KNOWLEDGE REPRESENTATION OF RISK AND OPPORTUNITIES ARISING FROM ENVIRONMENTAL, SOCIAL AND GOVERNANCE ISSUES: A BAYESIAN NETWORK APPROACH.

Renaud MacGilchrist – Irena Descubes – François Lépineux

Abstract

The very nature of Corporate Social Responsibility (CSR) has brought about the need for common tools which can be used to represent, test and synthesize knowledge and beliefs from diverse and amorphous origins. This paper investigates the use of Bayesian networks for knowledge representation of Risk and Opportunities Arising from Environmental, Social and Governance issues. It is based on an example that uses the data from MSCI, which provides research, ratings and analysis of corporate management of environmental and social risk factors on over 5,000 global companies. Bayesian networks are often used for eliciting, confronting and synthesizing both expert and layman knowledge. They enable us to update our beliefs and expectations regarding complex systems in a rigorous way given new or conjectured evidence and they provide a practical decision support system for elaborating strategies and policy making in the area of the corporate social responsibility.

Key words: Corporate Social Responsibility, Bayesian networks, risks and opportunities

JEL Code: M14, C39, Z13

Introduction

This paper investigates the potential use of Bayesian networks for modelling the knowledge of risk and opportunities arising from environmental, social and governance issues. It is exploits data from MSCI which provides research, ratings and analysis of corporate management of environmental and social risk factors on over 5,000 global companies. The challenge is both formal, in the logical sense, and semantic in terms of dealing with both imprecision and uncertainty. These issues of Causal Calculus were first addressed in the seminal papers by Trygve Haavelmo (Pearl, 2013) and Herbert Simon (1953), but it was Pearl who first provided the formal and logical apparatus for causal calculus and inference under imprecision and uncertainty. Bayesian networks are probabilistic graphical models that can be used to represent belief networks, which are in essence joint probability distributions, based on empirical evidence and/or expert knowledge. These models provide graphical structures for representing knowledge about complex systems in a very efficient and intuitive manner. Each node in the graph generally represents an observable or latent random variable. The relations between the nodes represent probabilistic dependencies, i.e. conditional probability distributions between the linked variables. An essential characteristic of Bayesian networks, when properly designed, is their ability to portray simple conditional independence and causal relationships between variables. Hence instead of dealing with a highly complex interconnected graph each node is only connected to nodes that directly affect or are directly affected by it (please, see the theoretical section below). We note that it is possible to elaborate many possible or matching networks given a set of observable data. However, networks that reflect the underlying causal relationships between variables often provide the simplest and most practical networks. The design and elaboration of causal networks generally necessitates expert knowledge of sorts and cannot be extracted automatically using learning algorithms. Bayesian networks have been used for eliciting, confronting and synthesizing both expert and consumer perception, e.g. Cain (2001) summarized them in a system of concrete steps to capture and represent the world as described by different stakeholders. They also enable us to update our beliefs and expectations regarding complex systems in a rigorous way given new or conjectured evidence and they provide a practical decision support system for elaborating strategies and policy making.

