Bayesian networks for the evaluation of complex systems' availability

Ait Mokhtar El Hassene, Laggoune Radouane
Unité de recherche LaMOS
Université de Bejaia 06000, Bejaia Algérie
aitmokhtar_elhassene@hotmail.com

Chateauneuf Alaa
Institut Pascal
Université blaise Pascal, Clermont- Ferrand, France

Unlike the simple systems, very few methodologies treat the evaluation of the dependability of complex systems, especially those configured as networks, where it is difficult to take into consideration the different links and factors that can affect the availability and reliability of such systems. In this context, Bayesian networks is a very interesting tool. In fact, they permit the modelling of systems configured as network and the computation of marginal probabilities of the nodes of the system using prior and conditional probabilities. In this paper, we propose an original approach based on the factor of conditional availability for the evaluation of the availability of the drinkable water distribution network of Bejaia city. And this by taking into consideration the different links and interaction between the pumping stations of this network.

Availability evaluation, Bayesian networks, Complex systems, Availability reduction factor.

1. INTRODUCTION

In the most current papers, the evaluation of dependability methods (evaluation of reliability and availability) are generally reserved for the simple systems (series and parallel systems) or for the components. But, for the most part of industrial systems, their components are configured as networks, where the interactions between the components are defined by logical or physical links which complicate the evaluation of the dependability of these kinds of systems. In this framework, Bayesian networks (BNs) are very useful since they permit a qualitative and quantitative representation of the relations between the variables of the model. The structure of the network reflects the conditional dependencies between the variables, while the prior and conditional probabilities are used to quantify them [1].

Many papers have proposed approaches to evaluate the availability and reliability of complex systems by using BN modelling. In [2], the authors have combined the evidence theory with BNs in order to create an effective tool for the reliability analysis of systems under random uncertainties. The system reliability is evaluated the basic of “Dempster Shafer” theory. In [3], the studied system is constituted of two parallel sub-systems and each sub-system is composed of two components in series. The data used are the time to failure knowing that two components follow the Weibull distribution and the two others follow the exponential distribution. The output variable is the overall system reliability; the obtained results can be updated after the availability of new data by adding binary nodes (yes/no) that describe the system state on a given time.

Dynamic oriented object Bayesian networks (DOOBN) is another type of BNs which is also used in dependability analysis of complex systems [4], the study proposed in [4] allows to simulate failures of different components of a complex system in order to evaluate its reliability. In [5], the authors have used simulation technique to estimate the availability of a complex system according to four different scenarios. In the first scenario, the conditional probability tables (CPT) are known, in the second the CPT are unknown, the third case shows the contribution of adding additional data, and in the last case, they estimate the reliability using data collected and added over the time.

BNs can be also used to evaluate the availability of systems. In [6], authors have applied hybrid Bayesian networks (HBN) since the different causes that have influence on the availability assessment are continuous variables (time to repair, programmed preventive maintenance times and delays). BNs are also used for redundant systems with improvements of the complex systems modelling by adding a “coverage factor” [7], this factor represents the probability that a simple failure of a redundant component causes the overall system failure. It can be modelled by FT [8], but according to [7], it seems to be more useful and more meaningful to use BNs.

The first section of this paper is reserved to the Bayesian networks and the modelling of complex systems using Bayesian networks. In the second one, we present the Bayesian inference which aims to compute the marginal probabilities of the nodes, in this section we introduced a new notion based of the “availability reduction factor” for the creation of the conditional probability tables. The third and last section is an application. It aims on the evaluation of average availability of the water distribution system of Bejaia city by applying the methodology developed here.

2. BAYESIAN NETWORKS, DEFINITIONS AND PROPERTIES

2.1 Definition 1 (Bayesian networks)

A Bayesian network $\mathfrak{B} = \{\mathcal{G}, \mathbb{P}\}$ is defined by

- A directed acyclic graph $\mathcal{G} = (X, E)$ where X is a set of nodes (or vertices) and E is a set of directed links (or edges);
- A probability space $(\Omega, \mathbb{P})$;
- A set of random variables $X = \{X_1 \ldots X_n\}$ associated with the graph’s nodes $(\Omega, \mathbb{P})$ such as $\mathbb{P}(X_1 \ldots X_n) = \prod_{i=1}^{n} \mathbb{P}(X_i | Pa(X_i))$ where $Pa(X_i)$ is the set of the parent’s nodes of the node X_i in $\mathcal{G}$.

