

A Causal Dependencies Identification and Modelling Approach for Redesign Process

Thierno M. L. Diallo¹, Marc Zolghadri

Quartz Laboratory, Supmeca, 3, rue Fernand Hainaut, 93407 Saint-Ouen, France

thierno.diallo@supmeca.fr; marc.zolghadri@supmeca.fr

Abstract. Systems and products are changed throughout their lifecycle to adapt to users' needs changes or to technological advances, among other reasons. The redesign process consists in modifying one or several parameters to reach the awaited redesign targets (better performance for instance). However, due to dependencies among parameters, changing one parameter may have unintended impacts on others. The problem we study in the redesign process concerns its underlying process of change propagation through the so called dependency model. The dependencies among parameters are correlation or causal. As a first contribution, the paper argues that the most interesting links to identify, model and work on are causalities. Therefore, the challenge to overcome is to identify the existing causal links among parameters using data exploration or expert knowledge mappings. The second contribution discusses a Causal dependencies identification and modelling approach for Redesign process, CaRe in short, which uses the Bayesian Network theory. CaRe is made to generate a causal Bayesian Network that allows evidential and causal inferences, supporting redesign decision-makings. The steps of CaRe are discussed in detail and future research works are presented at the end of the paper.

Keywords: redesign, dependency model, change propagation, causality, Bayesian networks

1 Introduction

This research deals with change propagation during the redesign process. Throughout its lifecycle, a system or product undergoes many redesign or renewal and upgrade operations [1]. Redesign or upgrade consists in partial or total redesign of the system triggered by causes such as user needs, obsolescence of some functions or components, etc., see [2, 3].

A system consists of a set of physical and functional interdependent elements. Furthermore, the system and its environment influence on each other. Changes propagation is the fact that an engineering change to one component (physical or functional) of the

¹ Corresponding author at: Supméca - Institut supérieur de mécanique de Paris
3 rue Fernand Hainaut, 93400 Saint-Ouen, France

system or its environment results in one or more additional changes to other components, when those changes would not otherwise have been required or observed [4, 5].

The redesign consists of changing one or more of system parameters. However, due to their (explicit or implicit) dependencies, changing one parameter would generate others. The prediction and management of the impacts of these undesired changes represent a great challenge for complex systems [2]. The aim of this research work is to contribute to the better identification of these impacts through identifying, appropriately modelling and exploiting causal dependencies among the system parameters. Having a better understanding of the possible (wanted and unwanted) consequences allows a better (re-)design decision-making. The unwanted consequences should be controlled far in advance before any real change in the system; they do not have to violate the system constraints.

As far as we were able to find out, in the engineering change management literature, no study does focus on the very nature of dependency links. This paper studies both causal and correlation dependency links used in a redesign process as a fundamental element of the dependency model. Nevertheless, the main target is put on the causal dependencies that provide more capabilities to designers.

To obtain such causal dependency model, we propose an approach based on the system architecture. The system architecture is defined as a physical structure, the functional structure and the mapping between these two structures [6]. Determining by this way, the architecture facilitates the identification of the causal dependencies model. In some cases, the correlation links may persist in the model depending on the quality of used data or expert knowledge. Once, in hands, such causal dependencies model, if modelled by an appropriate dependency language, may be used for redesign decision-makings. In our research we use the Bayesian Network as the dependency theory for modelling, mainly due to the various reasoning possibilities it offers.

The remainder of this paper is organized as follows. First, some main change management concepts and change propagation methods are presented in Section 2. This section will end up with a summary conclusion about the dependency modelling and engineering change management. The main motivations of our research are explained there. The Bayesian approach is then explained very briefly to allow the understanding of the approach (Section 3). In section 4, we present in detail the proposed approach, called “Causal dependencies identification and modelling approach for Redesign process”, “CaRe” in short. The paper ends with a summary of our findings and discusses the future works in section 5.

