Latent Variable Bayesian Networks Constructed using Structural Equation Modelling

Alta de Waal*† Keunyoung Yoo‡
*Statistics Department, University of Pretoria, Pretoria, South Africa, alta.dewaal@up.ac.za
† Centre for Artificial Intelligence (CAIR), Meraka Institute, Council for Scientific and Industrial Research, South Africa
‡ Economics Department, Newsis Seoul, Republic of Korea, kyyou92@gmail.com

***Abstract*—Bayesian networks in fusion systems often contain latent variables. They play an important role in fusion systems as they provide context which lead to better choices of data sources to fuse. Latent variables in Bayesian networks are mostly constructed by means of expert knowledge modelling. We propose using theory-driven structural equation modelling (SEM) to identify and structure latent variables in a Bayesian network. The linking of SEM and Bayesian networks is motivated by the fact that both methods can be shown to be causal models. We compare this approach to a data-driven approach where latent factors are induced by means of unsupervised learning. We identify appropriate metrics for URREF ontology criteria for both approaches.**

Index Terms – latent variables, Bayesian networks, structural equation modelling, URREF ontology

I. Introduction

Bayesian networks (BNs) are commonly used in information fusion as it is extremely powerful in integrating information from various sources and disciplines. In order to utilise Bayesian networks for decision-making, problem perspectives, values and goals of the stakeholders need to be included as variables in the model as they provide crucial context to the problem domain. Most often these variables are included in the BN structure as latent variables - they are factors that contribute to the system for which data do not exist. There are two main phases in a fusion life cycle, namely the inception and design phase, and the routine operation (runtime) phase [1]. Latent variables play an important role in both these phases. In the design phase they provide context which lead to better choices of data sources to fuse. During routine operation, fusion can take on three forms, namely context fusion, observation fusion and latent fusion. Latent variables provide a mechanism for all three forms of fusion [2]. Apart from the role they play in information fusion, there are several additional advantages to having latent variables in BNs [3], [4]: It leads to more parsimonious models, problem perspectives, goals and values are integrated as variables into the model and it controls the size of conditional probability tables.

It is widely advocated that BNs can be designed from multiple information types such as empirical data, sensors, expert knowledge and literature. A BN consists of a causal structure and a set of parameters over this structure. The structure can be designed by a group of experts and the probabilities can be estimated using the EM algorithm, or both the structure and probabilities can be machine learned, or both can be elicited from expert knowledge. In the case of expert knowledge modelling, the causal structure most often contains latent variables for which empirical data is not available. Incorrect expert knowledge handling could result in epistemic uncertainty because by definition it is not the true system, but knowledge about the true system [2]. Some typical mistakes made with expert knowledge modelling are: 1) Misunderstanding the problem context, 2) adding complexity without value, 3) confusion about what the variables represent and 4) biased estimation of probabilities [4].

Many knowledge engineering processes exist in order to provide guidelines on BN design such as *Knowledge Engineering with BNs process* (KEBN) [4], *Iterative BN Development Cycle* [5], and *Bayesia Expert Knowledge Elicitation Environment* (BEKEE) [6]. Although knowledge engineering processes provide guidelines and support on the identification of variables, structural design and parameter elicitation, the main focus is on parameter elicitation as this is perceived as the main entry point of epistemic uncertainty into the system [7].

Numerous studies in ecological modelling and environmental management have highlighted problems with the design of the BN structure. First, the BN structure is created according to previously accumulated subject-matter information which is time-consuming and error-prone [8], [9]. Secondly, in multidisciplinary applications with many contentious subjects, it is easy to generate overly complex models [10]. One might argue that BN structures can be learned from data, but causal relationships cannot be ascertained from statistical data alone [11], [12]. Furthermore, as mentioned earlier, there are several advantages to the inclusion of latent variables in a BN which cannot be assumed through automated structural learning. This might result in a limited understanding of the system and flawed theoretical explanations [11], [13].

In the case where no statistical data is available, other knowledge representations such as concept and morphological maps [14], or qualitative probabilistic networks [15] can be used to design the BN structure. However, when statistical data is available, structural equation modelling (SEM) can be used to combine expert knowledge and data in a causal structure. The idea of this approach is to use the tested causal structure from a SEM as the structure for the BN. The BN modelling tasks left to do are the discretisation of nodes, the number of states and the cutoff values of the discrete

978-0-9964527-7-9 ©2018 ISIF

states. This approach of linking SEM and BNs is illustrated in [16] and [11]. The application areas are ecological and consumer modelling respectively. Both these studies argue that SEM and BNs can be viewed as causal models under certain conditions and because of this analogousness, the two methods can be linked. In fact, Druzdel and Simon [17] showed that the mechanism-based view of causality in SEM is directly applicable to BNs and that the prior theoretical knowledge about a domain, captured in a SEM can be used as the BN structure. Anderson and Vastag [13] further reviewed SEM and BNs in the context of a unified causal modelling framework which also motivates the feasibility of linking the two methods.