In other words, A Bayesian network is a graph where the nodes represent random variables (continuous or discrete) and the edges represent the influences between the variables of the graph. We associate the random variable X to its modalities ($X = x_1; X = x_2; \ldots X = x_n$ if X can takes n values). About the edges, they represent the causalities which can be deterministic or probabilistic. For an edge, linking the fact A and the fact B , there is a relation which is the conditional probability noted $P(B|A)$, it represents a probabilistic relation of a node known its nodes parent. For the nodes without parents, named “root” nodes, a prior probability will be assigned to them. Generally Bayesian networks (BNs) are mostly used as an efficient framework for decision-making with uncertain knowledge [7]. They describe the system as a directed acyclic graph (DAG), and not as a tree, and represent a powerful mathematical formalism to model the complex stochastic processes. They allow for exact calculation of the influences of dependent components or events on the system reliability, unlike other methods as fault tree (FT) or Petri networks.

2.2 Bayes theorem

The Bayesian networks are developed thanks to the Bayes theorem. It is a basic result in probability theory, and comes from the works of Thomas Bayes (1702 - 1761).

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \qquad (1)$$

Where $P(A)$ is the prior probability, $P(B)$ is the observations (or evidence) and the posterior probability is given by $P(A|B)$.

2.3 Complex systems modelling as Bayesian network

In the literature, several papers discuss the methods of BN construction. When modeling of complex systems in order to optimize the maintenance or to evaluate their reliability and availability, two information sources are generally considered: the expert judgment and the statistical data on the system [7]. It is also important to specify that the modeling of complex systems by BNs is a difficult and very time consuming task. The steps of BN construction are:

(i) ***Specify what we want to model:*** identify the limits of the study by defining what to include and what is not.

(ii) ***Definition of variables:*** Select the important variables of the system to take into consideration in the BN. At this step, we have also to specify the range of continuous variables and the states of discrete variables.

(iii) ***Qualitative step:*** This step aims at connecting the different nodes to each other, by directed edges, in order to express the dependencies and independencies between the nodes.

(iv) ***Quantitative step:*** It consists in creating the probability tables: the prior probability tables for the root nodes and the conditional probability tables for the other nodes. To do it, we can use the statistical data of the system or the estimations of the experts. Note that their value must be normalized; they must be between 0 and 1 and their sum must be equal to 1.

(v) ***Verification:*** It is generally made by performing sensitivity analysis and behavior tests by simulating know scenarios.

Note: to facilitate the determination of the different linking and dependencies between the nodes of the BN, we can use the FMECA analysis (Failure Modes, Effects and Criticality Analysis) or Fault tree analysis.

2.4 Bayesian network and dependability analysis assessment of complex systems

In literature, except the BNs, there are three traditional methods used in dependability of complex systems; the fault tree (FT), Markov chains (MC) and Petri networks (PN) [9]

2.4.1. Fault tree

This method gives important results of modelling since it permits the integration of different kinds of knowledge (organisational, decisional, technical and human aspect) and allows considering the dependencies between events. It also gives an exact computation thanks to its Boolean representation of the elementary events.
However, when the system is affected by multiple failures with several consequences (which is generally the case of the industrial complex systems) the model needs a representation with multi-state variables. In this case FT can't be used. We can also add that the FT method permit the analysis of one event. Contrariwise, BN allows the use of multi-states variables and the analysis of several events in the same model.
Many papers have proposed methods that permit the transformation of FT to BN [9].

2.4.2. Markov chains

This method is adequate for the reliability and availability analysis of systems; it allows exact analysis of the failure probability even when the system components are dependent between them. It also permits the representation of multi-state variables, however, to reproduce the different interdependencies and links between the system variables we need to use a very large number of variables and the modelling becomes very difficult and leads to a combinatory explosion of the number of states [10], this gap is the main defect of this method. According to [11], thanks to the use of conditional probability tables, BN permit to avoid this combinatory explosion.

2.4.3. Petri Networks

It is a traditional method of the dependability modelling; it is also used in the domain of dynamic reliability and maintenance optimization policy. It is based on the simulation procedures like Monte Carlo analysis and other variants of this method which leads to the following constraints [9]:

- Inefficient consideration of low frequency events (accidents).
- They do not allow easily integrating evidence.

