



**HAL**  
open science

## Integrating ontological knowledge for iterative causal discovery and vizualisation

Montassar Ben Messaoud, Philippe Leray, Nahla Ben Amor

► **To cite this version:**

Montassar Ben Messaoud, Philippe Leray, Nahla Ben Amor. Integrating ontological knowledge for iterative causal discovery and vizualisation. ECSQARU 2009, 2009, Verona, Italy. pp.168-179, 10.1007/978-3-642-02906-6\_16 . hal-00412286

**HAL Id: hal-00412286**

**<https://hal.science/hal-00412286>**

Submitted on 17 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Integrating Ontological Knowledge for Iterative Causal Discovery and Visualization

Montassar Ben Messaoud<sup>1</sup>, Philippe Leray<sup>2</sup>, and Nahla Ben Amor<sup>1</sup>

<sup>1</sup> LARODEC, Institut Supérieur de Gestion Tunis

41, Avenue de la liberté, 2000 Le Bardo, Tunisie.

benmessaoud.montassar@hotmail.fr, nahla.benamor@gmx.fr

<sup>2</sup> Knowledge and Decision Team

Laboratoire d'Informatique de Nantes Atlantique (LINA) UMR 6241

Ecole Polytechnique de l'Université de Nantes, France.

philippe.leray@univ-nantes.fr

**Abstract.** Bayesian networks (BN) have been used for prediction or classification tasks in various domains. In the first applications, the BN structure was causally defined by expert knowledge. Then, algorithms were proposed in order to learn the BN structure from observational data. Generally, these algorithms can only find a structure encoding the right conditional independencies but not all the causal relationships. Some new domains appear where the model will only be learnt in order to discover these causal relationships. To this end, we will focus on discovering causal relations in order to get Causal Bayesian Networks (CBN). To learn such models, interventional data (i.e. samples conditioned on the particular values of one or more variables that have been experimentally manipulated) are required. These interventions are usually very expensive to perform, therefore the choice of variables to experiment on can be vital when the number of experimentations is restricted. In many cases, available ontologies provide high level knowledge for the same domain under study. Consequently, using this semantical knowledge can turn out of a big utility to improve causal discovery. This article proposes a new method for learning CBNs from observational data and interventions. We first extend the greedy approach for perfect observational and experimental data proposed in [13], by adding a new step based on the integration of ontological knowledge, which will allow us to choose efficiently the interventions to perform in order to obtain the complete CBN. Then, we propose an enriched visualization for better understanding of the causal graphs.

## 1 Introduction

Over the last few years, the use of ontologies is becoming increasingly widespread by the computer science community. The main advantage of ontologies is essentially that they try to capture the semantics of domain expertise by deploying knowledge representation primitives enabling a machine to understand the relationships between concepts in a domain [3]. A lot of solutions based on ontologies

have been implemented in several real applications in different areas as natural language translation, medicine, electronic commerce and bioinformatics.

In the other hand, causal discovery remains a challenging and important task [12]. One of the most common techniques for causal modeling are *Causal Bayesian Networks* (CBNs), which are probabilistic graphical models, where causal relationships are expressed by directed edges [18]. Like classical Bayesian Networks, CBNs need *observational data* to learn a partially directed model but also they require *interventional data*, i.e. samples conditioned on the particular values of one or more variables that have been experimentally manipulated, in order to fully orient the model.

Several researches have proposed algorithms to learn the structure of a CBN from observational and experimental data.

This paper goes farther by integrating ontological knowledge for more efficient causal discovery and proposes, henceforth, an enriched visualization of causal graphs based on semantical knowledge.

Less works has been done on this aspect of combining the power of Bayesian networks and ontologies. And most of them focus on developing probabilistic frameworks for combining different ontology mapping methods [16, 22].

The remainder of this paper is organized as follows: Section 2 provides some notations and basic definitions on Bayesian networks, Causal Bayesian Networks as well as ontologies. Section 3 describes the new approach for causal discovery based on ontological knowledge. Finally, section 5 proposes different tools to obtain an enriched visualization of CBN.