In this paper, we motivate and describe the construction of latent variable BNs using SEM. Because it is a SEM structure applied to a probabilistic framework, we call this approach *probabilistic SEM*. We also describe an alternative approach based on unsupervised learning. The second approach is probabilistic latent factor induction as described in [18]. In this approach the latent variables and structure are learned from observed data. We investigate uncertainty induced in the model and propose appropriate metrics for URREF criteria under the Representation and Data criteria groups for both approaches. Finally we illustrate the two approaches by means of an application.

II. Approach I: Probabilistic SEM

A. Causal Models

Defining both SEM and BNs as causal models is crucial in the argument to link the two methods. The fundamental questions of causality are: 1) What empirical evidence is needed in order to infer cause-effect relationships? 2) Once we have accepted the legitimacy of the cause-effect relationship in 1, what inferences can we draw from the information [12]? The appropriate approach to answer these questions has been debated mainly because of a lack of clear semantics for causal claims and well-founded mathematical tools for deriving causal answers [12]. More recent advances in graphical models transformed causal models by identifying causality under explicit weaker assumptions than generally made [19]. These assumptions are causal sufficiency, the causal Markov and faithfulness principle, and independence of specified and unspecified causes [11], [13]. Under these assumptions it is possible to elucidate potential causal relationships from data, derive causal relationships from a combination of knowledge and data, predict the effect of actions and evaluate explanations for observations and scenarios. In Pearl's words, causality has been mathematised [12].

Apart from SEM and BNs being viewed as causal models upon meeting the required assumptions [13], other similarities between the two methods exist. SEM and BN are both graphical models in the sense that they use nodes to represent variables and directed arcs to represent causal links between the variables. Also, they can estimate underlying latent variables in the system using observed variables [11].

B. Structural Equation Modelling

SEM is often used in the fields of social science, business and marketing to model abstract concepts such as intelligence, attitude, and inclination [19]. It is a multivariate statistical method that incorporates regression, path anaysis and factor analysis to explore relationships among correlated variables [11]. SEM can be summarised as a linear model concerned with the modelling of causal relationships between variables (both observed and latent) [19]. The initial intent of SEM was to combine qualitative cause-effect information with statistical data, implying that in the equation $y = \beta x + \epsilon$, the equality sign conveys 'is determined by', rather that algebraic equality [12]. The mathematical framework of causality allows the SEM practitioner to return to this initial intent. Due to its flexibility, SEM is preferred by researchers in modelling situations where one cannot simply design and conduct experiments, for example, because of ethical concerns, or when data is not observable [20]. Typically an observable dataset consists of a test or a survey that measures different aspects of an unobservable concept. Although we cannot directly measure someone's inclination towards becoming an entrepreneur, we can be fairly confident that a person is inclined to become an entrepreneur if we know that the respondent strongly agrees to being his/her own boss and taking a lot of risk, when given a survey. The SEM practitioner performs factor analysis to determine the appropriate number of factors to include in the model. It takes sufficient knowledge of the data to understand what each factor represents as well as to establish the path between the factors. The model is then fitted and evaluated and the practitioner can interpret the result.

C. Bayesian networks

BNs need no introduction in the field of information fusion and numerous applications has been reported on [5], [21], [22]. Unlike SEM, where a theoretical causal structure is developed, the BN structure can be learned from data or elicited from expert knowledge. Although a theoretically valid structural model can be forced as a BN, the resulting BN is not as capable as SEM for theoretical explanations [13]. A limitation from a social science perspective is that they do not differentiate between a latent construct and its measures (observed variables) [16]. The main role of BNs is to facilitate the analysis of actions as they are suited for prediction and diagnosis, rather than theory confirmation.

D. From SEM to BN

The output of a SEM is an empirically validated model based on theoretical construction [16]. The causal path between factors and manifest variables are all specified according to the relevant theory and this structure is used as the BN structure. The only tasks left in the BN design phase are the discretisation of nodes, number of states and the estimation of probabilities. In [16], latent scores computed for SEM are used as raw data for probability estimation in the BN. The probabilities, when the network structure is specified, are learned by maximum likelihood, where the occurrence of the variable states are simply counted [6].

III. Approach II: Probabilistic latent factor induction

Probabilistic Latent Factor Induction is a workflow within the BayesiaLab software package [6]. It has the same objective as SEM in the sense that it aims to find causal relationships between variables, but is based on principles derived from information theory. While correlation and covariance are the central measures for SEM, probabilistic factor induction is based on measures from information theory.