We can note that the modelling objective of the BN and PN is the same but the way to deal with the issue is very different.

In this paper, we have opted for the use of Bayesian networks since they permit:

- The use of imprecise of historical data.
- The use of expert judgement to complete the lack of data.
- The use of multi-states variables which are useful to model the event with several effects.

3. BAYESIAN INFERENCE

Bayesian networks are essentially used to compute the marginal and posterior probabilities of events connected between each other by relations of cause and effect. And this, by using prior probability tables for root nodes and conditional probability tables for the other BN nodes, this use is called “inference”. The model represented by a BN is not a statistical closed model; in fact, we can integrate new information. By changing the likelihood of certain nodes, the posterior probability of the system will be changed (data updating) [12]. This property (updating) is very interesting of the diagnostic application, where its appreciation will change according to one or many observations [13].

There are two kind of Bayesian inference, exact and approximate inference method. For the first kind, we can find two classes: message passing method introduced by Pearl [14], which is used for networks configured as tree or poly-tree and the methods using grouping nodes like the junction tree method of Jensen [15]. The main problem of the direct inference methods is the computing time, since the BNs are generally used for complex systems with e great number of variables, so the BN size of this kind of the complex systems is very large. And the execution time of the exact inference algorithms is very important according to the complexity of the graph (the number of variables and their modalities) [16]. To deal with this problem, the approximate inference method is very interesting then the exact inference methods regarding the computation time.

We have to note also that for some kind of BNs (BN that contain continuous and discrete nodes: hybrid BNs), we can just use the approximate inference methods. These methods are generally based on stochastic methods type MCMC (Monte Carlo Markov Chain) [17].

In this paper we have opted for the use of the Junction tree method (also called clustering or clique-tree propagation algorithm) introduced by Jensen in 1990 [15]. This method can be applied for all the DAG structures.

3.1 Junction tree algorithm

This algorithm can be divided into two phases, the junction tree construction phase and the message propagation phase.

3.1.1. Construction of the junction tree

Moralisation: the first step of the transformation of the graph is the moralisation. It consists on connecting two by two the parents of each node by non-directed edges. After having moralised the graph, we finished the transformation by deleting the direction of each edge.

Triangulating the moral graph: a non-directed graph is triangulated if every cycle of length four or greater contains an arc that connects two nonadjacent nodes in the cycle. It is made according to the following steps:

i. Associate to each node of the BN $\boldsymbol{X_i}$ a "weight" equal to the product of the modalities of $\boldsymbol{X_i}$and its neighbours;
ii. Select the node $\boldsymbol{X_i}$ whose the weight is minimal and which caused the least number of edge to add (to form a clique $\boldsymbol{C_i}$ of cycle lower or equal to 3);
iii. Remove the selected node and its adjacent edges and update the weights of the rest of the nodes.

Repeat this operation until there are no nodes. The $\boldsymbol{C_i}$ are the cliques of the junction tree.

Construction of an optimal tree:

i. For each pair of clique $\boldsymbol{X}$ and $\boldsymbol{Y}$, create a separator $\boldsymbol{S_{XY}}$ equal to $\boldsymbol{X} \cap \boldsymbol{Y}$ (we will have $\boldsymbol{n-1}$ separators, where $\boldsymbol{n}$ is the number of cliques).
ii. Select the separator $\boldsymbol{S_{XY}}$ with the greatest weight and inset it between the cliques $\boldsymbol{X}$ and $\boldsymbol{Y}$. Repeat the operation until all the separators will be inserted.
iii. The resulted graph is called "junction tree"

Note: when two or many separators have the same weight, we choose the separator with the smallest cost.

"The cost" of $\boldsymbol{S_{XY}}$ is the weight of $\boldsymbol{X}$ plus the weight of $\boldsymbol{Y}$.

3.1.2. Inference on the junction tree

In this phase, potentials are attributed to the components of the junction tree, then a series of calculation is performed in order to compute the marginal probabilities of the BN nodes. The different steps of this phase are developed below.

i. Initialisation

In this step, we assign potentials for the junction tree by using the probability tables of the BN.

