Subjective Bayesian Networks and Human-in-the-Loop Situational Understanding[⋆]

Dave Braines[1,2], Anna Thomas[1], Lance Kaplan[3], Murat Şensoy[6,2],
Jonathan Z. Bakdash[4,5], Magdalena Ivanovska[7], Alun Preece[2], and Federico Cerutti[2]

[1] IBM Hursley Park, UK
[2] Cardiff University, UK
[3] U.S. Army Research Laboratory, USA
[4] U.S. Army Research Laboratory South Field Element at the University of Texas Dallas, USA
[5] Texas A&M Commerce, USA
[6] Ozyegin University, Turkey
[7] University of Oslo, Norway

Abstract. In this paper we present a methodology to exploit human-machine coalitions for situational understanding. Situational understanding refers to the ability to relate relevant information and form logical conclusions, as well as identify gaps in information. This process for comprehension of the meaning information requires the ability to reason inductively, for which we will exploit the machines' ability to 'learn' from data. However, important phenomena are often rare in occurrence with high degrees of uncertainty, thus severely limiting the availability of instance data for training, and hence the applicability of many machine learning approaches. Therefore, we present the benefits of Subjective Bayesian Networks—i.e., Bayesian Networks with imprecise probabilities—for situational understanding, and the role of conversational interfaces for supporting decision makers in the evolution of situational understanding.

1 Introduction

Human situational understanding is filled with inductive reasoning. Say you just landed at Heathrow Airport in London, UK: the sun is blazing in the sky and a glorious warm temperature of 23 Celsius (74 Fahrenheit) welcomes you in the South of Britain. On the basis of this observation, it is rational to conclude that usually the South of Britain enjoys lovely weather, especially if the same happens the second day, the third day, and the fourth day of your visit. From a human perspective, general rules and thus understanding are, therefore, often derived on the basis of scarce data/information. Consequently,

human decision making frequently exhibits heuristics and biases rather than following rationality [11].

Sometimes limited data is not a problem, especially in those cases where we can have access to an *oracle*, mostly an expert in the domain. You might receive a useful piece of information from a friend who lived in the South of Britain for years, or you can access historical data and statistics showing that the South of Britain does not usually enjoy lovely weather, and therefore this apparent normality is in fact an exception. Oracles can help in overcoming scarcity of actual data through access to other information or rules that are relevant to the domain.

As humans we, therefore, apply analyses and judgements to relevant information "to determine the relationships of the factors present and form logical conclusions concerning threats, opportunities, and gaps in information" [7]. This is *situational understanding*.

Machine learning approaches are potentially powerful allies in situational understanding [3]. This is because machine learning algorithms are able to efficiently handle large quantities of information, which is extremely useful to support inductive reasoning in situational understanding, as well as deriving logical conclusions. However, they are generally useless for identifying gaps in information as well as in providing insights, such as those that could be provided by oracles. Moreover, many of the best algorithms for machine learning often assume the existence of a large training set with independent and identically distributed (IID) data. Algorithms for data with limited instances (class imbalanced) is a specific research area [5], as are algorithms for non-IID data [24]. Unfortunately, the assumption of large amounts of balanced and IID data tends to unrealistic in the real–world. Rare events often arise from multiple *dependent* factors: for example, the risk of political instability is a combination of corruption, illicit activities, and organised crime [8].

The need for less training data and modeling of underlying dependencies is particularly important in situational understanding problems where many important phenomena will be rare in occurrence, severely limiting the availability of instance data and, hence, the applicability of many machine learning approaches, including Bayesian and Deep Learning [16] approaches. Coupled with this, supporting human analysts in terms of more effective communication of uncertain information is also a key issue in situational understanding problems [6].

In this paper—that is an extended version of [2]—we propose a human-machine coalition partnership for real-world situational understanding by exploiting the strengths of each member in the coalition. Machines' strengths are linked to data analysis, and we explicitly address the unrealistic assumption of large training sets that could undermine the role of machine agents in such a human-machine coalition. Moreover, human experts are usually considered useful oracles, and we need to provide useful human-machine interfaces in order to support co-design and co-evolution of the coalition for situational understanding. Specifically, we consider a system within which the human agents can contribute to or correct the machine agent parts of the system.