## 2 Background and notations

This section gives basic definitions on Bayesian networks, causal Bayesian networks and ontologies. The following notations and syntactical conventions are used: Let  $X = \{X_1, X_2, \dots, X_N\}$  be a set of variables. By  $x$  we denote any instance of  $X$ . For any node  $X_i \in X$  we denote by  $\text{Pa}(X_i)$  (resp.  $\text{Ne}(X_i)$ ) the parents (resp. neighbors) of the variable  $X_i$ .  $P(X_i)$  is used to denote the probability distribution over all possible values of variable  $X_i$ , while  $P(X_i=x_k)$  is used to denote the probability that the variable  $X_i$  is equal to  $x_k$ .

### 2.1 Bayesian networks

A *Bayesian Network* (BN) [17] consists of a Directed Acyclic Graph (DAG) and a set of conditional probability tables (CPTs) such that each node in the DAG corresponds to a variable, and the associated CPT contains the probability of each state of the variable given every possible combination of its parents states i.e.  $P(X_i|\text{Pa}(X_i))$ . Bayesian networks are very suited for probabilistic inference, since they satisfy an important property known as *Markov property*, which states that each node is independent of its non-descendants given its parents and leads to a direct factorization of the joint distribution into the product of the conditional distribution of each variable  $X_i$  given its parents  $\text{Pa}(X_i)$ . Therefore, the

probability distribution relative to  $X=(X_1, X_2, \dots, X_n)$  can be computed by the following chain rule:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1..n} P(x_i | Pa(X_i)). \quad (1)$$

Note that several BNs can model the same probability distribution. Such networks are called equivalent or Markov equivalent [24].

**Definition 1.** *A Complete Partially Directed Acyclic Graph (CPDAG) is a representation of all equivalent BNs. The CPDAG contains the same skeleton as the original DAG, but possesses both directed and undirected edges. Every directed edge  $X_i \rightarrow X_j$  of a CPDAG denotes that all DAGs of this class contain this edge, while every undirected edge  $X_i - X_j$  in this CPDAG-representation denotes that some DAGs contain the directed edge  $X_i \rightarrow X_j$ , while others contain the oppositely orientated edge  $X_i \leftarrow X_j$ .*

Under Causal sufficiency assumption (i.e. there are no latent variables that influence the system under study), many structure learning techniques using perfect *observational data* are available and can be used to learn CPDAG and then choose a possible complete instantiation in the space of equivalent graphs defined by this CPDAG. These techniques can be classified into two groups, namely *score-based* and *constraint-based* algorithms.

*Score-based* algorithms [4, 5] attempt to identify the network that maximizes a scoring function evaluating how well the network fits the data while *constraint-based* algorithms [12, 21] look for (in)dependencies in the data and try to model that information directly into the graphical structure.

## 2.2 Causal Bayesian Networks

A *Causal Bayesian Network* (CBN) [18] is a Bayesian network with the added property that all edges connecting variables represent autonomous causal relations. Given a CBN, we can go further than probabilistic inference to perform causal inference. Pearl has introduced the *do operator* as standard notation for external intervention on causal models. In fact, the effect of an action "do( $X_j=x_k$ )" in a causal model corresponds to a minimal perturbation of the existing system that forces the variable  $X_j$  to the value  $x_k$ . In other terms, causal inference is the process of calculating the effect of manipulating some set of variables  $X_i$  on the probability distribution of some other set of variables  $X_j$ , this is denoted as  $P(X_i | do(X_j=x_k))$ .

Several researches in the literature have proposed algorithms to learn CBN's structure. We can, in particular, mention Tong and Koller [23] and Cooper and Yoo [6], which developed a score-based techniques to learn a CBN from a mixture of experimental and observational data. Eberhardt et al. [9] performed a theoretical study on the lower bound of the worst case for the number of experiments to perform to recover the causal structure.

Recently, Meganck et al. [13] proposed a greedy constraint-based learning algorithm for CBNs using experiments. The MyCaDo (My Causal DiscOvery) algorithm, which represents the main application of their approach, is a structure learning algorithm able to select appropriate interventions or experiments (i.e. randomly assigning values to a single variable and measuring some other possible response variables) needed to build a CBN. As input, the MyCaDo algorithm needs a CPDAG. Meganck et al. proposed to use the PC algorithm proposed by Spirtes et al. [21] with modified orientation rules to learn the initial CPDAG, but other structure learning algorithms can be taken into consideration.

When applying PC algorithm, we start with a complete undirected network. Then, we remove edges when independencies are found; we call such step Skeleton discovery. The edge orientation, in such algorithm, is based on v-structure discovery and edge propagation. Based on the already oriented edges, an inferred edge step will apply some orientation rules until no more edges can be oriented.