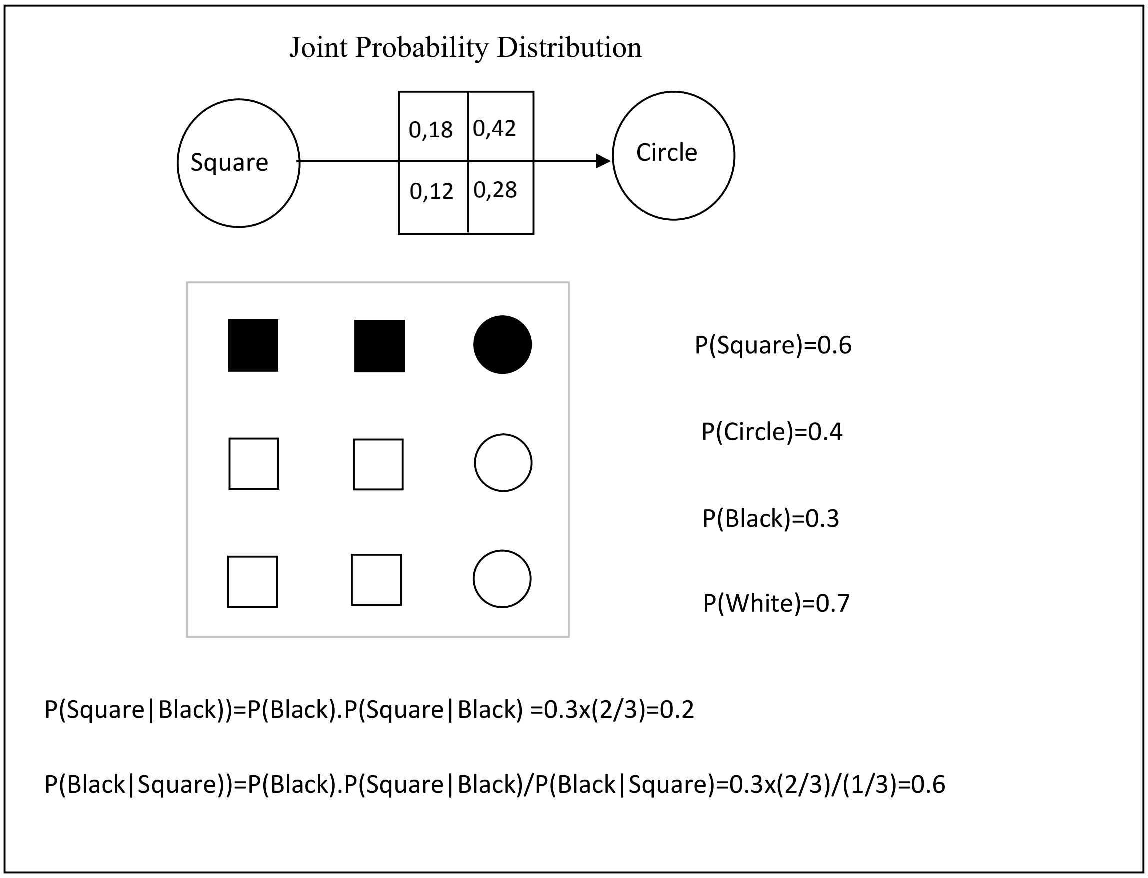
1 Theoretical Background

This section provides a review of the fundamental concepts that underpin Bayesian networks as well as illustrative examples. Bayesian reasoning is based upon Bayes' Rule, which can derive from the basic rules of probability calculus.

$$P(a,b) = P(a \mid b) * P(b) \qquad (1)$$

In Equation 1, P(a,b) is the joint probability of both events a and b occurring, P(a|b) is the conditional probability of event a occurring given that event b occurred, and P(b) is the probability of event b occurring.

Figure 1: Simple Example of Bayesian Reasoning with associated Bayesian Network



Symmetry Equation leads to conditional probability of b given a single piece of information a.

$$P(b \mid a) = \frac{P(a \mid b) * P(b)}{P(a)} \qquad (2)$$

Figure 1 provides a simple example of a Bayesian network with two nodes i.e. random variables Shape:{"Square" and "Circle"} and Colour:{Black, White}) and 9 elements in the sample data. In this example all of the variables are binomial (i.e. equivalent to Boolean). Note, however, that Bayesian networks can easily be extended to variables with multiple states or classes. The directed arcs (arrows) between the nodes generally represent causal relationships between those nodes. The relationship however is not necessarily causal which can lead to Bayesian networks which are more difficult to interpret. The existence of a relationship between two nodes implies the existence of conditional probability distribution between the two variables concerned. When more than n variables are involved the distribution becomes n-dimensional and more complex. Efficient and tractable Bayesian networks generally have more parsimonious relationships that avoid the risk of combinatory explosion.

The absence of a directed arc between two nodes implies conditional independence. Conditional independence of two random variables a and c given b holds when

$$P(a,c \mid b) = P(a \mid b) * P(c \mid b) \qquad (3)$$

Equivalently, we could have said that a and c are conditionally independent given b when c doesn't tell us anything more about a if we already know b:

$$P(a \mid b,c) = P(a \mid b) \qquad (4)$$

Equations (3) and (4) can be shown to be equivalent by using the basic relations of probability (see equations 5.).

$$P(a,b \mid c) = \frac{P(a,b,c)}{P(a)} = \frac{P(a \mid b,c) * P(b,c)}{P(a)} = P(a \mid b,c) * P(b \mid c) \qquad (5)$$

Table 1 and Fig. 2: Example of Conditional Independence and Equivalent JPD provide an example of population with particular characteristics that can be modelled using conditional independence and the corresponding Bayesian network Table 1 synthesizes this population of 30 individuals (which in effect constitute a learning set) with three possible characteristics (Y, R and B) which may or may not be possessed by each individual.