1. For each clique and separator X fix $\emptyset_X(X)$ to 1;
2. For each variable V of the BN; assign to V one clique that contain its family (V and its parents), then multiply $\emptyset_X$ by $P\big(V\big|Pa(V)\big)$;

This step must verify the following equation:

$$\frac{\prod_{i=1}^{N}\emptyset_{Xi}}{\prod_{j=1}^{N-1}\emptyset_{Si}} = P(U) \qquad (2)$$

ii. Global propagation

In this step, we perform an ordered series of local manipulations, called "message passes". The message passes rearrange the junction tree potentials and they become locally consistent; thus, the result of the global propagation is a "consistent" junction tree. This step can be divided into two phases: the collect and distribution phase. In the first phase, the messages are sent from the leaf cliques to the chosen clique. In the second phase, the messages are sent from the chosen clique to the leaf cliques.

1. $\emptyset_{Si}^{*} = \sum_{Ci\setminus Si}\emptyset_{Ci}\,, i = 1, \dots, n$
2. $\emptyset_{C}^{*} = \emptyset_C \prod_{i=1}^{n}\frac{\emptyset_{Si}^{*}}{\emptyset_{Si}}$
3. $\emptyset_{Si} = \emptyset_{Si}^{*}, i = 1, \dots, n$
4. $\emptyset_C = \emptyset_C^{*}$

The junction tree is consistent if the following equation is verified:

$$P(U) = \frac{\sum_i \emptyset_{Ci}}{\prod_j \emptyset_{Sj}} \qquad (3)$$

iii. Marginalisation

Since the junction tree is become consistent, we can now compute the marginal probability $P(V)$ of each node of the BN as the following:

1. We define a clique or separator that contain the node V;
2. We compute $P(V)$ by marginalising $\emptyset_X$ as the following equation:

$$P(V) = \sum_{X\setminus\{V\}}\emptyset_X \qquad (4)$$

4. METHODOLOGY AND APPLICATION

In this part, we present a methodology that aims to evaluate the availability of complex systems using Bayesian networks. This methodology is applied on a real system (the water distribution system of Bejaia city).

4.1 Construction of the Bayesian network

To model the studied system as a BN, three kinds of data has been used; the arrangement of pumping stations scheme of the system and the expert advice for the creation of the BN structure. We have also used the statistical data and the expert judgment for the creation of the probability tables of the BN. On the fig1 and fig2 are represented the water distribution system and its corresponding BN.

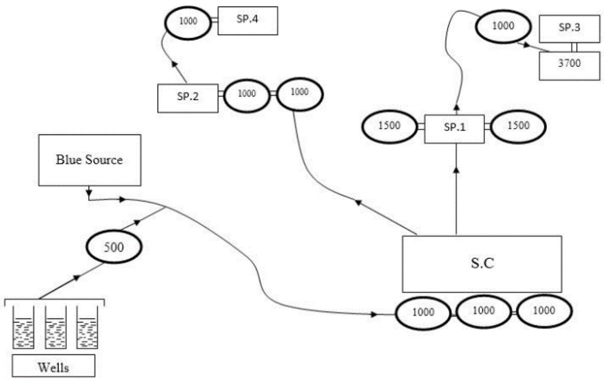

Figure 1: Water distribution system scheme.

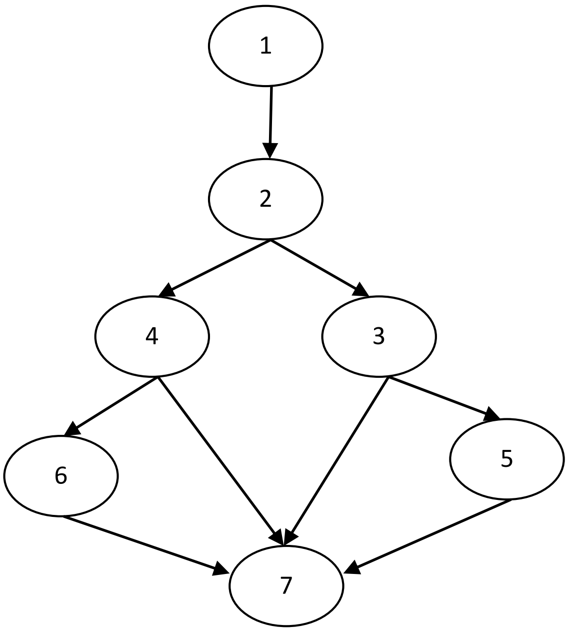

Figure 2: The Bayesian network corresponding to the figure 1.