2 Literature review

An engineering change is any change or modification made in the form, fit, material, dimension and function of a product or a component [7]. Change management is defined by [8] as “the comprehensive evaluation and approval or disapproval of a change that takes into consideration all effects of the change”. Change management is studied by lots of authors, see [2-5]. The change propagation is studied using various tools. For instance, Clarkson and his colleagues [2, 4, 5] use DSM (Design Structure Matrix) and

other similar matrix-based tools (MDM-Multiple-Domain Matrix, DMM-Domain Mapping Matrix, HoQ - House of Quality). However, no DSM-derived approach studies the nature of the links that could be correlation and causality.

Other proposed methods are specialized for geometrical change propagation. Cohen et al. [3] propose the C-FAR (Change Favourable Representation) methodology to trace and predict change propagation in computer-aided mechanical design. Most of the CAD (Computer-Aided Design) tools has a kind of change propagation features. Their main drawback is the fact that they propagate only a specific type de change.

The relationships between parameters may be a correlation or causality. However, "Correlation does not imply causation" [9]. When there is a correlation link between 2 parameters, this means that if a value of one of the parameters is observed, it is also expected to observe a given value of the second. The existence of causal link between 2 parameters implies that the modification of the cause will have an impact on the effect.

Correlation relationships express what we know or believe about the world, while causal relationships describe physical constraints. Correlation relationships characterize static conditions while the causal analysis deals with dynamic situations [10].

Relying on these distinctions, we postulate that the most interesting links to identify for changes propagation during the redesign process are causal. The redesign process involves actions or interventions on the system (change of the architecture or the functions, replacement of a component, etc.). Intervention (surgery on mechanisms) is one of the fundamental characteristics of causality [11]. The identification of the causality makes it possible to predict the consequences on the effects while acting on their cause parameter. Therefore, the first challenge is to identify the existing causal link among parameters. The research we presented here is focused on determination of causalities and assess their consequences during the redesign or upgrade process.

3 main approaches have been proposed to model causal relationships and to make causal inferences namely: the Neyman–Rubin causal model, the Structural Equation Model and Directed Acyclic Graphs also called Bayesian Networks (BN). It has been demonstrated that BN is the most commonly used tool to represent causal relationships and to make causal inferences [11].

3 Bayesian Network

A Bayesian Network (BN) is a Directed Acyclic Graph (DAG) represented by the pair (V, E) where V is a set of vertices and E a set of directed edges connecting vertices [12]. A marginal or conditional probability distribution table is associated to each vertex. The set of probabilities of the network, also called the network parameters, is denoted by P . The couple (G, P) , with $G = (V, E)$ a DAG, is a Bayesian network if it satisfies the Markov condition.

Edges in a BN do not necessary represent causal relationships among variables [11]. There are causal and belief (non-causal) networks. Non-causal networks express only the conditional independences between the nodes of the graph induced by the Bayes'

theorem and the Markov condition. Causal networks, in addition to independence relations, express cause-effect relationships between the nodes.

3.1 Causal discovery from observational data

The BN can be obtained thanks to expert knowledge. But, this becomes hard or unfeasible for large systems for practical reasons. However, it is also possible to make causal inference from the observation data [11, 13, 14]. This possibility is of great interest for complex systems. Performing causal inference based on observational data requires some assumptions [14, 15]:

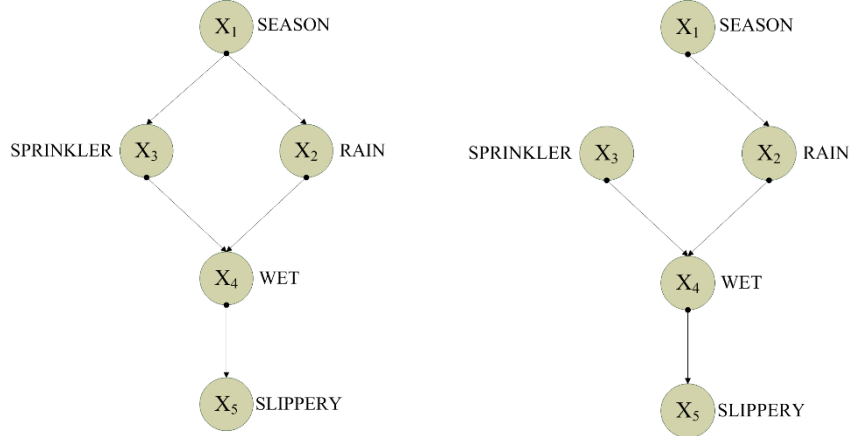
- We must assert explicit causal assumptions about the process that generated the observed data.
- Compliance with Causal Markov Condition. To fulfill Causal Markov Condition :
 - There must be no hidden common causes. No two variables in the set of observed variables V can have a hidden common cause, or, if they do, it must have the same unknown value for every unit in the population under consideration.
 - Absence of selection bias which is dependency created between two variables having a common effect when this common effect is instantiated.
 - Absence of causal feedback loops: if X has a causal influence on Y , Y does not have a causal influence X .

Several causal BN construction algorithms are proposed in the literature. This involves either constraint-based learning algorithm (e.g. PC algorithm [13]) or hybrid learning algorithms such as Min-Max Hill Climbing algorithm [16]. Since most of these construction algorithms are heuristics, their use often requires the validation by experts of the constructed network.

3.2 Evidential reasoning and causal reasoning

The difference between a causal and a non-causal Bayesian network can also be observed in the inference process. To illustrate the difference between evidential (observational) and causal reasoning, let us consider the causal Bayesian network represented in Figure 1 [17]. The full joint distribution of the left-side BN (Figure 1 (a)) is:

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1) \cdot P(X_2 | X_1) \cdot P(X_3 | X_1) \cdot P(X_4 | X_2, X_3) \cdot P(X_5 | X_4) \quad (1)$$



(a) BN representing causal relationship among season (X1), it is raining or not (X2), the sprinkler is on or not (X3), the pavement is wet or not (X4), and the pavement is slippery or not (X5) (b) Modified Bayesian network (removal of the causal link between the season X1 and the sprinkler X3) following an intervention on the sprinkler (turned on)

Fig. 1. Evidential and causal reasoning (adapted from [17]).

Evidential reasoning.

From the graph presented in Fig.1 and the joint distribution probability, it is possible to determine any conditional probability given some evidences. For example, the probability that the sprinkler is on, after observing that the pavement is slippery, is:

$$\begin{aligned}
 P(X_3=\text{on}|X_5=\text{true}) &= \frac{P(X_3=\text{on}, X_5=\text{true})}{P(X_5=\text{true})} \quad (2) \\
 &= \frac{\sum_{X_1, X_2, X_4} P(X_1, X_2, X_3=\text{on}, X_4, X_5=\text{true})}{\sum_{X_1, X_2, X_3, X_4} P(X_1, X_2, X_3, X_4, X_5=\text{true})} \\
 &= \frac{\sum_{X_1, X_2, X_4} P(X_1)P(X_2|X_1)P(X_3 = \text{on}|X_1)P(X_4|X_2, X_3 = \text{on})P(X_5 = \text{true}|X_4)}{\sum_{X_1, X_2, X_3, X_4} P(X_1)P(X_2|X_1)P(X_3|X_1)P(X_4|X_2, X_3)P(X_5 = \text{true}|X_4)}
 \end{aligned}$$

Causal reasoning.

The causal reasoning is the reasoning process that takes place when a certain intervention occurs on one of the model parameters. For example, what happens if one turn on the sprinkler? Contrary to the observation phenomenon, the intervention will have two effects: 1) modification of the value of the parameter that undergoes the action ($X_3 = \text{on}$) and 2) modification of the structure of the graph (removal of the causal link between the season X_1 and the sprinkler X_3 (see Figure 1 (b)). The joint probability distribution becomes:

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1) \cdot P(X_2|X_1) \cdot P(X_4|X_2, X_3=\text{on}) \cdot P(X_5|X_4) \quad (3)$$