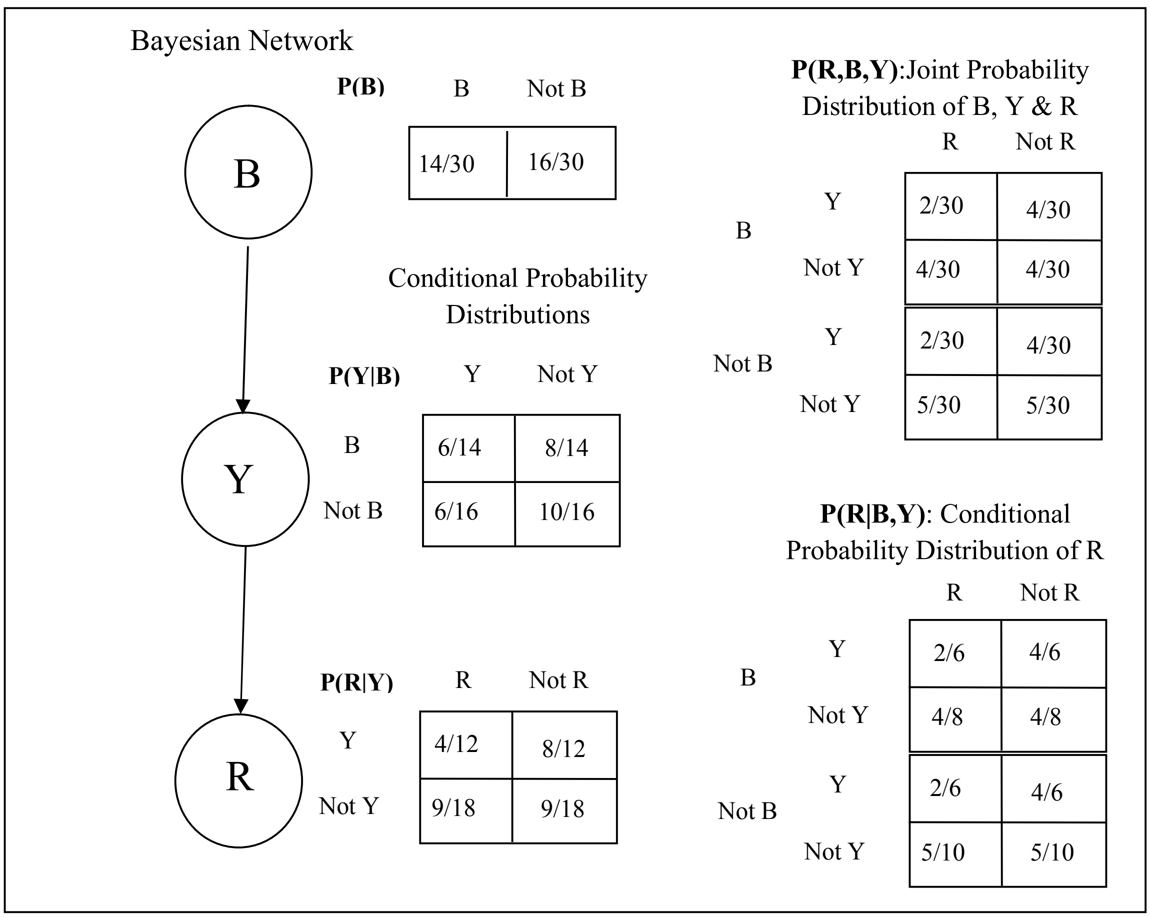
Note that the tabular representation is merely for compactness and convenience and has no particular significance. The existence of conditional independence for both Y and Not Y can be checked using equation 3.

$$P(R,B \mid Y) = P(R \mid Y) * P(B \mid Y) = 2/12 = 4/12 * 6/12$$
$$P(R,B \mid \neg Y) = P(R \mid \neg Y) * P(B \mid \neg Y) = 4/18 = 9/18 * 8/18$$

Table 1: Sample of 30 Individuals with 3 Different Characteristics (Random Variables:Y,B,R)

		,R,	,R,	,R,
	Y, ,	Y,R,	Y,R,	,R,
	Y, ,	Y, ,	Y,R,	,R,
	Y, ,B	Y, ,B	Y,B,R	,B,R
, ,B	Y, ,B	Y, ,B	Y,B,R	,B,R
,B	,B	,B	,B,R	,B,R

Fig. 2: Example of Conditional Independence and Equivalent JPD



The probability distribution of a node can be determined by considering the distributions of its parents. In this way, the joint probability distribution for the entire network can be specified. This relationship can be captured mathematically using the chain rule in Equation 3.

$$P(x) = \prod_{i=1}^{n} P(x_i \mid parents(x_i)) \qquad (6)$$

Equation 6 states that the joint probability distribution for node x is equal to the product of the probability of each component x_i of x given the parents of x_i.

There are several measures that can be used to evaluate the strength of the link between two nodes. A common metric is the measure of conditional mutual information (Meretakis & Wüthrich, 1999).

For any two variables A_i, B_j, conditional mutual information is defined as

$$I(A_i; B_j \mid C) = \sum_{\substack{x \in val(A_i), \\ y \in val(B_j), \\ c_m \in val(C)}} P(x, y, c_m) \log \frac{P(x, y \mid c_m)}{P(x \mid c_m) P(y \mid c_m)} \qquad (7)$$

Intuitively this function measures how well $P(A_i, B_j|C)$ can be approximated *on average* by $P(A_i|C)$ and $P(B_j|C)$.