In this step, we need also experimental data to perform interventions on the system. To learn a CBN from interventional data, three major parts can be distinguished in the MyCaDo algorithm:

- First of all, it tries to maximize an utility function based on three variables:  $gain(exp)$ ,  $cost(exp)$ ,  $cost(meas)$ , respectively, the gained information, the cost of performing an experiment and the cost of measuring other variables, to decide which experiment should be performed and hence also which variables will be measured.

If we denote performing an experiment on  $X_i$  by  $A_{X_i}$ , and measuring the neighboring variables by  $M_{X_i}$ , then the utility function will be as follows:

$$U(A_{X_i}) = \frac{\alpha gain(A_{X_i})}{\beta cost(A_{X_i}) + \gamma cost(M_{X_i})} \quad (2)$$

where  $\alpha, \beta$  and  $\gamma$  are measures of importance for every term.

In this formula,  $gain(A_{X_i})$  takes into consideration the number of undirected neighbors  $Ne_U(X_i)$  (e.g. nodes that are connected to  $X_i$  by an undirected edge) and the amount of possible inferred edges, after performing an experiment on  $X_i$ .

Three decision criteria were proposed *Maximax*, *Maximin* and *Expected Utility* depending on the type of situation in which to perform the experiments it might be advantageous to choose a specific criterion.

- Secondly, the selected experiment will be performed.
- Finally, the results of this experiment, will be analyzed in order to direct a number of edges in the CPDAG.

Note that the amount of edges of which the direction can be inferred after performing an experiment is entirely based on the instantiation (i.e. assignation of a direction) of the undirected edges connected to the one being experimented on.

This process will be iterated until we obtain the correct structure of the CBN. Note that we can have a non-complete causal graph (i.e. not all edges are oriented) as final output of MyCaDo algorithm.

Borchani et al. [2] have also described another approach for learning CBNs from incomplete observational data and interventions.

Furthermore, other approaches were proposed to learn graphical models that can handle latent variables. For this kind of hidden variable modeling, Meganck et al. [14, 15] and Maes et al [11] studied which experiments were needed to learn CBN with latent variables under the assumption of faithful distribution (i.e. the observed samples come from a distribution which independence properties are exactly matched by those present in the causal structure of a CBN). Their solution offers a new representation based on two paradigms that model the latent variables implicitly, namely *Maximal Ancestral Graphs* and *Semi-Markovian Causal Models*, in order to perform probabilistic as well as causal inference.

### 2.3 Ontology

An **ontology** [10] is defined as a formal explicit specification of a shared conceptualization. In other terms, it is a formal representation of a set of concepts within a domain and the relations between these concepts. The relations in an ontology are either taxonomic (e.g. is-a, part-of) or non-taxonomic (i.e. user-defined). The taxonomic ones are the commonly used relations. In the simple case, the ontology takes the form of a tree or hierarchy representing concept taxonomy. Formally:

**Definition 2.** A **Concept Taxonomy**  $H = (C, E, Rt)$  is a directed acyclic graph where,  $C = \{c_1, c_2, \dots, c_n\}$  is the set of all concepts in the hierarchy,  $E$  is the set of all subsumption links (is-a) and  $Rt$  is the unique root of this DAG. The primary structuring element of concept taxonomy is the subsumption relationship, which supposed that if the concept  $c_i$  is a child of concept  $c_j$  then all properties of  $c_j$  are also properties of  $c_i$  and we say that  $c_j$  subsumes  $c_i$ .

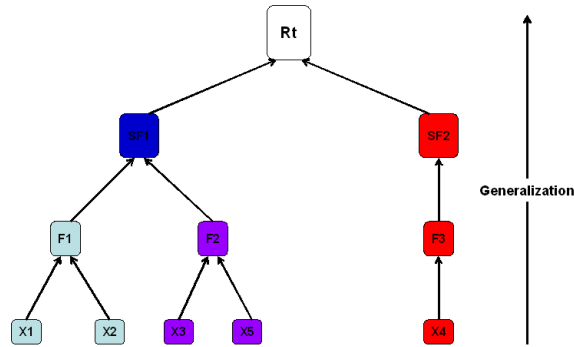
Regarding any concept taxonomy hierarchy, we can give the following notations:

- $pths(c_i, c_j)$ : the set of paths between the concepts  $c_i$  and  $c_j$ , where  $i \neq j$ ,
- $len_e(e)$ : the length in number of edges of the path  $e$ ,
- $mscs(c_i, c_j)$ : the most specific common subsumer of  $c_i$  and  $c_j$ , where  $i \neq j$ .