One node in the BN can include several components (water storage tank, pumps, pipes …). The node "1" represents the source station, the node "2" is the central pumping station, the nodes "3, 4, 5, 6" represent the secondary pumping stations which are linked to the node "7" which represent the client.

4.2 Probabilities assessment

The Bayesian theory is essentially based on the mathematical concept of probability. In this paper, we talk about the capability of a system to be in the state that permit to perform a required function under a given conditions (the availability).

$$\boldsymbol{A} = \frac{\boldsymbol{available\ time}}{\boldsymbol{available\ time + unavailable\ time}} \quad (5)$$

For a BN, we have to establish a probability tables for each node. The prior probability tables for the root nodes and the conditional probability tables for the other nodes. For the prior probability tables, we can use directly the relation (5). So, the probability table of a given root node ''x'' is as follow:

Table 1: *Prior probability table of a root node ''x''*

x	0	1
	$1 - A$	A

Where 0 and 1 represent the component state: 1 for the operation state and 0 for the failure state.

For the other BN nodes, the conditional probability concept will be applied. So, the availability of a given node has to be evaluated by knowing the state of its parent nodes. For that, a new concept is introduced herein: the factor of availability reduction due to the failures and the factor of availability reduction due to the PM actions.

- **Availability reduction factors**

The complex systems are subjected to various kinds of failure. By considering the causes of these failures, it is usually found that most of them are caused by the failure of another component or sub-system. So, it becomes very important to take in consideration this observation to compute the availability. For this reason, we have introduced a new concept; the availability reduction factor for the creation of the conditional probability tables. This factor can be defined as the proportion of availability of a given node (component or sub-system) affected by the failure of its parent nodes and not by its own failure or the failure of one of its components. This factor is computed from the historical data of the maintenance actions and the PM plan.

For a node "x" knowing that "y" is one of its parent nodes, the availability reduction factor is given by:

$$\boldsymbol{P_{x|y}} = \boldsymbol{1} - \frac{\boldsymbol{TID_{x|y}}}{\boldsymbol{T - TID_x}} \quad (6)$$

Whit: T is the inspection period, TID_x is the unavailable time of the node ''*x*'' caused by its failure and $TID_{x\backslash y}$ is the unavailable time of the node ''*x*'' caused by the failures of its parent node ''*y*''.

For the studied system, the availability factors of the BN nodes are computed from the historical data of the different pumping stations of the system. The availability reduction factors, caused by failures, of the BN nodes are in the table 2.

Table 2: *Availability reduction factors caused by failures*

Node	Factor	Value
2	$P_{2\|1}$	0,9324
3	$P_{3\|2}$	0,954
4	$P_{4\|2}$	0,9532
5	$P_{5\|3}$	0,9596
6	$P_{6\|4}$	0,9118
7	$P_{7\|3}$	0,9608
	$P_{7\|4}$	0,9149
	$P_{7\|5}$	0,9593
	$P_{7\|6}$	0,9254

So, the conditional probability tables of a given node are as the following:

Table 3: *Conditional probability table of a node "x" knowing the state of its parent node "y"*

	node *x*	
y	**0**	**1**
0	$P_{x\|y}(1-A)$	$P_{x\|y}A$
1	$1-A$	A

4.3 Conditional probability tables of the studied system

The conditional probability tables of the nodes of the studied system are as the following:

Table 4: *Conditional probability tables of the studied system*

node 1

0	1
0,0116	0.9884

node 2

node 1	0	1
0	0,1216	0,8784
1	0,0579	0,9421

node 3

node 2	0	1
0	0,118	0,882
1	0,0754	0,9246

node 4

node 2	0	1
0	0,1246	0,8754
1	0,0817	0,9183

node 5

node 3	0	1
0	0,083	0,917
1	0,0444	0,9556

node 6

node 4	0	1
0	0,1247	0,8753
1	0,04	0,96

node 7

node 3	node 4	node 5	node 6	0	1
0	0	0	0	0,2587	0,7413
0	0	0	1	0.1989	0,8011
0	0	1	0	0,2272	0,7728
0	0	1	1	0,1649	0,8351
0	1	0	0	0,1897	0,8103
0	1	0	1	0,1244	0,8756
0	1	1	0	0,1553	0,8447
0	1	1	1	0,0872	0,9128
1	0	0	0	0,2284	0,7716
1	0	0	1	0,1662	0,8338
1	0	1	0	0,1957	0,8043
1	0	1	1	0,1308	0,8692
1	1	0	0	0,1567	0,8433
1	1	0	1	0,0887	0,9113
1	1	1	0	0,1209	0,8791
1	1	1	1	0,05	0,95