A. Structural learning on observed variables

Structure learning involves a data driven process to find the optimal structure on the observed (manifest) variables. The aim is to find a tree (a graph without cycles) structure for the network which restricts the number of parents for each node to 1, making the search much simpler. This algorithm is known as the Chow-Liu algorithm or Kruskal's algorithm and it produces the maximum weight spanning tree for the network [23]. Once a structure has been obtained we can induce small changes to the network in order to find the local maximum. These small changes are defined as adding, removing or reversing an edge to the network, while making sure that no cycles are created by these changes [24]. It is not guaranteed that a better network will be found after all the local modifications have been performed, in which case the initial network is considered the best network. Alternatively, one can use techniques such as random restarts [24] which starts the process of local modification with a different initial network (possibly by adding a small noise to the data) to see if many different starting points lead to the same or similar local optimal network.

Two graphs which are Markov equivalent contain the same conditional independence assumptions. This means the graphs which are Markov equivalent belong to the same Markov equivalence class and this happens when the undirected skeleton and the v–structures of the graphs are the same. Put differently, "different graphs that share exactly the same d-separation properties are said to be Markov equivalent" [23]. We can use this fact to narrow our search space, which entails exploring the space of equivalence classes of BN as opposed to the entire space of BN [27]. Essentially, we are searching for a group of networks that belong to the same family, instead of directly looking for the single best network. By doing so, we can bypass many redundant calculations and only go into detailed modifications once we have obtained the optimum equivalence class. After the optimum equivalence class has been found, we can search for the local optimum by adding or removing directed or undirected edges [25]. The Equivalent Class (EQ) algorithm can be used alongside the minimum weight spanning tree (MWST) algorithm, on a fully unconnected dataset. If both algorithms return the same network, it is quite possible that the network obtained is the optimal network.

B. Clustering

Here, the nodes within the network are clustered, based on the Kullback-Leibler (KL) divergence, using hierarchical agglomerate clustering. The KL metric is calculated for every pair of nodes, which measures how close a node is to every other. Although the KL metric is not symmetrical - and therefore not a distance it is a sufficient metric to measure the impact of removing a link between nodes. At the start of the process, the nodes are all treated as a cluster on its own and two clusters with the smallest distance are joined into a single cluster. This process is repeated either until a satisfactory number of clusters have been obtained or until no clusters are deemed close enough to be joined into one cluster. This is similar to performing an exploratory factor analysis where the researcher tries to find the optimal number of factors within the data.

Cross-validation is performed to check whether the clustering groups the same nodes together frequently. To do this we start off by adding small noise to the data, create the network structure, perform clustering on the network and repeat this process many times. If the results show that same nodes are clustered into the same group many times over the iterations, we can safely assume that it is indeed the most likely scenario of clustering outcome. Pair-wise KL divergence is calculated for all variables in the network and those values are used to perform hierarchical clustering.

C. Multiple Clustering

We are now interested in finding factors from the clustering data, with as many factors being introduced into the model as the number of clusters. This process entails inducing factors with discrete states for each cluster, making sure that the factors have high mutual information with their children nodes [26] The imputing of factor states to each observation can be done using maximum likelihood [18]. The representative values for each state can be obtained by calculating the weighted average of the manifest variables' means given the specific states, where the weights are relative significance in relation to the factor.

So far we have only dealt with how the observed variables interact with factors. This is what is known as the measurement model in the field of SEM [19] . We will now move onto the structural part which deals with how latent variables interact with each other.

D. Structural learning on latent variables

Here we must find the relationships between the factors while keeping the relationship between factors and their manifest variables intact. In order to enforce this restriction, we must resort to using the Tabu algorithm, which searches for the local optimum while verifying at every step that it does not belong to a set of restricted networks [27]. Establishing the relationships between the factors returns the structural part of the model and this completes the network structure learning of the data. Table I summarises the steps for PSEM and Latent factor induction.

TABLE I
STEPS IN TWO APPROACHES

PSEM	Latent factor induction
1. Covariance matrix between observed variables	1. Learn structure from observed variables
2a. Confirmatory factor analysis (CFA) to confirm number of clusters 2b. Path analysis to confirm the structure	2. Cluster observed variables
3. Define latent factors using latent factor scores (create variables states)	3. Define latent factors for each cluster (create variables states)
4. Parameterise the model using EM algorithm	4. Learn structure between latent variables

IV. Evaluation under URREF - Uncertainty in Representation

At the core of this study is the investigation of alternative approaches to expert knowledge modelling to construct latent variables in BNs. In the context of information fusion, uncertainty propagates throughout the life cycle of the fusion system [1] and it is important to characterise this uncertainty in order to understand decision consequences. It is for this purpose that the Uncertainty Representation and Reasoning Evaluation Framework (URREF) [28] ontology was created in order to define and link uncertainty concepts in a fusion system. URREF has four criteria groups namely *Representation, Reasoning, Data* and *Data Handling*. We argue that *Representation* and *Data* criteria groups are most relevant to evaluate uncertainty that propagates through the BN structure development: The structure is a representation of the system and it gets informed by data. The criteria for *Representation* are compatibility, knowledge handling, adaptability, simplicity and expressiveness. The criteria for *Data* are weight of evidence, relevance to problem, credibility and quality. In this section we match relevant metrics to criteria.