Entropy is a well-established concept for measuring the amount of information contained in a probability distribution and comes from the work of Shannon on Information Theory. Shannon's entropy for discrete distribution with *n* outcomes with probability p_i is defined as:

$$H(p) = -\sum_x p(x) \ln(p(x)) \qquad (8)$$

This concept can be extended to several distributions. The joint entropy of two random variables is defined as:

$$H(p(x, y)) = -\sum_{x,y} p(x, y) \ln(p(x, y)) \qquad (9)$$

If two random variables are independent their joint entropy is the sum of their two entropies. To show this we recall that the joint distribution of two independent random variables is given by

$$p(x, y) = p(x) p(y). \qquad (10)$$

Using equation 10 we have

$$H(p(x, y)) = -\sum_{x,y} p(x, y) \ln(p(x, y)) = -\sum_{x,y} p(x) p(y) \ln(p(x) p(y)) = \sum_{x,y} p(x) p(y) (\ln(p(x)) + \ln p(y))$$

$$H(p(x, y)) = -\sum_x p(x) \ln(p(x)) \sum_y p(y) - \sum_y p(y) \ln(p(y)) \sum_x p(x) = H(p(x)) + H(p(y)) \qquad (11)$$

The concept of entropy can be extended to that of relative entropy and which can be regarded as a measure of the distance between distributions. If two distributions are defined in exactly the same outcome space then distance between two distributions is known as Kullback-Leibler (1951) distance or divergence between two distributions is as follows:

$$D_{KL}(p, q) = \sum_x p(x) \ln(p(x) / q(x)) \qquad (12)$$

Although it is called a distance this entity is not symmetric. i.e.

$$D_{KL}(p,q) \neq D_{KL}(q,p) \qquad (13)$$

K-L distance is a fairly intuitive measure of the information that is shared between two factors or variables.

To resume, Bayesian models and the key concept of conditional independence provide graphical and mathematical structures for representing qualitative and quantitative knowledge about complex systems in a very efficient and intuitive manner. In mathematical terms, Bayesian networks enable what would otherwise be a highly complex generally intractable joint probability distribution to be broken down into much simpler and understandable local probability distributions. These structures call upon Bayesian probabilities which can combine "objective" and "subjective" probabilities that is direct observations together with constructs and conjectures.

2 Analysis of the MSCI data base

To investigate the applicability of Bayesian networks in CSR data from MSCI Intangible Value assessment (IVA) data base, was analyzed using the EQ unstructured automated learning algorithm. This database provides research, ratings and analysis of corporate management of environmental and social risk factors on over 5,000 global companies.

The MSCI IVA database provides information and rating regarding three key questions (MSCI, 2013):

• What are the key ESG (Environmental, Social and Governance) risks and opportunities most applicable to each industry sector?

• Do companies have risk management strategies commensurate with the ESG risks they face?

• Do companies have strategies to capture potential opportunities in the ESG space?