The major contribution of using concept taxonomies is, essentially, to present the domain knowledge in a declarative formalism.

For instance, the Gene Ontology (GO) [25] is one of the principal knowledge resource repository in the bioinformatic field and represents an important tool for the representation and processing of information about genes and functions. It provides controlled vocabularies for the designations of cellular components, molecular functions and biological processes.

**Example 1.** Figure 1 shows an ontology toy example or more precisely, an is-a tree, where the leaf nodes ( $X1, X2, \dots, X5$ ) are genes. Their direct subsumers ( $F1, F2, F3$ ) represent the biological functions i.e. every set of genes in a GO share a very specified function. Also, the biological functions are subsumed by their biological super-functions ( $SF1, SF2$ ). Note that we can use the term of generalization as shown in figure 1 to refer the subsuming process. For this case of is-a tree, only one path can be found between two different concepts. To more illustrate this purpose, let us consider the two concepts  $X1$  and  $SF2$ . Here,  $pths(X1, SF2)$  represents the set of the edges  $X1-F1, F1-SF1, SF1-RT$  and  $RT-SF2$ . Consequently the length in number of edges of the corresponding path will be equal to four.



**Fig. 1.** Example of hierarchy representing a taxonomy of concepts.

Recently, several works highlighted the importance of evaluating the strength of the semantic links inside domain ontologies. We can distinguish three major classes of semantic measures, namely *semantic relatedness*, *semantic similarity* and *semantic distance*, evaluating, respectively, the closeness, the resemblance and the disaffection between two concepts.

The *semantic similarity* represents a special case of *semantic relatedness*. For instance, if we consider the two concepts *wind turbine* and *wind*, they would be more closely related than, for example the pair *wind turbine* and *solar panel*. However the latter concepts are more similar. Therefore, all pairs of concepts with a high semantic similarity value (i.e. high resemblance) have a high semantic relatedness value whereas the inverse is not necessarily true. In the other hand, the semantic distance is an inverse notion to the semantic relatedness.

The major approaches of measuring semantic distance are *Rada et al.'s distance*, *Sussna's distance* and *Jiang and Conrath's distance*. For the semantic similarity, we find *Leacock and Chodorow's similarity*, *Wu and Palmer's similarity* and *Lin similarity*, while for semantic relatedness, the most used one is *Hirst and St Onge's relatedness*. See [1] for a comparative study of these measures.

In what follows, we will focus on semantic distances and in particular on the classical *Rada et al.’s distance* [19], which can be replaced by any other semantic distance.

This distance is based on the shortest path between the nodes corresponding to the items being compared such that the shorter the path from one node to another, the more similar they are. Thus, given multiple paths between two concepts, we should take the length of the shortest one. Formally, given two concepts  $c_i$  and  $c_j$  the Rada et al.’s distance is defined by:

$$dist_{rmbb}(c_i, c_j) = \min_{p \in pths(c_i, c_j)} len_e(p) \quad (3)$$

Consequently, we will compute a distance matrix, giving us all distances between all pairs of concepts.

**Table 1.** Rada et al.’s distance matrix between the nodes given in Fig.1

	Rt	SF1	SF2	F1	F2	F3	X1	X2	X3	X4	X5
Rt	0	1	1	2	2	2	3	3	3	3	3
SF1	1	0	2	1	1	3	2	2	2	4	2
SF2	1	2	0	3	3	1	4	4	4	2	4
F1	2	1	3	0	2	4	1	1	3	5	3
F2	2	1	3	2	0	4	3	3	1	5	1
F3	2	3	1	4	4	0	5	5	5	1	5
X1	3	2	4	1	3	5	0	2	4	6	4
X2	3	2	4	1	3	5	2	0	4	6	4
X3	3	2	4	3	1	5	4	4	0	6	2
X4	3	4	2	5	5	1	6	6	6	0	6
X5	3	2	4	3	1	5	4	4	2	6	0

**Example 2.** In figure 1, every term has at least one generalization’s path back to the top node. The full Rada et al.’s distance matrix concerning the concept taxonomy is presented in Table 1. For instance,  $dist_{rmbb}(SF1, X2) = 2$  since the shortest path between the two concepts SF1 and X2 is composed of the two edges SF1-F1 and F1-X2.