A. PSEM

1) Reliability: When working with SEM, their reliability is typically confirmed using average variance explained (AVE) and composite reliability (CR). Reliability is a measure which quantifies the percentage of variance in an observed variable explained by the latent variable [19]. If there are little measurement errors, the reliability coefficient will be high. The minimum recommended values are 0.5 and 0.7 for AVE and CR, respectively. The significance of factor loading in the measurement model should be tested using the t-statistic. One would hope to see values higher than 1.96 in absolute value for significance at 5% or higher than 2.58 in absolute value for significance at 1%. It is stated in [29] that reliability is an attribute of an information source,and measures the consistency of a measure of some phenomenon. It forms part of the credibility criterion. In this case the two metrics for reliability (credibility) are AVE and CR.

2) Accuracy and Simplicity: Chi-square test tests for significance between actual covariance matrix and estimated covariance matrix. Null hypothesis assumes no significant difference between the two. This is rarely met because of sample size sensitivity (small difference is seen as significant in large samples). It also requires the condition of multivariate normality, hence it is no longer seen as viable option. Three types of indices are reported frequently: absolute t index, incremental t index and parsimony-adjusted index. Absolute index takes on values between 0 and 1 and it can be thought of as an R^2. But instead of measuring how much variance is explained by the model, it measures how much the variance-covariance matrices correspond to each other. Naturally, values closer to 1 are preferred. Incremental t index shows how the model has improved relative to the baseline model which essentially assumes the value of 0 for covariances [30]. Parsimony index, as the name suggests, allows us to identify the simpler model among the available models. If all models yield satisfactory results that are of similar level, parsimony index prefers the simplest model [30]. The absolute and incremental t indices act as metrics for accuracy (under the *Data* criteria group). The parsimony index acts as metric for simplicity (under the *Representation* criteria group).

B. Latent factor induction

MWST is a simple, data driven method of finding a network structure. Of course, using different algorithms for structure learning will most likely lead to slightly different models suggested by each algorithm. The minimum description length (MDL) for each model can be evaluated to select the best model. MDL operates under the logic that regularities within the data can be compressed, meaning certain symbols can be used to describe the data in a more compact way than the actual data. Highly regular data can therefore be highly compressed [31]. MDL consists of two parts as stated in the following equation:

$$MDL(B,D) = \alpha DL(B) + DL(D|B) \tag{1}$$

where B is the model (Bayesian Network) and D is the observed data, hence $DL(B)$ is the complexity (number of bits) of the suggested model and $DL(D|B)$ is the number of bits to describe the log-likelihood of the data given (with the help of) the model, which is none other than the error [31]. Eq. 1 states that the MDL score is the sum of the complexity of the

model and the complexity of the errors. Generally, if the model is highly accurate it will need much description (as it will have many terms) but the resulting error will be small. Conversely, if the model is very simple, its description will be very short but we will need a lot of information to describe its errors. The α is a structural coefficient, or simply a weight, which we can use to reduce the impact of model complexity. Even if the model is highly complex and $DL(B)$ is high, by making α small, model complexity will have reduced impact on the overall MDL [5]. A fully unconnected network will translate to the minimum value for $DL(B)$ and a fully connected network will translate the minimum value for $DL(D|B)$. Thus obtaining minimum MDL finds the right balance between the two extremes. If we start off with a blank network, an edge will connect two nodes only if the decrease in $DL(D|B)$ is larger than the increase in $DL(B)$ [6].
The MDL score acts as metric for simplicity (under the *Representation* criteria group).

C. Ambiguity

For the PSEM approach ambiguity can occur in the Representation criterion group. In factor analysis, a latent variable is a novel construct from the manifest variables. In many factor models, however, the factors are not uniquely constructed and therefore indeterminate [16]. This leads to a potential infinite number of sets of that can be computed from any given analysis that satisfy the stipulations of the common factor model. Although not solved, measures have been put in place to mitigate the problem [32]. Factor indeterminacy is relevant during the design phase of a PSEM model, and we therefore suggest that ambiguity needs to be added to the *Representation* criteria group [1].

Table II summarises the evaluation metrics relevant to URREF criteria in the respective approaches.

TABLE II
SUMMARY OF EVALUATION METRIC RELEVANT TO URREF CRITERIA

URREF Criteria	**PSEM**	**Latent factor induction**
Reliability	AVE CR	KL divergence
Accuracy	Chi-square R^2 (absolute fit index) CFI (incremental fit index)	
Simplicity	RMSEA (parsimony index)	MDL
Ambiguity	Factor score indeterminacy	

V. APPLICATION

In this section we present an application in order to illustrate the two approaches. This application falls in the domain of consumer science. SEM is a popular modelling technique in consumer science in which many theories are validated with observed data in the form of questionnaires. The latent constructs (or variables) in the theory drive the questionnaire design.