To exemplify our proposal, we discuss a running example about the German stock market in Section 2, and in Section 3 we exploit one of the machines' strengths: performing inductive reasoning with quantitative measures such as probabilities. We dis-

cuss a robust approach to handling uncertain information from a rather scarce dataset, namely Subjective Bayesian Networks, an extension of Bayesian Networks using uncertain probabilities. This helps us towards overcoming one of the main issues related to Bayesian networks: the lack of information about the certainty of the trained model.

We then show, in Section 4, that Subjective Bayesian Networks are well suited for situational understanding. Our tests show that they provide more accurate results compared to other approaches to Bayesian networks with uncertain probabilities, such as Credal networks [23] and belief networks [20].

Finally, in Section 5 we summarise an evaluation we performed with a focus group on the usage of conversational interfaces for co-designing a Subjective Bayesian Network and using it for situational understanding.

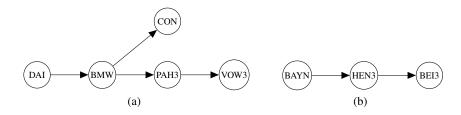## 2 Human-Machine Coalitions for Situational Understanding

Fig. 1. German automotive (a) and cosmetic (b) company dependency networks provided as input

	Company	Comment
BAYN	Bayer	Pharmaceutical company
BEI3	Beiersdorf	Cosmetic company
BMW	BMW	Automotive manufacturer
CON	Continental	Tyre manufacturer
DAI	Daimler	Automotive manufacturer
HEN3	Henkel	Cosmetic company
PAH3	Porsche	Automotive manufacturer
VOW3	Volkswagen	Automotive manufacturer

Table 1. Companies considered from the German stock market in Figures 1-2
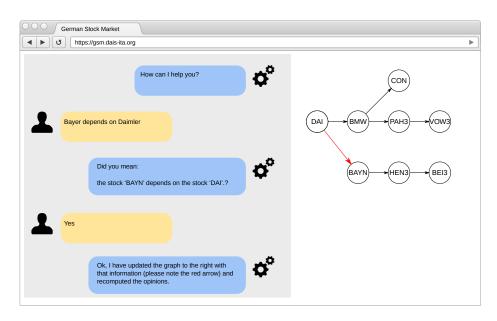
Fig. 2. Mockup depicting the action of updating a dependency network through our proposed conversational interface. Other speech acts envisaged for such an interface include "Explain dependencies. . . " and "What happens if . . . "

Let us suppose you are an advisor for investors who want to enter the German stock market. For brevity, let us suppose that a colleague has provided the two high-level dependency networks depicted in Figure 1, showing on the one hand dependencies between Daimler, BMW, Continental, Porsche, and Volkswagen (automotive companies); and on the other hand dependencies between Bayer, Henkel, and Beiersdorf (cosmetic companies). These dependencies suggest that the stock prices of those companies are linked such that a significant variation of the stock price of Daimler will influence a variation in the stock price of BMW.

Let us suppose you have the privilege of using our conversational interface for interacting with such dependencies networks, see Figure 2. Among other activities, such as explaining the dependencies and exploring what-if scenarios such a conversational interface would allow you also to express additional information, in particular that there is a dependency between Bayer and Daimler thus *de facto* providing a machine with domain knowledge unavailable before. This enables the human user to, therefore, act as an *oracle*, contributing relevant information to the machine agent based on their wider knowledge of the domain in question. Indeed, Daimler and Bayer are regularly traded by over-the-counter (OTC) list shares[8] such as INTL FCStone Financial.[9]

---

[8] OTC trades refers to stock trades via a dealer network.
[9] https://goo.gl/lTruuv (on 4th May 2017).

