Dependency Models

Jiří Vomlel

22 January 1996

1 Introduction

Definition 1 (Dependency model) Let V be a basic index set and $T(V)$ the set of all ordered triplets $(A, B|C)$, where $A, B, C \subseteq V$ are pairwise disjoint and A, B nonempty. *Dependency model over* V is every subset of $T(V)$.

Definition 2 (Regular inference rule) *Regular inference rule* with r antecedents ($r \geq 1$) is an $(r+1)$-nary relation on $T(V)$ specified by the following items:

- A finite set of symbols $\mathcal{S} = \{A_1, \ldots, A_n\}$ possibly including special symbol $\emptyset$ is given ($n \geq 3$).
- For each $k = 1, \ldots, r+1$ an ordered triplet $(\mathcal{S}_1^k, \mathcal{S}_2^k | \mathcal{S}_3^k)$ of nonempty disjoint subsets of $\mathcal{S}$ is given. Whenever any set $\mathcal{S}_i^k$ contains symbol $\emptyset$ then no other symbol is contained in this set $\mathcal{S}_i^k$. To avoid redundancy we suppose that each symbol from $\mathcal{S}$ is contained in a set $\mathcal{S}_i^k$ and that no couple of different symbols is contained in exactly same collection (?) of sets $\mathcal{S}_i^k$. The syntactic record of such an inference rule is
$$[(\mathcal{S}_1^1, \mathcal{S}_2^1 | \mathcal{S}_3^1), \ldots, (\mathcal{S}_1^r, \mathcal{S}_2^r | \mathcal{S}_3^r)] \rightarrow (\mathcal{S}_1^{r+1}, \mathcal{S}_2^{r+1} | \mathcal{S}_3^{r+1}).$$
- For every basic set V a set of $(r+1)$-tuples of triples from $T(V)$ is specified as follows: whenever there exists a substitution map m assigning to every symbol $A \in \mathcal{S}$ a subset $m(A) \in V$ ($m(\emptyset) = \emptyset$) such that:

  $\{m(A_i); i = 1, \ldots, n\}$ is a disjoint collection, and

  for each $k = 1, \ldots, r+1 \quad \bigcup\{m(A); A \in \mathcal{S}_1^k\} \neq \emptyset$ and $\bigcup\{m(B); B \in \mathcal{S}_2^k\} \neq \emptyset$

  i.e. the triplets
$$(\bigcup\{m(A); A \in \mathcal{S}_1^k\}, \bigcup\{m(B); B \in \mathcal{S}_2^k\} | \bigcup\{m(C); C \in \mathcal{S}_3^k\})$$
  form an $(r+1)$-tuple of elements of $T(V)$ called *inference instance*. First r triplets are called *antecedents* and the last one is called *consequent*.

The described *regular inference rule* is the set of all its *inference instances*.