Social Networking Sites (SNS) represent a great opportunity today for companies to advertise their products and services as well as target and personalise their messages based on the data people declared online. However, how people perceive the new advertising methods employed in SNS is still largely unknown, as well as whether they consider such advertisements to be an intrusion on their private, although social, space. In addition, very little empirical research of causal relationships exists, leaving unanswered questions about how SNS users perceive advertisements posted on their own online social profile.

The conceptual framework is based on Theory of Planned Behaviour (TPB) [33]. The components of TPB are four general constructs: behavioural intention, attitude, subjective norm and perceived behavioural control. Following the TPB, we thus predict that behaviour towards SNS advertising will be influenced by the user's attitude towards SNS advertising, the subjective norms and perceived behavioural control. The attitude is itself dependent on four main beliefs: trust, attitudes towards advertising in general, advertising value and advertising intrusiveness, which are themselves dependent on other antecedent variables as described in the literature. The study population comprised of active adult (18 years and older) Facebook users. The survey was developed in English for both South Africa and Australia and delivered online. Sampling in both countries involved the use of market research firms holding consumer panels where the firms' provided a link to the survey. Participants were incentivised by the market research firms in accordance with their normal practices and a sample of 401 were realised in both countries respectively, resulting in a total of 802 respondents.

There are 9 factors in this SEM, namely Privacy Concern; Trust; Ad [2] intrusiveness; Behaviour towards brand; Behaviour towards ad Perceived behaviour control; Attitude towards ads; Attitude towards FB ad; and Ad values. The observed data consists of 46 manifest variables which we don't discuss for brevity. In all figures we only display latent variables and not manifest variables.

A. SEM

The SEM, as well as estimated standardised path coefficients and squared multiple correlation for endogenous variables were constructed in SPSS AMOS [34]. All path coefficients as well as error variance were significant. The resulting model (manifest variables ommitted) is shown in Figure 1

The value of -.13 between *Ad intrusiveness* and *Attitudes towards FB ad* indicates that there is an inverse relationship, albeit relatively weak, between the two variables, where an increase of 1 standard deviation in *Advertising intrusiveness* will lead to a decrease of 0.13 standard deviations in *Attitudes*

[1]Ambiguity is an entity in the *UncertainEvidence* group, but that is not relevant to *Representation*

[2]For the sake of brevity in figures, *advertisement* is abbreviated to *ad*

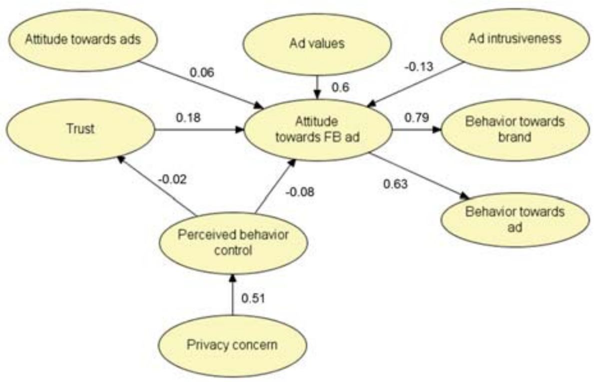

Fig. 1. SEM

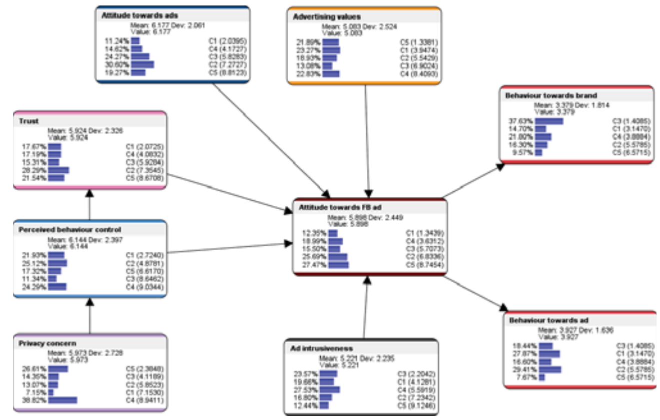

Fig. 2. Probabilistic SEM

TABLE III
GOODNESS-OF-FIT SEM

Model fit index	Index value
CFI	0.926
RMSEA	0.051
RMSEA upper 90%	0.054
RMSEA lower 90	0.049

towards FB ad. A strong positive relationship exists between *Attitudes towards FB ad* and *Behaviour towards ad*, indicated by the coefficient value of .79. The value of .63 for squared multiple correlation of *Behaviour towards ad* shows that *Attitudes towards FB ad* explains 63% of the variance in *Behaviour towards ad*. Other values in the diagram can be interpreted in the same way.