4 Proposed approach for change propagation analysis model construction

The causal dependencies identification and modelling approach for redesign process, CaRe, is based on the determination of causal links among the system parameters. We use the BN formalism to model the causal dependencies exploiting expert knowledge and data exploitation. The CaRe approach is composed of five steps (see Figure 2):

- Step 1. Definition of the system boundary
- Step 2. System architecture identification
- Step 3. Determination of interfaces and exchanges between components via these interfaces
- Step 4. Quantification of exchanges between components and construction of the causal model
- Step 5. Exploitation of the change propagation model

4.1 Step 1 - Definition of the system boundary

The goal is to identify the system elements to consider. In particular, one must ensure that the causal Markov condition is fulfilled, that is, there are no hidden common causes (left-side situation in Fig.2) or selection bias or causal feedback loops (right-side situation in Fig.2). This step is critical for the reliability of the model. Its realization requires the participation of experts in the different areas covered by the system parameters.

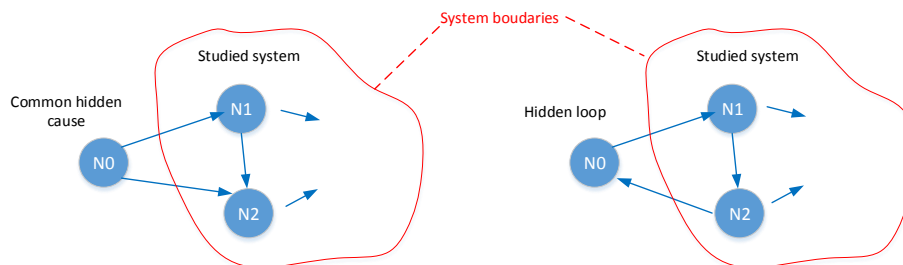


Fig. 2. System boundary

4.2 Step 2 - System architecture identification

Product architecture is the scheme by which the function of a product is allocated to physical components [6]. Here, we define first the functions of the system. The system physical components are then identified. Finally, their mapping is established. This step allows to determine components involved in the realization of each of the functions. Functions should be characterized by quantitative indicators.