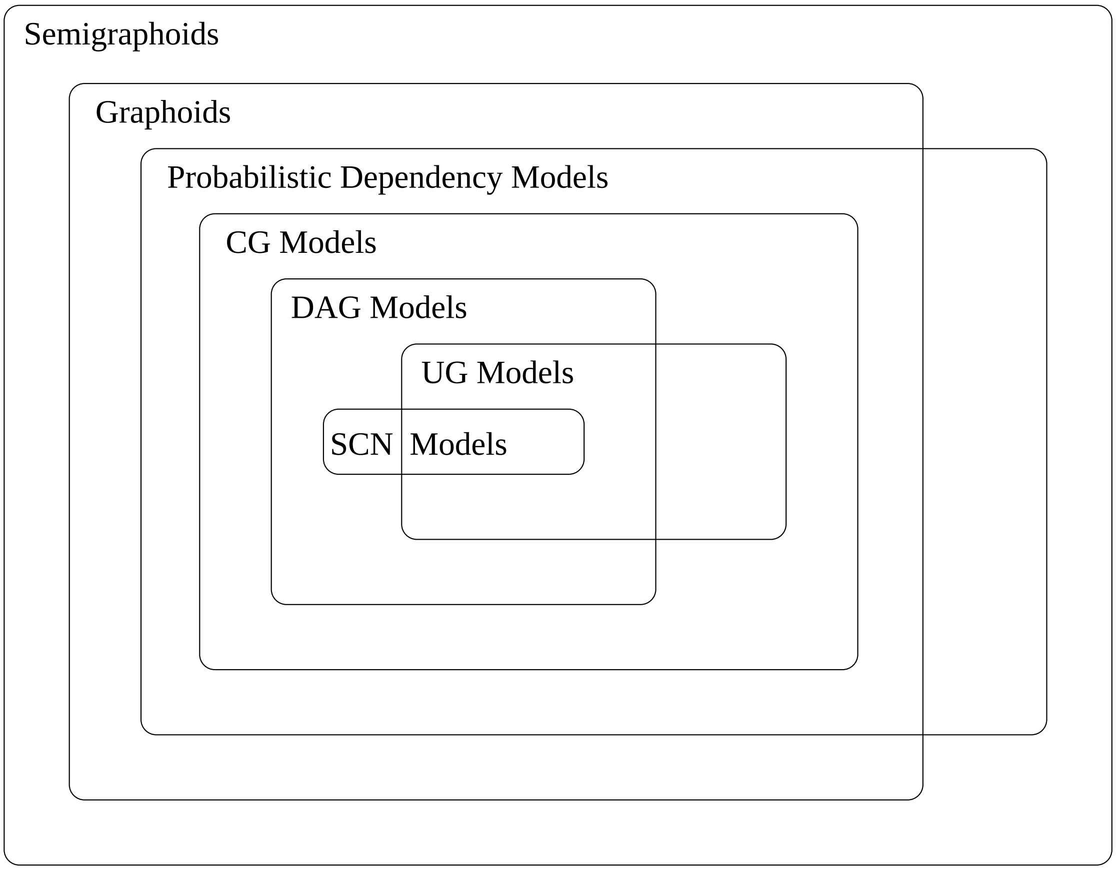
Definition 3 (Dependency model closed under an inference rule) Having a regular inference rule and a dependency model $I \subset T(V)$ we say that the dependency model$I$ is *closed under* that inference rule if the consequent of each inference instance of that inference rule belongs to dependency model I provided that antecedents of that instance belong to dependency model I.

In this article we are going to deal with the following types of dependency models:

- Semigraphoids
- Graphoids

- Probabilistic Dependency Models (CI Models)
- Non-Extreme Probabilistic Dependency Models
- Undirected Graph Models (UG Models)
- Directed Acyclic Graph Models (DAG Models or Markov Fields)
- Chain Graph Models (CG Models)
- Chordal Graph Models (Decomposable Models)
- Singly Connected Network Models (SCN Models)

Their interrelationship is shown in figure 1.



Non-Extreme Prob. Dep. Models	=	Graphoids $\cup$ Prob. Dep. Models
Chordal Graph Models	=	DAG Models $\cup$ UG Models

Figure 1: Dependency models

2 Graphoids and Semigraphoids

Definition 4 (Graphoids and Semigraphoids) A dependency model is called *graphoid* if it verifies following five inference rules:

(A1) Symmetry: $(X, Y|Z) \rightarrow (Y, X|Z)$

(A2) Decomposition: $(X, Y \cup W|Z) \rightarrow (X, Y|Z)$

(A3) Weak Union: $(X, Y \cup W|Z) \rightarrow (X, W|Z \cup Y)$

(A4) Contraction: $(X, Y|Z)$ and $(X, W|Z \cup Y) \to (X, Y \cup W|Z)$

(A5) Intersection: $(X, Y|Z \cup W)$ and $(X, W|Z \cup Y) \to (X, Y \cup W|Z)$

If a dependency model verifies only first four inference rules (A1), (A2), (A3), and (A4) then it is called *semigraphoid.*

3 Probabilistic Dependency Models

Let V denote a finite index set of a system $\{X_i\}_{i\in V}$ of finite-valued variables and let $P((X_i)_{i\in V})$ denote a joint probability distribution defined on the cartesian product $\times_{i\in V}\mathcal{X}_i$. For each $i \in V$, $\mathcal{X}_i$ denotes a finite set of values of the variable X_i.