### 3 Our approach for causal discovery and visualization based on ontological knowledge

Causal Bayesian networks are used to model domains under uncertainty [17]. In many cases, available ontologies provide consensual representation of the same domain and a full description of the knowledge model. Basically, ontologies attempt to find the complete set of concepts covering any domain as well as the relationships between them. Such domain knowledge can be efficiently used to



enrich the learning process of Causal Bayesian networks and optimize the causal discovery.

In order, to illustrate this idea we propose to extend the MyCaDo (My Causal DiscOvery) algorithm [13] used to learn causal structures with experiments by integrating the ontological knowledge, extracted using Rada et al. semantic distance calculation, for causal discovery and visualization.

We make the assumption that we consider a concept taxonomy hierarchy, in which the leaves are also the nodes of the Bayesian network. However, the causal relations we want to discover and model with our CBN did not exist in the corresponding ontology, therefore, each model representation complete the other.

### 3.1 Integrating ontological knowledge for causal discovery

In what follows let us assume that after performing an experiment on a variable  $X_i$ , we can measure all neighboring variables  $Ne_U(X_i)$ .

Our utility function  $U(.)$  is an extension of the one proposed in [13] (see subsection 2.2) by generalizing the first term  $Ne_U(X_i)$  and replacing it by the semantical inertia, denoted by  $SemIn(Ne_U(X_i))$ .

This notion will enable us to integrate our ontological knowledge extracted by calculating semantic distances from ontologies and guide the choice of the best experiment. In many situations, we can have a node  $X_i$  with a high number of neighbors but all those neighbors are very close semantically. This implies that we will have, exactly, two alternatives (i.e.  $X_i$  is the direct cause of all his neighbors or the inverse).

For instance, in bio-informatics, biologists would discover more causal relations between biological functions. Considering a neighborhood which is very close semantically will reduce the number of functions under study and consequently the informative contribution of the experiment would be very low. However, in the case of distant neighbors, there will be more cause-to-effect relations between biological functions to find. Such causal discoveries have an important scientific contribution in the bio-informatic field.

The semantical inertia will be as follows:

$$SemIn(M) = \frac{\sum_{X_i \in M} \min_{p \in pths(X_i, mscs(M))} (len_e(p))}{card(M)} \quad (4)$$

Let us consider the CPDAG in figure 2 in order to illustrate our approach and compare it to the original MyCaDo. We will also use the concept taxonomy hierarchy presented in figure 1. Suppose that we will perform an action on  $X_2$ . Figure 3 summarizes all possible instantiations of the edges  $X_i-Ne(X_i)$ .

It is clear that the following edges  $X_2 - X_3$  and  $X_2 - X_5$  have the same direction because the two neighbors  $X_3$  and  $X_5$  share the same function in the corresponding ontology. In this case, we will have three instantiations whereas, by applying MyCaDo we will find exactly six. Consequently, we reduced considerably the set of possible instantiations in the particular PDAG (i.e. CPDAG

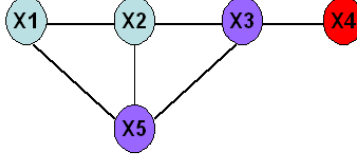


Fig. 2. An example of CPDAG

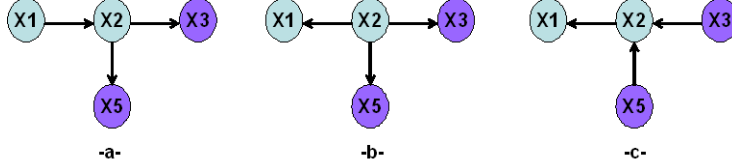


Fig. 3. All possible instantiations for  $X_2 - Ne_U(X_2)$

with some oriented edges).