Furthermore, Table III shows values for the model goodness of fit. Bentler's confirmatory fit index (CFI) is larger than 0.9 while the Root Mean Square Error of Approximation ($RMSEA$) is less than 0.055, which are both indicative of an adequate overall model fit.

B. Probabilistic SEM

The PSEM and probabilistic latent factor induction models were constructed in BayesiaLab 7.0.7 (www.bayesia.com). Here, the path between factors and manifest variables are all specified according to the relevant theory and the theoretical or expert-developed SEM is turned into a BN for inference. The corresponding structural model of the probabilistic SEM (PSEM) is shown in figure 2.

The graph displays small windows which show the posterior probability visually as well as numerically: the column of numbers on the right lists the different states of the nodes and the column on the left lists the respective default probability of belonging to the state. A high valued state indicates a favourable or a strong opinion towards the specific variable. The probabilities, when the network structure is specified, are learned by maximum likelihood, where the occurrence of the variable states are simply counted [18]. Learning probabilities in this manner also makes the network less responsive to changes, as the conditional probabilities will only reflect the changes if there are observations in the data which possess the given state value for the variable.

C. Latent Factor Induction

Figure 3 illustrates the output of the manifest variable cluster of the probabilistic factor induction approach (Step 1, column 2 in Table I). Figure 4 shows the clustering output (Step 2, column 2 in Table I). From the manifest variable names, it can be seen that question groups were clustered together. For example, a manifest variable starting with 'Q8' belongs to a group of questions designed to address 'Trust'. This information, however, is not available to the clustering algorithm. Whereas 9 variables were identified in the SEM approach, the clustering algorithm could only elucidate 7 variables. Figure 5 illustrates the final structure where the relationships between the latent factors are learned. Only the latent variable structure is shown. The lean structure typical of the MWST algorithm where all nodes only have one parent is evident.*Behaviour towards ad* and *Behaviour towards brand* were merged. Figure 6 illustrates the marginal probabilities of the latent factors.

D. Discussion

The PSEM takes the SEM developed from theory by the researcher and converts it into a BN. This allows SEM practitioner to take a step forward by adding the capability to perform what-if analysis onto the network. On the other hand we have the latent factor induction, which can be used without any prior knowledge regarding the data, as the process is purely data-driven. This can assist the practitioner in dynamically exploring the data. This particular application does not contain a target node, which makes it difficult to directly compare the performance of the two approaches. It is, however, encouraging that similar clustering results were obtained by the latent factor induction than was suggested by the theoretical SEM construct.

VI. CONCLUSIONS

In this paper we introduced two alternative approaches to expert knowledge modelling to construct latent variables in