Considering any joint probability distribution $P((X_i)_{i\in V})$ and a subset $C \subseteq V$, the symbol P^C denotes the marginal distribution of P defined as follows

$$P^C(A) = P(A \times (\times_{i\in V\setminus C}\mathcal{X}_i))$$

whenever A is a subset of $\times_{i\in C}\mathcal{X}_i$).

Definition 5 (Probabilistic conditional independence) Suppose that $P((X_i)_{i\in V})$ is a joint probability distribution. Having a triplet $(A, B|C)$ of pairwise disjoint subsets of V, where A and B are nonempty, we say that $\{X_i\}_{i\in A}$ is *conditionally independent* of $\{X_i\}_{i\in B}$ given $\{X_i\}_{i\in C}$ iff

$$\forall (x_i)_{i\in V} \in \times_{i\in V}\mathcal{X}_i \quad P^{A\cup B\cup C}((x_i)_{i\in A\cup B\cup C}) \cdot P^C((x_i)_{i\in C}) = P^{A\cup C}((x_i)_{i\in A\cup C}) \cdot P^{B\cup C}((x_i)_{i\in B\cup C}).$$

Definition 6 (Probabilistic Dependency Model) A dependency model $I \subseteq T(V)$ is the *probabilistic dependency model* iff I is the set of all and only triplets representing valid conditional statements for a probability distribution P.

Definition 7 (Non-Extreme Probabilistic Dependency Model) A *non-extreme probabilistic dependency model* is any probabilistic dependency model associated with a probability distribution P which is restricted to the open interval $(0, 1)$ i.e excluding 0's and 1.

Definition 8 (Probabilistically sound inference rule) An inference rule is *probabilistically sound* iff every probabilistic dependency model is closed under it.

Definition 9 (Minimal inference rule) Consider a probabilistically sound inference rule with r antecedents. In case it has at least one inference instance $[t_1, \ldots, t_{r+1}] \in T(V)^{r+1}$ such that no proper subset of the antecedent set $\{t_1, \ldots, t_r\}$ has the property that each probabilistic dependency model containing it contains also the consequent t_{r+1}, then we will say that the inference rule is *minimal.*

Definition 10 (Perfect inference rule) Consider a probabilistically sound inference rule with r antecedents. In case it has at least one inference instance $[t_1, \ldots, t_{r+1}] \in T(V)^{r+1}$ where each proper subset of the antecendent set $\{t_1, \ldots, t_r\}$ is a probabilistic dependency model then the inference rule will be called *perfect.*

Example 1 (Minimal but not perfect inference rule)

$$[(A, B|E), (A, C|BE), (A, D|CE)] \to (A, D|E)$$

is a minimal probabilistically sound inference rule. But it is not perfect since e.g. an inference instance $\{(\{1\}, \{2\}|\emptyset), (\{1\}, \{3\}|\{2\})\}$ is not a probabilistic dependency model.

The question whether probabilistic dependency models can be characterized by a finite axiomatic system is answered by following theorem proved in [Studený 92].

Theorem 1 (No finite axiomatic system for Probabilistic dependency models) *No finite systems of regular inference rules can characterize every probabilistic dependency model in the sense that for every dependency model I on any index set V the following holds:*
I is a probabilistic dependency model iff I is closed under a finite set of regular inference rules.

4 Graph Models

Definition 11 (I-map) Given a dependency model I we say that an graph G is an *Independence map* or *I-map* if every separation in graph G implies an independence in the dependency model I.

There will be defined different separation criteria for different type of graphs (undirected graphs, directed acyclic graphs and chain graphs) in the following sections.

Definition 12 (D-map) Given a dependency model I we say that an graph G is an *Dependency map* or *D-map* if every independence in the dependency model I implies a separation in the graph G.

Definition 13 (Perfect map) A graph G is *Perfect map* of a dependency model I if it is both an I-map and a D-map.

Definition 14 (Graph isomorphism) A dependency model I is said to be *graph-isomorphic* if a graph exists which is a perfect map of I.

5 Undirected Graph Models

Definition 15 (Separation in UG) If X, Y and Z are three disjoint subsets of nodes in a graph G then we say that set Z *separates* sets X and Y iff every path between nodes in X and nodes in Y contains at least one node from Z.