Here the most specific common subsumer of  $Ne_U(X_2) = \{X_1, X_3, X_5\}$  is the concept SF1. According to the ontology in figure 1 and table 1, the semantical inertia of the nodes  $X_1$ ,  $X_3$  and  $X_5$  is as follows:

$$SemIn(X_1, X_3, X_5) = \frac{\sum_{X_j \in Ne_U(X_2)} \min_{p \in pths(X_j, mscs(Ne_U(X_2)))} (len_e(p))}{\#Ne_U}$$

$$= \frac{2 + 2 + 2}{3} = 2$$

The semantical inertia presents three major characteristics. At the first glance, where all undirected neighboring nodes belong to the same biological function, the semantical inertia of the neighborhood will be equal to one. Secondly, the semantical inertia depends on the number of undirected neighbors. For example, if we eliminate  $X_1$  from the neighborhood of  $X_2$ , SemIn will automatically decrease.

$$SemIn(X_3, X_5) = \frac{1 + 1}{2} = 1$$

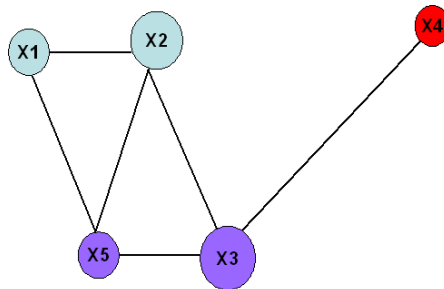
And, finally, the more the neighboring variables are distant according to the ontology, the more the semantical inertia will be important and the utility maximized. Here, if we replace the node  $X_3$  by  $X_4$ , which is more distant from the rest of neighbors of  $X_2$ , SemIn will increase considerably.

$$SemIn(X_1, X_4, X_5) = \frac{3 + 3 + 3}{3} = 3$$

It is clear now that the semantical inertia represent a generalization of  $\#Ne_U(X_i)$  and introducing it in the utility function will allow a better choice of experiments, based on ontological knowledge. With such method, we can focus the causal discovery on relations between distant concepts.

### 3.2 Integrating ontological knowledge for causal graph visualization

Visualizing large networks of hundreds or thousands nodes is a real challenging task. It is essentially due to the limitation of the screen, the huge number of nodes and edges and the limitations of human visual abilities. In this context, our approach can turn out of big utility to improve actual visual tools. The main purpose was to propose an enriched visualization of causal models by integrating the power of semantical knowledge. The Rada et al.'s distance matrix can implement different graph drawing algorithms among which MultiDimensional Scalling (MDS) and Force Directed Placement [7, 8]. More precisely, we will adjust the node's position in the screen, referring to the matrix distance calculated above. Similarly, in [20], Ranwez et al. used the ontological distance measures to propose an alternative information visualization on conceptual maps.



**Fig. 4.** Enriched visualization corresponding to the CPDAG in Figure 2.

In figure 4, we show an enriched visualization of the CPDAG used in the previous subsection. The biological functions are more distinguished and the causal relations between those functions are, visually, more revealed.

In the other hand, we propose to adapt the size of the nodes to the utility function. This method allows to biologists to determine the ideal node to make interventions.

While judicious use of semantical distances and node's size can help considerably the understanding of the causal bayesian network, we can go farther in the visualization by adding a zoom-in/zoom-out function that allows one to visualize either the global structure of the graph or just, smaller components reduced to more general concepts in the ontology. For example, we can pass from visualizing

the genes, in our example, to an abstraction reduced to only biological functions or even to the biological super-functions.

All those visualization attributes can lead to different valuable informations about the causal bayesian network. Moreover, we will be able to investigate further into the adapted structure of the network. Since the enriched visualization can offer assistance to the domain experts, we can change the MyCaDo algorithm into semi-automatic method, combining the use of optimal solution of utility and visualization of causal graph. Consequently, we can ease considerably the depiction of causality.

## 4 Conclusion

In this article, we discussed how ontological knowledge can be useful for iterative causal discovery. More precisely, we extend the MyCaDo algorithm via introducing the notion of the semantical inertia. By supporting the assignment of costs to experiments and measurements, our approach permit us to guide the choice of the best experiment in order to obtain the complete causal bayesian network. We then proposed an enriched visualization of causal models, using the power of semantical knowledge extracted from ontologies.

To this end, we considered a toy example ontology as a starting point to develop our approach but this choice does not exclude the application of our works to more realistic ontology's domains. We emphasize that the work described here represents a major step in a longer-term project focusing on the knowledge integration for causal bayesian networks learning.

Directions for future work include studying how ontological knowledge can be integrated to learn causal graphical models with latent variables, or other links between causal bayesian network learning and ontologies construction.