3 Reasoning under Uncertainty with Limited Data

3.1 Dealing with Uncertainty: Subjective Logic

Subjective logic is a formalism for reasoning under uncertain probabilistic information [10]. It expands the notion of a probability value to a distribution of possible probabilities. This paper considers binary variables such as X that can take on the value of true or false, i.e., $X = \mathfrak{x}$ or $X = \bar{\mathfrak{x}}$. The value of X does change over different instantiations, and there is an underlying ground truth value for the probability $p_X(x)$ of taking on the value in the domain $\mathbb{X} = \{\mathfrak{x}, \bar{\mathfrak{x}}\}$. In general, the variable can take on one of K mutually exclusive values.

A subjective opinion can be formed by directly observing N_{ins} independent instantiations of X. If over these instantiations, n_x times $X = \mathfrak{x}$, $n_{\bar{x}} = N_{ins} - n_x$ times $X = \bar{\mathfrak{x}}$ and assuming an uninformative uniform prior, then the posterior knowledge of the ground truth outcome probability of X is known to follow the beta distribution

$$f_{\boldsymbol{\beta}}(p_x|\omega_X) = \frac{1}{\beta(\alpha_x, \alpha_{\bar{x}})} p_x^{\alpha_x - 1}(1 - p_x)^{\alpha_{\bar{x}} - 1} \tag{1}$$

for $0 \leq p_x \leq 1$, where $\beta(\cdot)$ is the beta function and the beta parameters $\boldsymbol{\alpha} = [\alpha_x, \alpha_{\bar{x}}] = [n_x + 1, n_{\bar{x}} + 1]$ are one particular representation of the opinion ω_X. The opinion ω_X in belief space is a tuple of belief $b_X = \frac{n_x}{s_X}$, disbelief $d_X = \frac{n_{\bar{x}}}{s_X}$ and uncertainty $u_X = \frac{2}{s_X}$, where $s_X = \alpha_x + \alpha_{\bar{x}}$ is the Dirichlet strength. Therefore, a tuple $\langle b_X, d_X, u_X \rangle$ identifies a point in a 3D space. However, since the belief masses are positive and sum up to one, such a 3D space can be flattened into a 2D triangle, as depicted in Figure 3. Following [10, p. 49] we can partition the 2D space of subjective logic opinions for (lossy) representation using fuzzy natural language terms such as "High Confidence" and "Very Likely". Such terms can be made even more consumable for human users when embedded within larger natural language sentences such as: "When BAYN stock price changes, there is *high confidence* that HEN3 stock price is *very likely to change*" that can summarise the subjective opinion $\langle 0.8, 0.1, 0.1 \rangle$.

In this paper, it will be convenient to represent the subjective opinion ω_X by the mean and Dirichlet strength of the corresponding beta distribution. The mean represents the projected probability that converts the opinion into the pignistic probabilities, and is given by

$$P_X(\mathfrak{x}) = \frac{\alpha_x}{s_X} \quad \text{and} \quad P_X(\bar{\mathfrak{x}}) = \frac{\alpha_{\bar{x}}}{s_X}. \tag{2}$$

The variance of the corresponding beta distribution,

$$\sigma_X^2 = \frac{P_X(\mathfrak{x}) P_X(\bar{\mathfrak{x}})}{s_X + 1}, \tag{3}$$

is a function of the projected probabilities and Dirichlet strength of the subjective opinion. This expression is used in the experiments to predict the root mean squared error between the projected probability $P_X(x)$ and the actual ground truth $\rho_X(x)$. Subjective opinions naturally extend to subjective conditional opinions, where for example, the opinion for X conditioned on Y and Z is interpreted as the set $\{\omega_{X|y,z} : y \in \mathbb{Y}, z \in \mathbb{Z}\}$, and $\omega_{X|y,z}$ represents the effective number of times that $X = \mathfrak{x}$ or $X = \bar{\mathfrak{x}}$ when $Y = y$ and $Z = z$ while jointly observing X, Y, and Z.