Definition 16 (UG Model) A dependency model I is *Undirected Graph Model* if there exist an undirected graph G which is a perfect map of I.

Theorem 2 (Complete characterization of UG Models) *A dependency model is isomorphic to an undirected graph if and only if it verifies following five inference rules:*

(B1) Symmetry: $(X, Y|Z) \rightarrow (Y, X|Z)$

(B2) Decomposition: $(X, Y \cup W|Z) \rightarrow (X, Y|Z)$ *and* $(X, W|Z)$

(B3) Intersection: $(X, Y|Z \cup W)$ *and* $(X, W|Z \cup Y) \rightarrow (X, Y \cup W|Z)$

(B4) Strong Union: $(X, Y|Z) \rightarrow (X, Y|Z \cup W)$

(B5) Transitivity: $\forall \gamma \in V \setminus (X \cup Y \cup Z) \;\; (X, Y|Z) \rightarrow (X, \gamma|Z)$ *or* $(\gamma, Y|Z)$.

6 Directed Acyclic Graph Models

There are two equivalent separation criteria for directed acyclic graphs.

Definition 17 (d-separation) Given a DAG G, a chain C (a chain in a directed graph is a sequence of adjacent nodes, the direction of the arrows does not matter) from a node α to a node β is said to be blocked by the set of nodes Z if there is a node $\gamma \in C$ such that either

- $\gamma \in Z$ and arrows of C do not meet head to head at γ, or
- $\gamma \notin Z$, nor γ has any descendants in Z, and the arrows of C do meet head to head at γ.

Two subsets of nodes X and Y are said to be *d-separated* by Z, iff all chains between the nodes in X and the nodes in Y are blocked by Z.

In [Lauritzen 90] a criterion equivalent to d-separation was proposed.

Definition 18 (a-separation) If X, Y and Z are three disjoint subsets of nodes in a directed acyclic graph, then Z is said to *a-separate* X from Y, iff Z separates X from Y in the moral graph of the smallest ancestral set containing $X \cup Y \cup Z$.

Definition 19 (DAG Model) A dependency model I is *Directed Acyclic Graph Model* if there exist a directed acyclic graph G which is a perfect map of I.

Theorem 3 (Necessary condition for a DAG-isomorphism) *A necessary condition for a dependency model I to be isomorphic to a DAG is that (X, Y, Z) satisfies the following properties:*

(C1) Symmetry: $(X, Y|Z) \rightarrow (Y, X|Z)$

(C2a) Decomposition: $(X, Y \cup W|Z) \rightarrow (X, Y|Z)$ *and* $(X, W|Z)$

(C2b) Composition: $(X, Y|Z)$ *and* $(X, W|Z) \rightarrow (X, Y \cup W|Z)$

(C3) Intersection: $(X, Y|Z \cup W)$ *and* $(X, W|Z \cup Y) \rightarrow (X, Y \cup W|Z)$

(C4) Weak Union: $(X, Y \cup W|Z) \rightarrow (X, W|Z \cup Y)$

(C5) Weak Transitivity: $\forall \gamma \in V \setminus (X \cup Y \cup Z)$
$(X, Y|Z)$ *and* $(X, Y|Z \cup \gamma) \rightarrow (X, \gamma|Z)$ *or* $(\gamma, Y|Z)$

(C6) Chordality: $(\alpha, \beta|\gamma \cup \delta)$ *and* $(\gamma, \delta|\alpha \cup \beta) \rightarrow (\alpha, \beta|\gamma)$ *or* $(\alpha, \beta|\delta)$

Theorem 4 (Equivalence of two DAG models) *Two DAG models I_1 and I_2 are equivalent iff they have*

- *equivalent underlying undirected graphs G_1 and G_2 made from their respective directed graphs just by losing the orientation of their edges, and*
- *equivalent all complexes composed from three nodes connected with two oriented edges meeting in one of these nodes head to head.*

7 Chordal Graph Models

Definition 20 (Chordal Graph Model) A dependency model I is a *Chordal Graph Model* if there exist a chordal undirected graph G which is a perfect map of I.