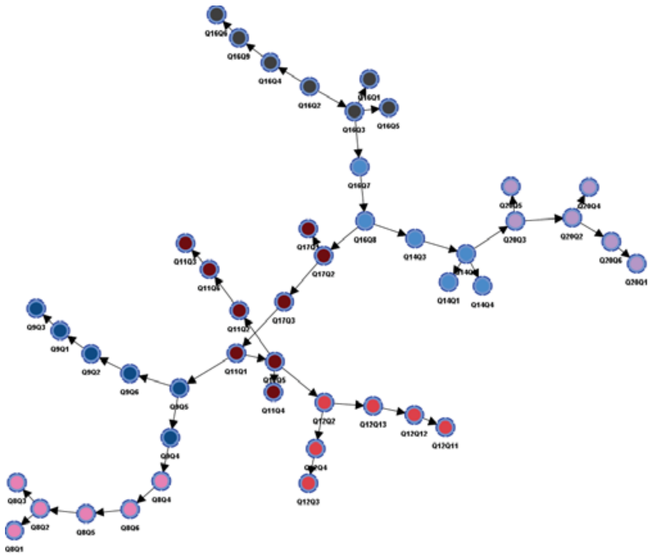

Fig. 3. Probabilistic latent factor induction: Step 1

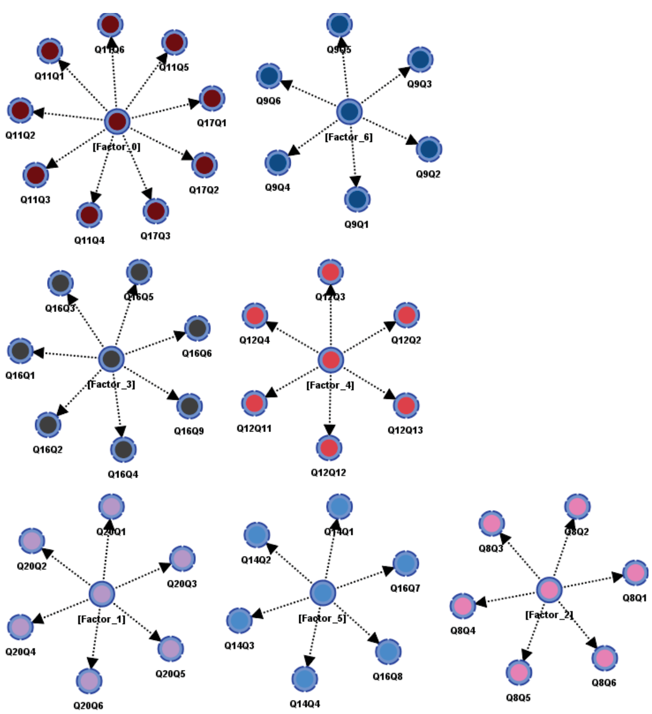

Fig. 4. Probabilistic latent factor induction: Step 2

BNs. The one approach (PSEM) is theory-driven and the other (Latent factor reduction) is data-driven. We motivate using SEM in the theory-driven approach as both SEM and BNs are causal models. We recommended metrics for URREF criteria in both approaches and finally we illustrated the output of the two techniques by means of a consumer science application.

The design of an evaluation framework using URREF metrics will enable us to test and compare different methods to construct latent variables. For future work, a first priority is to compare these two approaches with an expert knowledge modelling approach. On a technical level we would like to expand on the MWST algorithm to allow for more complex BN structures. Finally we plan to compare the methods on an end-to-end fusion system.

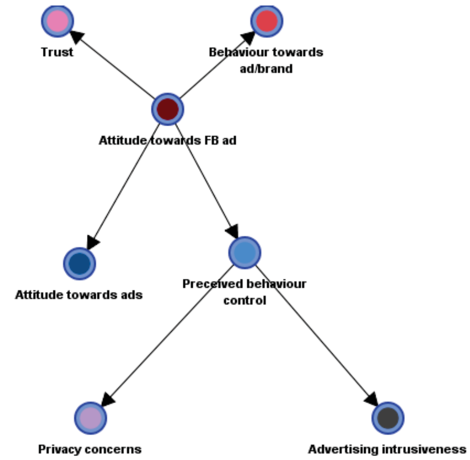

Fig. 5. Probabilistic latent factor induction: Step 3

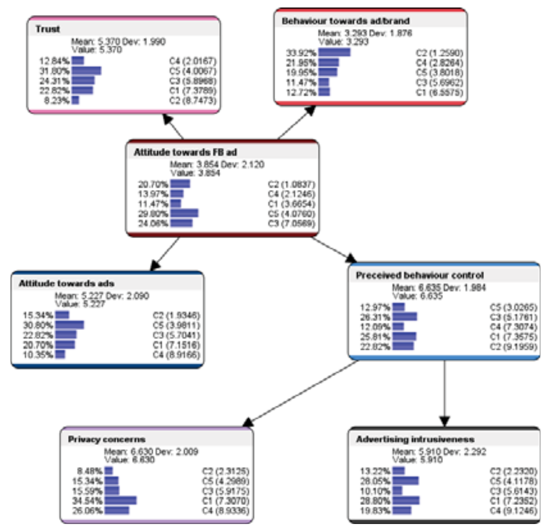

Fig. 6. Probabilistic latent factor induction: Step 4

REFERENCES

[1] J. P. de Villiers, K. Laskey, A. L. Jousselme, E. Blasch, A. D. Waal, G. Pavlin, and P. Costa, "Uncertainty representation, quantification and evaluation for data and information fusion," in *Information Fusion (Fusion), 2015 18th International Conference on*, July 2015, pp. 50–57.

[2] A. de Waal, H. Koen, P. de Villiers, H. Roodt, N. Moorosi, and G. Pavlin, "Construction and evaluation of Bayesian networks with expert-defined latent variables," in *Information Fusion (FUSION), 2016 19th International Conference on*. IEEE, 2016, pp. 774–781.

[3] B. Großmann-Hutter, A. Jameson, and F. Wittig, "Learning Bayesian networks with hidden variables for user modeling," in *IJCAI-99 Workshop" Learning About Users*, 1999, pp. 29–34.

[4] K. B. Korb and A. E. Nicholson, *Bayesian artificial intelligence*, 2nd ed. Boca Raton, FL: CRC Press, 2011.

[5] S. Johnson, K. Mengersen, A. de Waal, K. Marnewick, D. Cilliers, A. M. Houser, and L. Boast, "Modelling cheetah relocation success in southern Africa using an Iterative Bayesian Network Development Cycle," *Ecological Modelling*, vol. 221, no. 4, pp. 641–651, 2010.

[6] S. Conrady and L. Jouffe, *Bayesian Networks and BayesiaLab: A Practical Introduction for Researchers*. Bayesia USA, 2015.

