@utia.cas.cz