Theorem 5 (Decomposable probability distribution) *If I is chordal graph model with respect to a chordal undirected graph G, then the corresponding joint probability distribution P can be written as a product of the distributions of the cliques of G divided by a product of their intersections.*

This theorem together with its proof can be found in [Pearl 88].

8 Chain Graph Models

Definition 21 (c-separation) If X, Y and Z are three disjoint subsets of nodes in a chain graph, then Z is said to *c-separate* X from Y, iff Z separates X from Y in the undirected graph made from the smallest ancestral set containing $X \cup Y \cup Z$ by connecting all nodes having child nodes in the same component of this ancestral graph (nodes to be in the same component must be connected by unoriented paths).

Definition 22 (Chain Graph Model) A dependency model I is *Chain Graph Model* if there exist a chain graph G which is a perfect map of I.

9 Singly Connected Network Models

Definition 23 (Singly Connected Network Model) A dependency model I is *Singly Connected Network Model* if there exist a singly connected directed graph G (polytree or polyforest) which is a perfect map of I.

Theorem 6 (Complete characterization of SCN Models) *A dependency model is isomorphic to a singly connected directed acyclic graph if and only if it verifies following ten inference rules:*

(P1) Symmetry: $(X, Y|Z) \rightarrow (Y, X|Z)$

(P2a) Decomposition: $(X, Y \cup W|Z) \rightarrow (X, Y|Z)$ *and* $(X, W|Z)$

(P2b) Composition: $(X, Y|Z)$ *and* $(X, W|Z) \rightarrow (X, Y \cup W|Z)$

(P3) Intersection: $(X, Y|Z \cup W)$ *and* $(X, W|Z \cup Y) \rightarrow (X, Y \cup W|Z)$

(P4) Weak Union: $(X, Y \cup W|Z) \rightarrow (X, W|Z \cup Y)$

(P5) Contraction: $(X, Y|Z)$ *and* $(X, W|Z \cup Y) \rightarrow (X, Y \cup W|Z)$

(P6) Weak Transitivity: $\forall \gamma \in V \setminus (X \cup Y \cup Z)$
$(X, Y|Z)$ *and* $(X, Y|Z \cup \gamma) \rightarrow (X, \gamma|Z)$ *or* $(\gamma, Y|Z)$

(P7) Semi-strong Union: $\forall W, Z \subseteq V \setminus \{\alpha, \beta\}$
$(\alpha, \beta|Z)$ *and not* $(\alpha, \beta|\emptyset) \rightarrow (\alpha, \beta|Z \cup W)$

(P8) Semi-strong Atringuality: $\forall W \subseteq V \setminus \{\alpha, \gamma\}, \forall Z \subseteq V \setminus \{\beta, \gamma\}$
not $(\alpha, \gamma|W)$ *and not* $(\gamma, \beta|Z) \rightarrow (\alpha, \beta|\gamma)$ *or* $(\alpha, \beta|\emptyset)$

(P9) Shrinkage: $\forall Z_1, Z_2$ *such that* $Z = Z_1 \cup Z_2$
$(X, Y|Z) \rightarrow (X, Y|Z_1)$ *or* $(X, Y|Z_2)$

(P10): $\forall \delta \in V \setminus \{\alpha, \beta, \gamma\}, \forall W \subseteq V \setminus \{\alpha, \gamma\}, \forall Z \subseteq V \setminus \{\beta, \gamma\}$
not $(\alpha, \gamma|W)$ *and not* $(\gamma, \beta|Z) \rightarrow (\alpha, \delta|\gamma \cup \beta)$ *or* $(\beta, \delta|\gamma \cup \alpha)$ *or* $(\alpha \cup \beta, \delta|\emptyset)$

(P11) Marginal Chordality: $(\alpha, \beta|\emptyset)$ *and* $(\gamma, \delta|\emptyset) \rightarrow (\alpha, \beta|\gamma)$ *or* $(\alpha, \beta|\delta)$

Proof of this theorem can be found in [Campos 94].

10 Comparisons

Example 2 (UG Model which is not DAG Model) $I(\alpha, \delta|\beta \cup \gamma)$ and $(\beta, \gamma|\alpha \cup \delta)$
For corresponding graph see figure 2.