Fig. 3: Non Causal Bayesian Network for IVA Rating year 2012

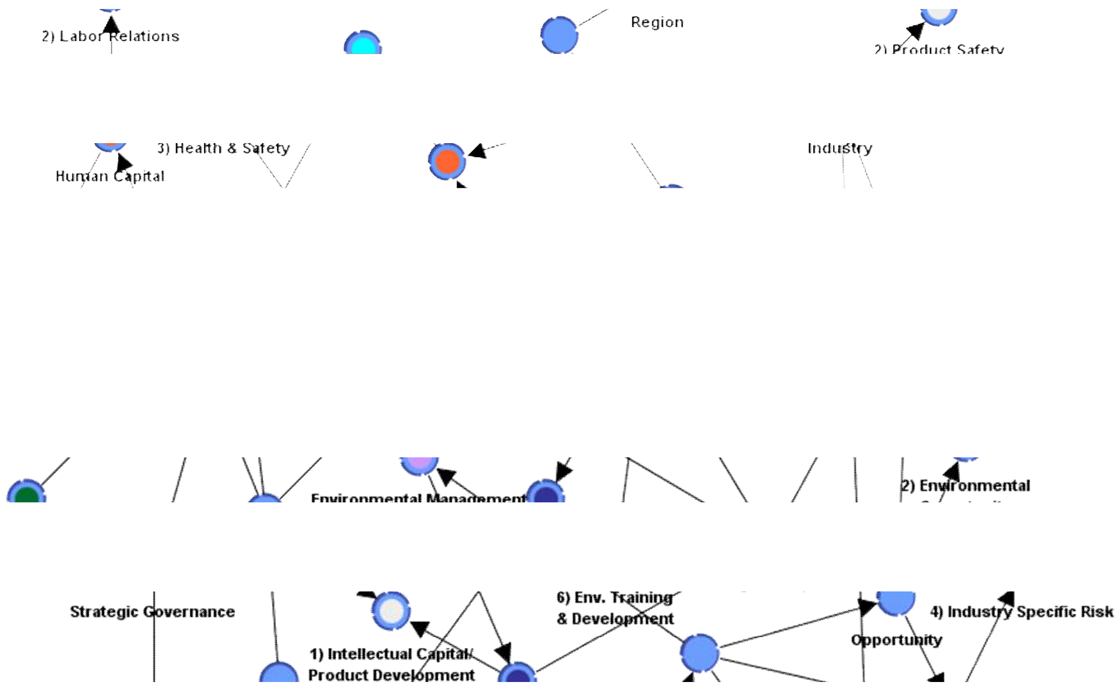


Fig. 3 shows the non-causal Bayesian network extracted, using automated learning, for 5,000 global companies for year 2012. Each node represents a field within the MSCI database. These ratings are arrived at using the MSCI IVA rating methodology which requires the input of many experts as well as the detailed analysis of company behaviour and practices. The colour codes correspond to MSCI groupings of variables, for example the "Sustainable Governance Factor" is divided into the sub factors 1) SG Strategy, 2) Strategic Capability / Adaptability and 3 Traditional Governance Concerns. We note that this network is non-causal; hence the relationships between the nodes only represent conditional probability distributions. In general, many equivalent Bayesian networks may exist for the same joint probability distribution.

This network was extracted using the EQ learning algorithm (Munteanu et al, 2001) which is one of many unsupervised structural learning algorithms for extracting Bayesian networks directly from data. We note that different automated learning techniques do not necessarily lead to the same Bayesian network and that these methods vary in terms of their statistical efficiency and calculation time.

We first illustrate the formal significance of each node by taking the example of the node entitled "Human Rights/ Child and Forced labour" we observe that 6.99% of the observed population have a score below 3.612. Fig. 4 shows the distribution of values of this variable, which has a range of 0 to 10, using four classes. Hence each node corresponds to a histogram or probability distribution.

Fig. 4: A Priori Distribution of Human Rights Variable

We observe that the node "Human Rights and Forced labour" is linked by a relation to "Oppressive Regimes".

Fig. 5 shows the A Priori distributions of the variables Oppressive Regimes and Human Rights, i.e. we suppose no knowledge about the regime. Fig. 6: Impact of Oppressive Regimes on Human Rights, however, shows how these two distributions are related and provide an example of "what if" analysis. We can see for example what the distribution of human rights is for the top 9.94% of individuals with an Oppressive Regime score of less than 4.045 (50.16%, 41.06%, 6.47%,2.32%). The same analysis is conducted for each class. It is important to n ote that it is also possible to examine the impact of any possible change in the values of the distributions.

Fig. 5: A Priori Distributions Oppressive Regimes and Human Rights

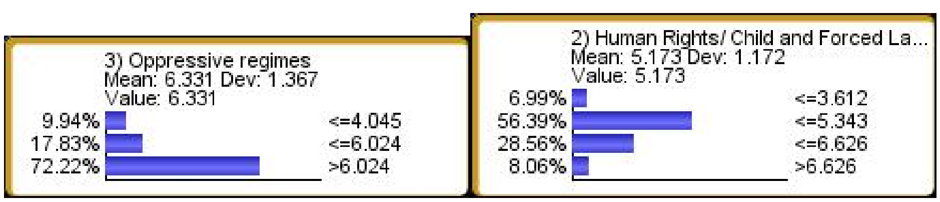


Fig. 6: Impact of Oppressive Regimes on Human Rights

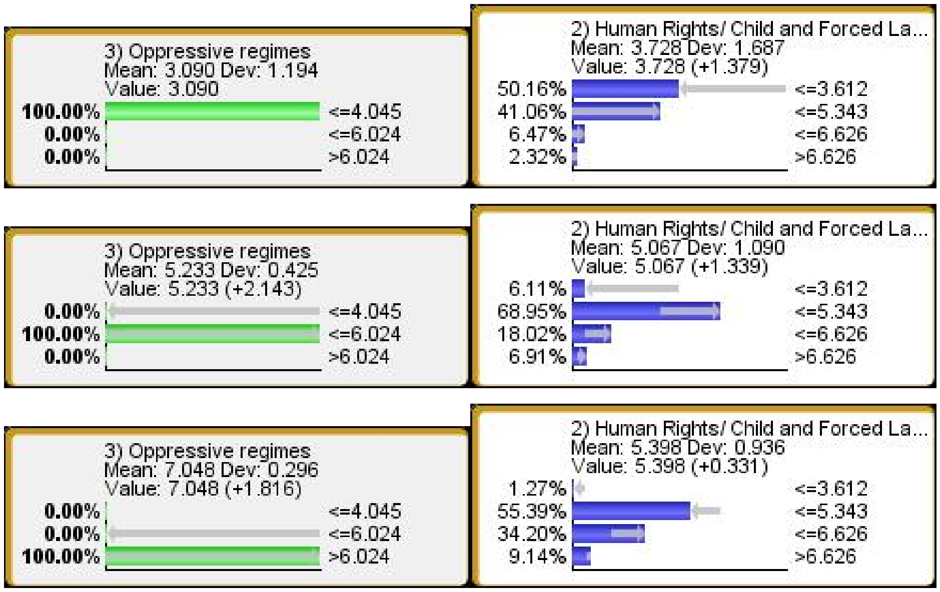


Fig. 7 shows the joint probability distribution between the nodes Risk Factor and Industry Specific Risk. This is equivalent, in this case, to a classical contingency table.