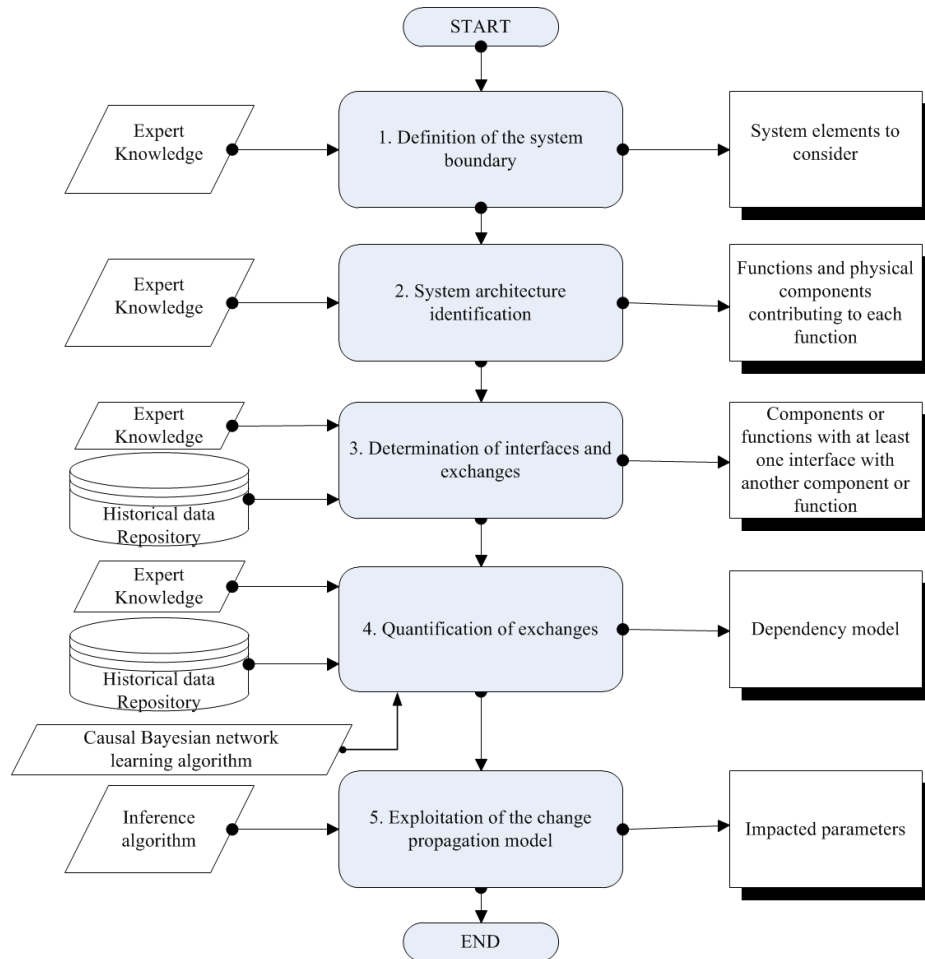


Fig. 3. CaRe: the proposed approach for the causal dependency modelling

4.3 Step 3 - Determination of interfaces and exchanges between components and functions via these interfaces

We then determine the relationship among components and functions. Two components or functions are linked when they share an interface. Four types of exchanges can be done through an interface [18]: Material, Information, Spatial, Energy (MISE in short) exchanges. The interfaces define the change propagation channels (see Fig.3).

It is also necessary to identify the exchanges between the functions of the system using functional DSM. This step allows to select the relevant parameters for the rest of the approach. Indeed, only components or functions with at least one interface with another component or function will be retained.

The components and their MISE exchanges, the functional model and their interdependencies, and finally the components-functions mapping allow to complete the architecture of the system.

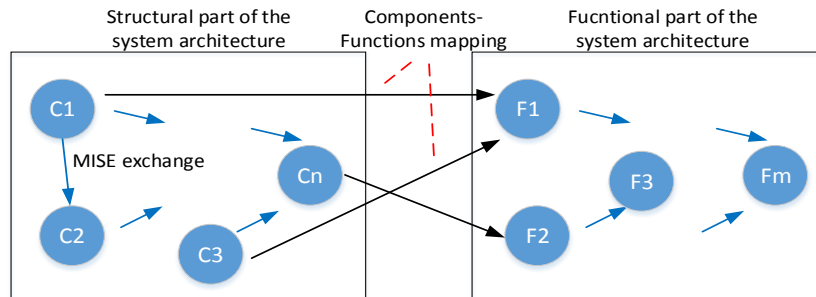


Fig. 4. Architecture and exchanges among components and functions and their mapping

4.4 Step 4 - Quantification of exchanges between components and construction of the causal model

The goal is to quantitatively assess the identified links and to build the causal model by exploiting the data representing the exchanges between functions/components.

Data collection.

In order to facilitate understanding of the data collection process, we consider the trailer exposed in Fig. 2 of [6]. This trailer is composed of 6 components (Box, Hitch, Fairing, Bed, etc.) and realize 6 functions (“Protect cargo from weather”, “Connect to vehicle”, “Minimize air drag”, “Support cargo load”, etc.). All parameters characterizing a function or a component or shared with other function/component are identified:

- Function: $F_x_MISE_xy$ (Parameter y characterizing or shared by function x).
For example, “Connect to vehicle” function (identified by x) might be characterized by the dimensions of the contact surface and the maximum force it is expected to sustain. Each of these parameters will be identified by a specific y (y_1 and y_2).
- Component: $C_x_MISE_xy$ (Parameter y characterizing or shared by component x).
We proceed in the same way as previously replacing functions by components.

Observed values of each parameter are then collected for different periods of time. The historical data table takes the form of Table 1.

	Functions						Components					
	F1			F2			C1			C2		
Exchanged parameters	F1 MISE 11	F1 MISE 12	...	F2 MISE 21	F2 MISE 22	...	C1 MISE 11	C2 MISE 12	...	C2 MISE 21	C2 MISE 22	...
Observed values												

Table 1. Historical data for causal model construction

Causal model building.