Example 3 (DAG Model which is not UG Model) $I(\alpha, \beta|\emptyset)$ and not $(\alpha, \beta|\gamma)$
For corresponding graph see figure 3.

Example 4 (CG Model which is neither UG Model nor DAG Model) $I(\alpha, \beta|\emptyset)$ and $I(\alpha, \delta|\beta \cup \gamma)$ and $(\beta, \gamma|\alpha \cup \delta)$
For corresponding graph see figure 4.

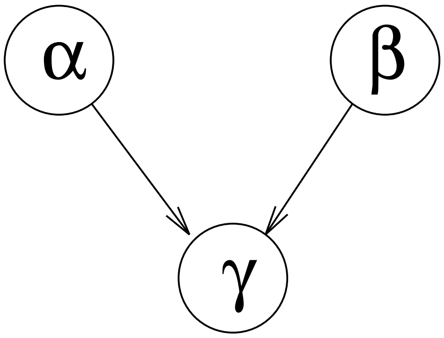


Figure 2: UG from example 2

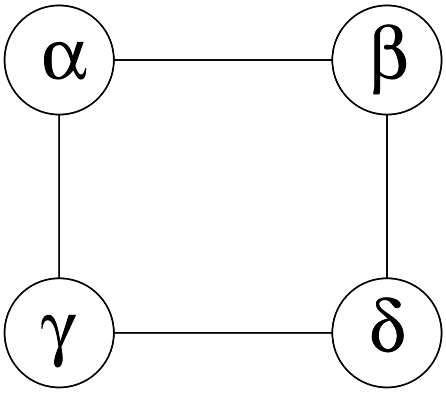


Figure 3: DAG from example 3

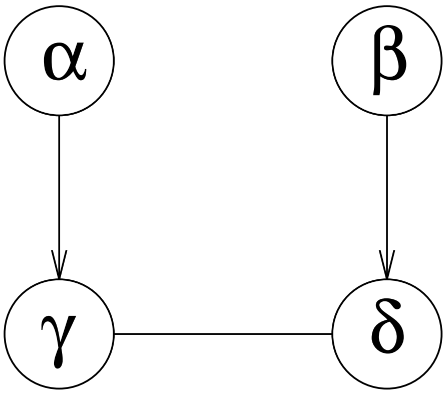


Figure 4: CG from example 4

Models	Number
Semigraphoids	26,424
Graphoids	
Probabilistic Dependency models	18,300
Non-extreme probabilistic dependency models	
Chain Graph (CG) models	? 117
Directed Acyclic Graph (DAG) models	? 108
Undirected Graph (UG) models	64
Chordal Graph models (decomposable models)	61
Singly connected network (SCN) models	

Table 1: Number of different models in the case of four variables

Models	Inference rules
Semigraphoids	A1 - A4
Graphoids	A1 - A5
Probabilistic dependency models	∞
Non-extreme probabilistic dependency models	? ∞
Chain graph (CG) models	unknown
Directed acyclic graph (DAG) models	unknown
Undirected graph (UG) models	B1 - B5
Chordal graph models (decomposable models)	? B1 - B5 and C6
Singly connected network (SCN) models	P1 - P11

Table 2: Necessary and sufficient inference rules to cover all possible models

References

[Campos 94] de Campos, L., M., 1994, *Independency Relations in Singly Connected Networks*, Techn. Report DECSAI-94214, Universidad de Granada.

[Lauritzen 90] Lauritzen, S., L., Dawid, A., P., Larsen, B., N., Leimer, H.-G., 1990, Independence properties of directed markov fields, *Networks* 20, 491-505.

[Pearl 88] Pearl, J., 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA.

[Studený 92] Studený, M., 1992, Conditional independence relations have no finite complete characterization, *Information Theory, Statistical Decision Functions and Random Processes*: transactions of the 11th Prague Conference vol. B (S. Kubík, J.A. Víšek eds.), Kluwer, Dordrecht - Boston - London (also Academia, Prague), 377-396.

[Studený 94] Studený, M., Boček, P., 1994, CI-models Arising among 4 Random Variables, *Proc. 3rd Workshop on Uncertainty Processing in Expert Systems*, Třešť: 268-282.