Fig. 7: Impact of Risk Factors on Industry Specific Risks

Risk Factors	<=2.262	<=4.222	<=6.563	>6.563
<=2.992	39.149	21.165	39.678	0.008
<=4.778	6.881	33.286	56.528	3.305
<=6.676	1.370	6.474	82.218	9.938
>6.676	0.007	2.160	45.448	52.384

Fig. 8 shows the joint probability distribution for a node Corporate Governance which has two parents Impact of Environmental Management Capacity and Traditional Governance Concerns. We note the augmented complexity of the table as the number of parents and classes increases. The probability in terms of percentage, for example, that the Corporate Governance score will be between 0 and 2.538 given that the Environmental Management Capacity score is between 0 and 2.905 and that the Traditional Governance score is between 0 and 3.526 is 61.23%.

Fig. 8: Impact of Environmental Management and Traditional Governance on Corporate Governance

Environmental Management Capacity	3) Traditional Governance Concerns	<=2.538	<=4.613	<=6.576	>6.576
<=2.905	<=3.526	61.234	35.414	3.340	0.013
	<=5.319	55.538	23.901	20.553	0.008
	<=6.915	45.110	28.835	22.742	3.313
	>6.915	61.984	21.486	4.800	11.729
<=4.788	<=3.526	24.360	61.317	11.026	3.296
	<=5.319	3.287	50.969	43.282	2.462
	<=6.915	2.414	10.444	74.314	12.828
	>6.915	11.179	27.182	29.742	31.896
<=6.251	<=3.526	31.324	41.049	20.456	7.171
	<=5.319	1.702	40.663	53.033	4.602
	<=6.915	0.003	2.804	76.059	21.134
	>6.915	4.587	11.756	21.817	61.840
>6.251	<=3.526	24.359	22.167	21.880	31.594
	<=5.319	0.008	43.294	40.439	16.260
	<=6.915	0.009	0.009	65.580	34.402
	>6.915	1.022	6.636	19.453	72.888