Confidence

⟨0,0,1⟩

No Confidence

5E

5D

Low

6D   4D

6C   5C   4C

7C   3C   Some

7B   6B   5B   4B   3B

8B   2B   High

9A 8A   7A   6A   5A   4A   3A   2A 1A   Total

⟨0,1,0⟩   ⟨1,0,0⟩

Absolutely Not / Very Unlikely / Unlikely / Somewhat Unlikely / Chances about Even / Somewhat Likely / Likely / Very Likely / Absolutely
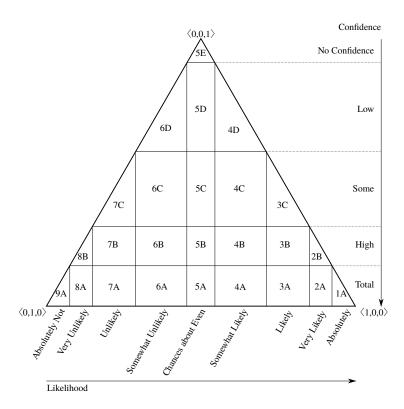
Likelihood

Fig. 3. Subjective Logic 2D triangle and areas for fuzzy labels, adapted from [10, p. 49].

3.2 Dealing with Limited Data: Subjective Bayesian Network

The Subjective Bayesian network (SBN) was first proposed in [9], and it is an uncertain Bayesian network where the conditionals are subjective opinions instead of dogmatic probabilities. In other words, the conditional probabilities are known within a beta distribution. A SBN reflects the knowledge about a Bayesian network when limited historical data is used to learn the conditionals. The inference in SBN leads to an opinion about the marginal probability of all the unobserved variables conditioned on the values of the observed variables. While different types of SBNs were discussed in [9], this paper focuses on the type that uses the beta distribution interpretation of the subjective opinion to compute uncertainty. This section reviews subjective belief propagation (SBP), which was introduced for trees in [12] and extended for singly-connected networks in [13] for this class of SBNs.

SBP extends the Belief Propagation (BP) inference method of Pearl [19]. In BP, π- and λ-messages are passed from parents and children, respectively, to a node, i.e., variable. The node uses these messages to formulate the inferred marginal probability of the corresponding variable. The node also uses these messages to determine the π- and λ-messages to send to its children and parents, respectively. In SBP, the π- and λ-

messages are subjective opinions characterized by a projected probability and Dirichlet strength.

The SBP formulation approximates output messages as beta-distributed random variables using the methods of moments and a first-order Taylor series approximation to determine the mean and variance of the output messages in light of the beta-distributed input messages. The details of the derivations are provided in [12, 13]. Given a node X with m parents U_i for $i = 1, \ldots, m$, the subjective opinions of the π-messages sent to X are characterized by the projected probabilities $\pi_{U_i,X}(x)$ and Dirichlet strengths $s_{\pi_{U_i,X}}$. Likewise, given that X has k children Y_j for $j = 1, \ldots, k$, the subjective opinions of the λ-messages sent to X are characterised by the projected probabilities $\lambda_{U_i,X}(x)$ and Dirichlet strengths $s_{\lambda_{U_i,X}}$. Node X processes these opinions to form the fused π opinion

$$\pi_X(x) = \sum_{u_1,\ldots,u_m} P(x|u_1,\ldots,u_m) \prod_{i=1}^{m} \pi_{U_i,X}(u_i), \tag{4}$$

$$s_{\pi_X} = \frac{\pi_X(x)(1 - \pi_X(x))}{\sigma_{\pi_X}^2} - 1, \tag{5}$$

where the variance $\sigma_{\pi_X}^2 = V_{\pi_X} - \pi_X^2(x)$,

$$V_{\pi_X} = \sum_{u_1,\ldots,u_m} \sum_{u_1',\ldots,u_m'} g(x,x;u_1,\ldots,u_m;u_1',\ldots,u_m') \cdot \prod_{i=1}^{m} h(u_i,u_i'), \tag{6}$$