## References

1. E. Blanchard, M. Harzallah, H. Briand and P. Kuntz. A typology of ontology-based semantic measures. In *2nd INTEROP-EMOI Open Workshop on Enterprise Models and Ontologies for Interoperability at the 17th Conference on Advanced Information Systems Engineering (CAISE'05)*, volume 160, pages 407-412. CEUR-WS, 2005
2. H. Borchani, M. Chaouachi and N. Ben Amor. Learning causal bayesian networks from incomplete observational data and interventions. In *Proceedings of Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 17-29, 2007.
3. M. Cannataro, A. Massara, P. Veltri. The OnBrowser ontology manager: Managing ontologies on the Grid. *International Workshop on Semantic Intelligent Middleware for the Web and the Grid*, Valencia, Spain, 2004.
4. D.M. Chickering. Optimal Structure Identification With Greedy Search. *Journal of Machine Learning Research*, pages 507-554, 2002.
5. G.F. Cooper and E. Herskovits. A Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning*, pages 309-347, 1992.
6. G.F. Cooper and C. Yoo. Causal discovery from a mixture of experimental and observational data. In *Proceedings of Uncertainty in Artificial Intelligence*, pages 116-125, 1999.

7. M. Crampes, S. Ranwez, J. Villerd, F. Velickovski, C. Mooney, A. Emery, N. Mille. Concept Maps for Designing Adaptive Knowledge Maps, in Concept Maps, A Special Issue of Information Visualization, Vol. 5, Issue 3, Guest Editors: S.-O. Tergan, T. Keller and R. Burkhard, Palgrave - Macmillan, 2006.
8. M. Crampes, S. Ranwez, F. Velickovski, C. Mooney, N. Mille. An integrated visual approach for music indexing and dynamic playlist composition, in proceedings of 13th Annual Multimedia Computing and Networking (MMCN'06), San Jose, CA, US, 2006.
9. F. Eberhardt, C. Glymour and R. Scheines. N-1 experiments suffice to determine the causal relations among n variables. Technical report, Carnegie Mellon University, 2005.
10. T. R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies* 43 , pages 907-928, 1993.
11. S. Maes, S. Meganck and P. Leray. An integral approach to causal inference with latent variables. In F. Russo and J. Williamson, Causality and Probability in the Sciences (Texts in Philosophy), pages 17-41. College Publications, 2007.
12. S. Mani and G.F. Cooper. Causal discovery using a Bayesian local causal discovery algorithm, Proceedings of MedInfo, pages 731-735. IOS Press, 2004.
13. S. Meganck, P. Leray, and B. Manderick. Learning causal bayesian networks from observations and experiments: A decision theoretic approach. In *Proceedings of the Third International Conference, MDAI 2006*, volume 3885 of Lecture Notes in Artificial Intelligence, pages 58-69, Tarragona, Spain. Springer, 2006.
14. S. Meganck, S. Maes, P. Leray and B. Manderick. Learning semi-markovian causal models using experiments. In *Proceedings of The third European Workshop on Probabilistic Graphical Models*, PGM 06, pages 195-206, 2006.
15. S. Meganck, P. Leray and B. Manderick. Causal graphical models with latent variables: Learning and inference. Ninth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007), pages 5-16, 2007.
16. R. Pan, Z. Ding, Y. Yu and Y. Peng: A Bayesian Network Approach to Ontology Mapping. In *Proceedings ISWC*, pages 563-577, 2005.
17. J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers, San Francisco (California), 1988.
18. J. Pearl. *Causality: Models, reasoning and inference*, MIT Press, 2000.
19. R. Rada, H. Mili, E. Bicknell and M. Blettner. Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, pages 17-30, 1989.
20. S. Ranwez, V. Ranwez, J. Villerd and M. Crampes. Ontological Distance Measures for Information Visualization on Conceptual Maps. In Proceedings of the OTM Workshops. LNCS 4278, Springer, pages 1050-1061, 2006.
21. P. Spirtes, C. Glymour and R. Scheines. *Causation, Prediction and Search*, MIT Press, 2000.
22. O. Sváb, V. Svátek. Combining Ontology Mapping Methods Using Bayesian Networks. In *Workshop on Ontology Matching at ISWC*, 2006.
23. S. Tong and D. Koller. Active learning for structure in bayesian networks. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 863-869, 2001.
24. T.Verma and J.Pearl, Equivalence and Synthesis of Causal Models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, 1990.
25. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology, *Nature Genet*, pages 25-29, 2000.