Fig. 9: Mutual Information for the IVA Bayesian Network

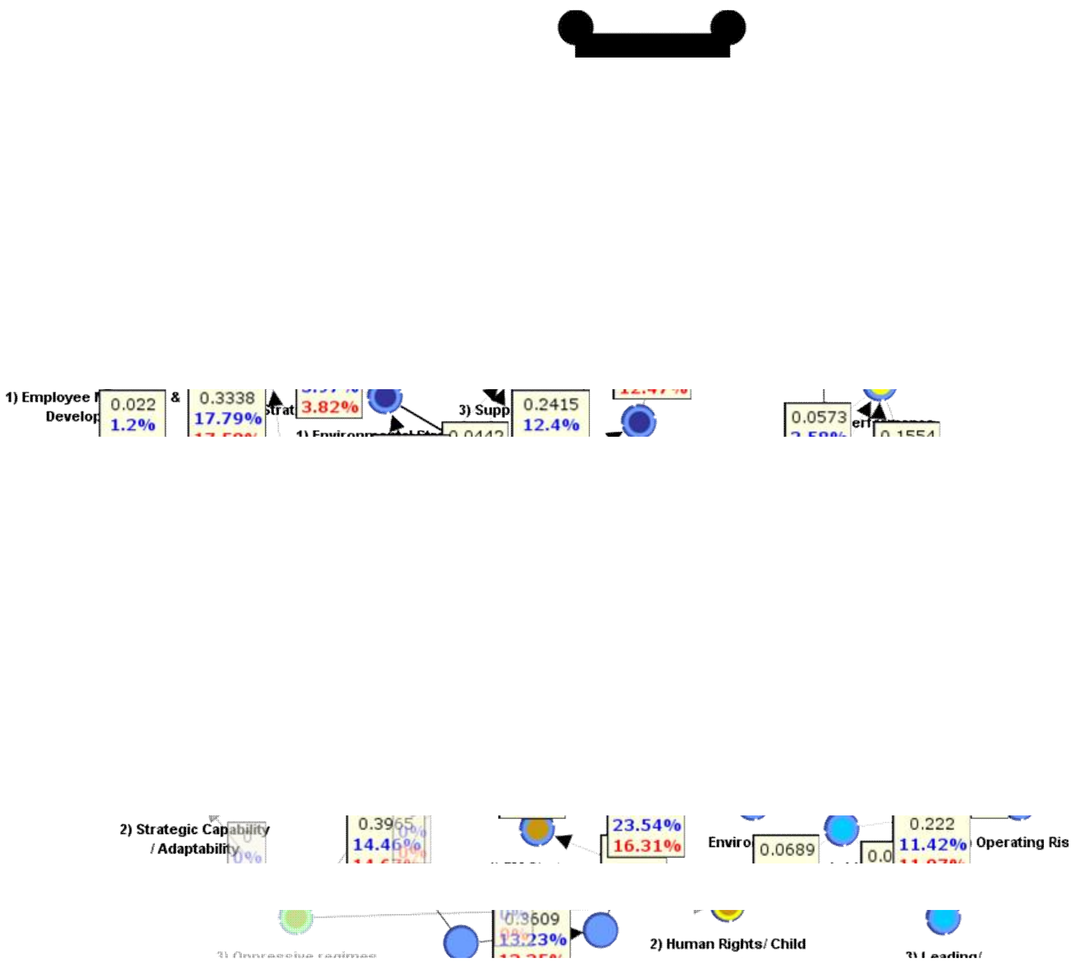
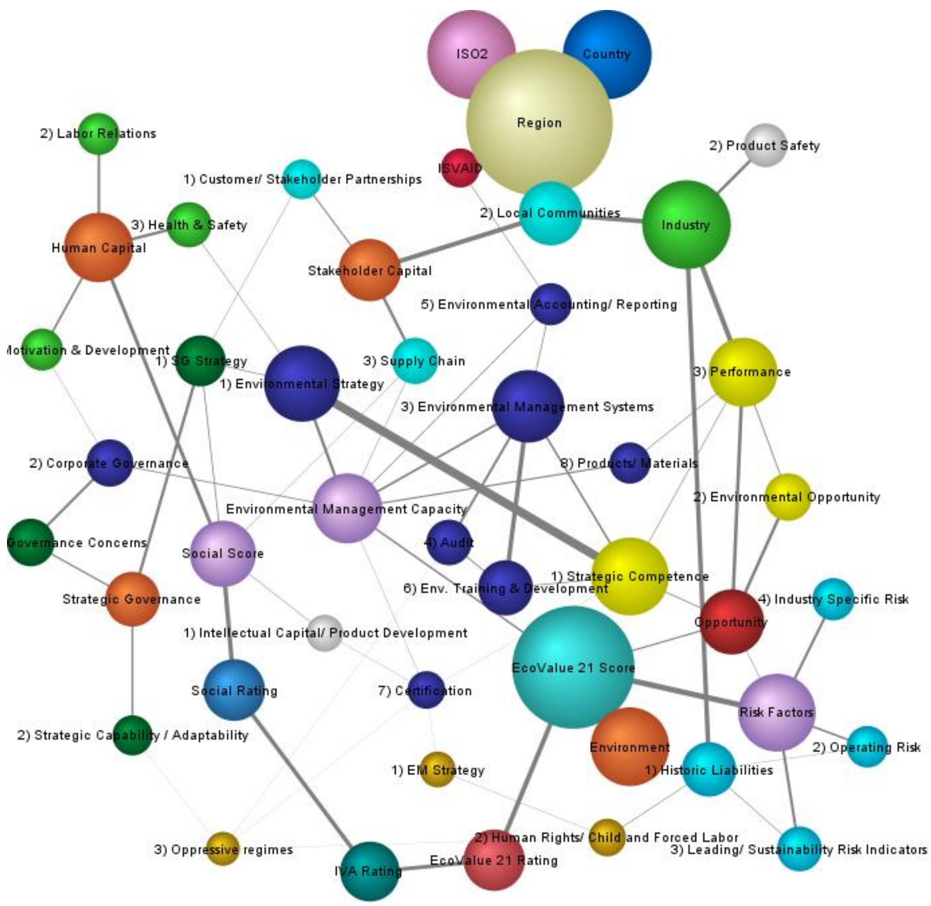


Fig. 9 shows the mutual information associated with the arcs of the IVA Bayesian Network. The figure in the top part of the each box situated on each arc gives the Mutual Information between the nodes. The second figure shows the relative Mutual Information (as a percentage) in the direction of the arc parent to child. The bottom figure is the relative mutual information in the opposite direction (child to parent. We observe, for example, that the "Performance" variable does not give a lot of information (7.45%) about the variable below "Industry" but that the reverse is not true (27.23%).

Fig. 10 displays the relative node strengths for the Bayesian network extracted from the IVA database. We recall that the strength of an arc can be measured in terms of the importance of the probabilistic relation between the nodes; typically we can use the mutual information measure given by equation 7. The global node force or strength is the sum of the force of both the entering arcs and the outing arcs; it provides some indication of the relative importance of a variable on its immediate neighbourhood. Unsurprisingly, Fig. 10 clearly indicates the importance of the variable Region on its neighbouring variables.

Fig. 10: Node Strengths for IVA Bayesian Network



3 Discussion and Conclusion

The power of Bayesian networks and their capacity to represent complex systems that bring into play both uncertainty and imprecision should not be underestimated. Much attention has already been given to issues of probability and the propagation of evidence in the form of new information through Bayesian networks. The semantic power of Bayesian networks and their capacity to deal with imprecise and overlapping concepts is, maybe still, not fully appreciated.