The causal Bayesian network representing causal relationships among parameters is learned from historical data using a causal Bayesian network learning algorithm. Building a Bayesian model requires the determination of its structure (vertices and edges) and the marginal and conditional probabilities distributions. But first, a choice has to be made on how to model vertices or nodes. Two choices may be made:

- A node represent the state (changed or not) of the corresponding parameter following a change. In this case, nodes are discrete (binary random variable) and the network parameters are conditional probability tables. In this configuration, it is relatively easy to convert a likelihood DSM into a Bayesian model, see [19]. However, this presupposes that a certain number of conditions are met, including: 1) Being able to define the states (changed or not) of each parameter; and 2) The possibility to an expert to quantifies all change propagation probability from one component to the other.
- A node represent the value of the correspond parameter. If we suppose that functions and components parameter are continuous, then nodes are continuous too. In most cases, these parameters are also expected to follow a normal distribution. In this case, root nodes parameters are marginal distributions and the other nodes parameters are expressed as a Gaussian linear function of parent nodes [20]. This approach requires only sufficient data. As parameters are supposed to follow a normal distribution, one can define a normal fluctuation interval $[\mu \pm k\sigma]$ for each parameter, μ being the mean and σ standard deviation.

We recommend to proceed by the second modeling approach. Considered parameters are supposed to be continuous and the researched BN is a Gaussian BN. Each node represents a random variable corresponding to the value of the corresponding parameter. The structure of the causal Bayesian network representing the change propagation model can then be learned using one constraint-based [13] or hybrid learning algorithm [16]. In addition to historical data, most of these learning algorithms allow to exploit expertise in order to reduce the learning process and/or to improve the quality of the built network. The result of this learning process is a network with nodes representing value of functions and components parameters and arcs corresponding to the causal link existing between the connected nodes.

4.5 Step 5 - Exploitation of the change propagation model

For this step, an inference process taking into account the causal aspect of the network is defined (cf. Section 3.2. Evidential reasoning and causal reasoning). For a given change made on a parameter, the inference algorithm update the parameters' value and identify all other impacted parameters following this change. These impacted parameters are identified by the calculation of the conditional probabilities after having modified the structure of the network and updated the value of the modified parameter (see example in Section 3.2.). Those parameters for which the updated value is outside the normal fluctuation interval $[\mu \pm k\sigma]$ are considered to be the impacted parameters. It is also possible to quantify this impact by calculating the difference between the updated value and the initial value.

5 Conclusion and future work

Existing systems are more and more subject to redesign process because of the rapid advances in technology and the constantly evolution of users' needs, among other reasons. Changes propagation management and the control of their consequences on the system functionalities and performances are a major challenge in many industrial sectors.

Change propagates along existing links between components or parameters of the system being redesigned. In this research, we analyze the two types of links that may exist among components or parameters, namely correlation and causality relationships. Based on this analysis, we postulate that the privileged changes propagation links during the redesign process are the causal links. We then review main existing methods for causal relationships identification and causal inferences. The findings suggest that Bayesian Networks are the most appropriate framework to represent causal relationships and to make causal inferences. An approach for change propagation analysis model construction, CaRe (for Causal dependencies identification and modelling approach for Redesign process) is proposed. This approach is based on Bayesian Network theory and uses DSM (Design Structure Matrix).

Some assumptions made in our approach require further research. All the parameters necessary to build the model are supposed known and their historical data available. However, this is not always the case for real system. Therefore, incomplete and missing data should be considered in the learning process. Parameter are supposed continuous variables and the researched BN is a Gaussian Bayesian network. Hybrid Bayesian network containing both discrete and continuous variables could be the most flexible solution for complex systems.