[7] M. Druzdel and L. C. Van Der Gaag, "Building probabilistic networks:" where do the numbers come from?"," *IEEE Transactions on knowledge and data engineering*, vol. 12, no. 4, pp. 481–486, 2000.

[8] J. Bromley, N. A. Jackson, O. Clymer, A. M. Giacomello, and F. V. Jensen, "The use of Hugin® to develop Bayesian networks as an aid to integrated water resource planning," *Environmental Modelling & Software*, vol. 20, no. 2, pp. 231–242, 2005.

[9] L. Uusitalo, "Advantages and challenges of Bayesian networks in environmental modelling," *Ecological Modelling*, vol. 203, no. 3-4, pp. 312 – 318, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304380006006089

[10] I. Alameddine, Y. Cha, and K. H. Reckhow, "An evaluation of automated structure learning with Bayesian networks: An application to estuarine chlorophyll dynamics," *Environmental Modelling & Software*, vol. 26, no. 2, pp. 163–172, 2011.

[11] X. Xu, J. Sun, H. Nie, D. Yuan, and J. Tao, "Linking structural equation modeling with Bayesian network and its application to coastal phytoplankton dynamics in the Bohai Bay," *China Ocean Engineering*, vol. 30, no. 5, pp. 733–748, 2016.

[12] J. Pearl, *Causality: models, reasoning, and inference*, 2nd ed. Cambridge, U.K.; New York: Cambridge University Press, 2009.

[13] R. D. Anderson and G. Vastag, "Causal modeling alternatives in operations research: Overview and application," *European Journal of Operational Research*, vol. 156, no. 1, pp. 92–109, 2004.

[14] M. J. Eppler, "A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing," *Information visualization*, vol. 5, no. 3, pp. 202–210, 2006.

[15] M. P. Wellman, "Fundamental concepts of qualitative probabilistic networks," *Artificial Intelligence*, vol. 44, no. 3, pp. 257–303, 1990.

[16] S. Gupta and H. W. Kim, "Linking structural equation modeling to Bayesian networks: Decision support for customer retention in virtual communities," *European Journal of Operational Research*, vol. 190, no. 3, pp. 818–833, 2008.

[17] M. J. Druzdzel and H. A. Simon, "Causality in Bayesian belief networks," in *Uncertainty in Artificial Intelligence, 1993*. Elsevier, 1993, pp. 3–11.

[18] S. Conrady and L. Jouffe, "Probabilistic latent factor induction with Bayesialab," 2014.

[19] R. H. Hoyle, *Handbook of structural equation modeling*. Guilford Press, 2012.

[20] B. M. Byrne, *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge, 2010.

[21] H. Koen, J. de Villiers, A. de Waal, G. Pavlin, F. Mignet, and P. de Oude, "A framework for inferring predictive distributions of rhino poaching events through causal modelling," in *Information Fusion (FUSION), 2014 17th International Conference on*. ISIF, July 2014.

[22] G. Pavlin, P. de Oude, M. Maris, J. Nunnink, and T. Hood, "A multi-agent systems approach to distributed bayesian information fusion," *Information fusion*, vol. 11, no. 3, pp. 267–282, 2010.

[23] T. Koski and J. Noble, *Bayesian networks: an introduction*. John Wiley & Sons, 2011, vol. 924.

[24] A. Darwiche, *Modeling and reasoning with Bayesian networks*. Cambridge University Press, 2009.

[25] P. Munteanu and M. Bendou, "The EQ framework for learning equivalence classes of Bayesian networks," in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*. IEEE, 2001, pp. 417–424.

[26] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[27] F. Glover, "Tabu search: A tutorial," *Interfaces*, vol. 20, no. 4, pp. 74–94, 1990.

[28] P. Costa, K. Laskey, E. Blasch, and A. L. Jousselme, "Towards unbiased evaluation of uncertainty reasoning: The URREF ontology," in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 2301–2308.

[29] E. Blasch, K. B. Laskey, A.-L. Jousselme, V. Dragos, P. C. Costa, and J. Dezert, "Urref reliability versus credibility in information fusion (stanag 2511)," in *Information Fusion (FUSION), 2013 16th International Conference on*. IEEE, 2013, pp. 1600–1607.

[30] R. B. Kline, *Principles and practice of structural equation modeling*. Guilford publications, 2015.

[31] P. Griinwald, "Introducing the minimum description length principle," *Advances in minimum description length: Theory and applications*, p. 3, 2005.

[32] F. Acito and R. D. Anderson, "A simulation study of factor score indeterminacy," *Journal of Marketing Research*, pp. 111–118, 1986.

[33] I. Ajzen, "The theory of planned behavior," *Organizational behavior and human decision processes*, vol. 50, no. 2, pp. 179–211, 1991.

[34] J. Arbuckle, "Amos (version 23.0)[computer program] ibm spss," 2014.