The IVA database example shows how easy it is to build a basic Bayesian network by directly exploiting an existing database. Clearly, further work can be done to transform this network into a causal model as well as integrate dynamic aspects, i.e; build a dynamic Bayesian network. Further, it is possible to combine this synthetic vision of key CSR issues with many of the existing databases relating to business and economics, etc.

The capacity to add and combine overlapping and imprecise constructs without impacting unduly on the overall performance of a network provides an important tool for confronting theories and theory development. Pushed to the extreme, workable networks can be built that combine knowledge from very different disciplines providing we are able to establish bridging concepts. In a similar fashion expert knowledge can be confronted with consumer perception using Bayesian Delphi methods that are constructed through the joint elaboration of belief systems of expert panels. This issue is of kew importance in CSR where conflicting positions can lead to very different policies.

The issue of conflicting belief systems leads us to a very pragmatic feature of Bayesian networks that is their capacity to analyze possible future scenarios in a consistent and coherent fashion. In fact, Bayesian networks provide a rigorous framework for studying "what if" scenarios. Hence it is possible to analyze the impact of new evidence as well a new theory on the performance of a network and more particularly target variables such as ratings

Another important feature of Bayesian analysis is the possibility of combining these networks with target analysis. This technique consists in optimising a network in such a manner that it explains in the behaviour of a target variable, for example a particular rating, in terms of the rest of the network. Further, once such an explanatory network has been built algorithmic tools exist that enable the target cell to be optimized. Hence it is possible to answer the question what are the most efficient set of actions to be taken to respect or achieve a certain level of standards. This makes Bayesian networks an extremely practical and actionable tool that can be exploited for policy making.

Bayesian networks have been claimed to be one of the most promising technological innovations of recent years that will have a major impact on Society and they are likely to constitute one of the most important tools for Knowledge Management *par excellence*.

References

Cain, J. (2001). *Planning Improvements in Natural Resources Management. Guidelines for Using Bayesian Networks to Support the Planning and Management of Development Programmes in the Water Sectors and Beyond*. Centre for Ecology and Hydrology, Wallingford.

Conrady, S., & Jouffe, L. (2013). *Tutorial on Driver Analysis and Product Optimization with BayesiaLab.* Retrieved from http://www.bayesia.us/white_papers/PSEM_v16.pdf, [cit. 2013-06-10]

Heckerman, D. A (2008). *Tutorial on Learning with Bayesian Networks*. Innovations in Bayesian Networks, 33–82.

Kullback, S., & Leibler, R.A. (1951). "On Information and Sufficiency". *Annals of Mathematical Statistics***,** 22 (1): 79–86.

Munteanu, P., & Bendou, M., (2001). *The EQ framework for learning equivalence classes of Bayesian networks*,. in Proc. Proceedings 2001 IEEE International Conference on Data Mining, pp.417,424.

MSCI. (n.d.). Retrieved from http://www.msci.com/resources/factsheets/MSCI_ESG_IVA_FactSheet.pdf , [cit. 2013-05-15]

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*., 2nd ed. Cambridge University Press.

Pearl, J., & Stuart, R. (2000). *Bayesian Networks.* UCLA Cognitive Systems Laboratory. Retrieved from http://bayes.cs.ucla.edu/csl_papers.html, [cit. 2013-06-17]

Pearl, J. (2013). Trygve Haavelmo and the Emergence of Causal Calculus, Technical Report R-391 Feb. 2013, *Econometric Theory*, special issue on Haavelmo Centennial.

Simon, H. (1953). Causal ordering and identifiability. In Hood, W.C.; Koopmans, T.C. *Studies in Econometric Method*. New York: Wiley. 49–74.

Mead, R., Paxton, J., & R. Sojda. (2006). Applications of Bayesian Networks in Ecological Modelling., Environmental Modelling and Simulation (EMS 2006), November 29 – December 1, 2006 **,** St. Thomas, USVI, USA

Meretakis, D., & Wüthrich, B. (1999). *Classification as Mining and Use of Labeled Itemsets*. in *Proc. ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery.*

Haenlein, M., & Kaplan A. M.,(1999). *A Beginner's Guide to Partial Least Squares Analysis*. Understanding Statistics, 3(4), 283–297

Contact

Renaud MacGilchrist
ESC Rennes School of Business
2 rue Robert d'Arbrissel, 35065 Rennes Cedex – France
renaud.macgilchrist@esc-rennes.fr

Irena Descubes
ESC Rennes School of Business
2 rue Robert d'Arbrissel, 35065 Rennes Cedex – France
irena.descubes@esc-rennes.fr

François Lépineux
ESC Rennes School of Business
2 rue Robert d'Arbrissel, 35065 Rennes Cedex – France
francois.lepineux@esc-rennes.fr