References

1. Zolghadri, M., et al., *Complex systems renewal: positioning, concepts and architectural issues*. IFAC Proceedings Volumes, 2014. **47**(3): p. 8731-8736.<http://dx.doi.org/10.3182/20140824-6-ZA-1003.01771>
2. Clarkson, P.J., C. Simons, and C. Eckert, *Predicting change propagation in complex design*. Journal of Mechanical Design(Transactions of the ASME), 2004. **126**(5): p. 788-797.[doi:10.1115/1.1765117](https://doi.org/10.1115/1.1765117)
3. Cohen, T., S.B. Navathe, and R.E. Fulton, *C-FAR, change favorable representation*. Computer-aided design, 2000. **32**(5): p. 321-338.[https://doi.org/10.1016/S0010-4485\(00\)00015-4](https://doi.org/10.1016/S0010-4485(00)00015-4)

4. Koh, E.C., N.H. Caldwell, and P.J. Clarkson, *A method to assess the effects of engineering change propagation*. Research in Engineering Design, 2012. **23**(4): p. 329-351.<https://doi.org/10.1007/s00163-012-0131-3>
5. Giffin, M., et al., *Change propagation analysis in complex technical systems*. Journal of Mechanical Design, 2009. **131**(8): p. 081001.DOI: 10.1115/1.3149847
6. Ulrich, K., *The role of product architecture in the manufacturing firm*. Research policy, 1995. **24**(3): p. 419-440.[https://doi.org/10.1016/0048-7333\(94\)00775-3](https://doi.org/10.1016/0048-7333(94)00775-3)
7. G.Q., H. and M. K.L., *Current practices of engineering change management in UK manufacturing industries*. International Journal of Operations & Production Management, 1999. **19**(1): p. 21-37.<https://doi.org/10.1108/01443579910244205>
8. Mooz, H., K. Forsberg, and H. Cotterman, *Communicating project management: the integrated vocabulary of project management and systems engineering*. 2003: John Wiley & Sons,
9. Aldrich, J., *Correlations genuine and spurious in Pearson and Yule*. Statistical science, 1995: p. 364-376,
10. Holland, P.W., *Statistics and Causal Inference*. Journal of the American Statistical Association, 1986. **81**(396): p. 945-960.[10.2307/2289064](https://doi.org/10.2307/2289064)
11. Pearl, J., *Causality*. 2009: Cambridge university press,
12. Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1988: Morgan Kaufmann Publishers, <http://books.google.fr/books?id=k9VsqN24pNYC>
13. Spirtes, P., C.N. Glymour, and R. Scheines, *Causation, Prediction, and Search*. 2000: MIT Press, <http://books.google.fr/books?id=vV-U09kCdRwC>
14. Neapolitan, R.E., *Learning Bayesian Networks*. 2003: Prentice-Hall, Inc., https://www.researchgate.net/profile/Richard_Neapolitan/publication/235410247_Learning_Bayesian_Networks/links/02e7e52dcf3617aa9c000000.pdf
15. Conrady, S. and L. Jouffe, *Bayesian Networks and BayesiaLab: A Practical Introduction for Researchers*. 2015: Bayesia USA, <https://books.google.fr/books?id=etXXsgEACAAJ>
16. Tsamardinos, I., L. Brown, and C. Aliferis, *The max-min hill-climbing Bayesian network structure learning algorithm*. Machine Learning, 2006. **65**(1): p. 31-78.[10.1007/s10994-006-6889-7](https://doi.org/10.1007/s10994-006-6889-7)
17. Pearl, J., *Bayesian networks*. Department of Statistics, UCLA, 2011, <https://escholarship.org/uc/item/53n4f34m>
18. Engel, A. and Y. Reich, *Advancing architecture options theory: six industrial case studies*. Systems Engineering, 2015. **18**(4): p. 396-414.<https://doi.org/10.1002/sys.21312>
19. Lee, J. and Y.S. Hong, *Bayesian network approach to change propagation analysis*. Research in Engineering Design, 2017. **28**(4): p. 437-455.[10.1007/s00163-017-0252-9](https://doi.org/10.1007/s00163-017-0252-9)
20. Scutari, M. and J.B. Denis, *Bayesian Networks: With Examples in R*. 2014: Taylor & Francis, <http://books.google.fr/books?id=v-XMAwAAQBAJ>