4.4 Application results

For the availability evaluation of the studied system, we have opted for the Bayesian inference by using the junction tree algorithm of Jensen. We have programmed this algorithm on the mathematical computing software MATLAB by using BNT tools. From this algorithm, we have computed the marginal probability of the node "7" which represents the client node. This probability represents the availability of the client node. It represent also the availability of the studied system, it is computed by taking into account the different interactions and links between the nodes (pumping stations) of the water distribution system of Bejaia city. This availability is equal to 0.9352.

5. CONCLUSION

In this paper we have proposed an original methodology, based on the availability factor caused by the maintenance actions, for the evaluation of the availability of the complex systems by using the Bayesian networks. Thanks to the Bayesian networks, we are able to model the real complex systems by including the different links and causalities that can exist between the system components. The Bayesian inference permit the computation of the marginal probability of a given node, which represent the availability of the system in our case, and always by taking into account the links and interaction of the system components. This methodology is applied on a real system, the drinkable water distribution system of Bejaia city.

As prospect, we plan to include this methodology in a maintenance cost model in order to optimize the maintenance of complex systems

6. REFERENCES

[1] Ait Mokhtar, E. Laggoune, R. Chateauneuf, A. (2013) Evaluation et optimisation des performances des systèmes complexes par les réseaux bayésiens. International Workshop on « Evaluation de Performance et Qualité de Service » (EPQOS'2013), Bejaia, Mai 2013, 257–262. LAMOS Editions.

[2] Simon C, Weber P, Evsukoff A. (2008) Bayesian networks inference algorithm to implement Dempster Shafer theory in reliability analysis. Reliability Engineering and System Safety, 93, 950–963.

[3] Marquez D, Neil M, Fenton N. (2010) Improved reliability modeling using Bayesian networks and dynamic discretization. Reliability Engineering and System Safety, 95, 412–425.

[4] Weber P, Jouffe L. (2006) Complex system reliability modelling with Dynamic Object Oriented Bayesian Networks (DOOBN). Reliability Engineering and System Safety, 91, 149–162.

[5] Wilson AG, Huzurbazar AV. (2007) Bayesian networks for multilevel system reliability. Reliability Engineering and System Safety, 92, 1413–1420.

[6] Bobbio A, Portinale L, Minichino M, Ciancamerla E. (2001) Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. Reliability Engineering and System Safety, 71, 249–260.

[7] Langseth, H. Portinale, L. (2007) Bayesian networks in reliability. Reliability Engineering and System Safety, 92, 92–108.

[8] Amari SVD, J.B.; Misra, R.B. (1999) A separable method for incorporating imperfect fault-coverage into combinatorial models. IEEE Transactions on Reliability, 48, 267–274.

[9] Weber, P. Medina-Oliva, G. Simon, C. Iung, B. (2012) Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. Engineering Applications of Artificial Intelligence, 25, 671–682.

[10] De Souza, E. Ochoa, P.M. (1992) State space exploration in Markov models. Performance Evaluation Review, 20 (1), 152–166.

[11] Weber, P. Jouffe, L. (2003) Reliability modeling with dynamic Bayesian networks. Reliability Engineering and System Safety, 91 (2), 149–162.

[12] Naïm, P. Wuillemin, P.H. Leray, P. Pourret, O. Becker, A. (2007) Réseaux bayésiens. Edition Eyrolles, Paris, France.

[13] Alyson, G. Wilson, Aparna, V. Huzurbazar. (2007) Bayesian networks for multilevel system reliability. Reliability Engineering and System Safety, 92, 1413–1420.

[14] Pearl, J. (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers Inc, San Francisco, USA.

[15] Jensen, F. Lauritzen, S. Olesen, K. (1990) Bayesian updating in recursive graphical models by local computations. Computational Statistical Quaterly, 4, 269–282.

[16] Smail, L. (2004) Algorithmique pour les Réseaux Bayésiens et leurs Extensions. Thèse de Doctorat. Université de Marne-La-Vallée.

[17] Langseth, H. Thomas, D. Nielsen. Rumi, R. Salmeron, A. (2009) Inference in hybrid Bayesian networks. Reliability Engineering and System Safety, 94, 1499–1509.