$$g(x,x';u_1,\ldots,u_m;u_1',\ldots,u_m') = p_{x|u_1\ldots u_m} p_{x|u_1'\ldots u_m'} + (-1)^{x \neq x'} \delta_{\mathbf{u},\mathbf{u}'} \frac{p_{x|u_1\ldots u_m}(1 - p_{x|u_1\ldots u_m})}{s_{X|u_1\ldots u_m} + 1}, \tag{7}$$

where $\mathbf{u}$ is an arbitrary joint assignment of the variables $U_1, \ldots, U_m$,

$$\delta_{\mathbf{u},\mathbf{u}'} = \begin{cases} 1, \text{ if } u_j = u_j', \text{ for } j = 1, \ldots, m \\ 0, \text{ otherwise} \end{cases}$$

is the Kronecker delta function, and

$$h_\pi(u_i,u_i') = \pi_{U_i,X}(u_i)\pi_{U_i,X}(u_i') + (-1)^{u_i \neq u_i'} \frac{\pi_{U_i,X}(u_i)(1 - \pi_{U_i,X}(u_i))}{s_{\pi_{U_i,X}} + 1}. \tag{8}$$

The fused λ-message is

$$\lambda_X(x) = \alpha_\lambda \prod_{j=1}^{k} \lambda_{Y_j,X}(x), \tag{9}$$

$$s_{\lambda_X} = \left(\sum_{j=1}^{k} \frac{\lambda_X(\mathfrak{x})\lambda_X(\bar{\mathfrak{x}})}{\lambda_{Y_j,X}(\mathfrak{x})\lambda_{Y_j,X}(\bar{\mathfrak{x}})} \frac{1}{s_{\lambda_{Y_j,X}} + 1} \right)^{-1} - 1,$$

where α_λ is a normalizing constant so that $\lambda_X(x)$ sums to one over its domain $\mathbb{X}$.

The π and λ-opinions are fused to determine the marginal opinion for node X:

$$P_X(x|o) = \alpha_f \pi_X(x)\lambda_X(x) \,, \tag{10}$$

$$s_X = \left(\frac{P_X(\mathfrak{r})P_X(\bar{\mathfrak{r}})}{\pi_X(\mathfrak{r})\pi_X(\bar{\mathfrak{r}})} \frac{1}{s_{\pi_X}+1} + \frac{P_X(\mathfrak{r})P_X(\bar{\mathfrak{r}})}{\lambda_X(\mathfrak{r})\lambda_X(\bar{\mathfrak{r}})} \frac{1}{s_{\lambda_X}+1} \right)^{-1} - 1,$$

where α_f is also a normalizing constant.

The opinion for the message that node X sends to parent U_i is

$$\lambda_{X,U_i}(u_i) = \alpha_b \sum_x \lambda_X(x) \sum_{\{u_1,\dots,u_m\}\backslash\{u_i\}} P(x|u_1,\dots,u_i,\dots,u_m) \cdot \prod_{j\neq i} \pi_{U_j,X}(u_j), \tag{11}$$

$$s_{\lambda_{X,U_i}} = \frac{\lambda_{X,U_i}(u_i)(1-\lambda_{X,U_i}(u_i))}{\sigma^2_{\lambda_{X,U_i}}} - 1, \tag{12}$$

where

$$\sigma^2_{\lambda_{X,U_i}} = \alpha_b^2 \left(\lambda^2_{X,U_i}(\bar{x})\sigma^2_{uu} + \lambda^2_{X,U_i}(x)\sigma^2_{\bar{u}\bar{u}} - 2\lambda_{X,U_i}(x)\lambda_{X,U_i}(\bar{x})\sigma^2_{u\bar{u}} \right), \tag{13}$$

$$\sigma^2_{zv} = \sum_x \sum_{x'} h_\lambda(x,x') \sum_{\{u_1,\dots,u_m\}\backslash\{z\}} \sum_{\{u'_1,\dots u'_m\}\backslash\{v\}}$$
$$g(x,x';u_1,\dots,z,\dots,u_m;u'_1,\dots,v,\dots,u'_m)\prod_{j\neq i} h_\pi(u_j,u'_j), \tag{14}$$

and

$$h_\lambda(x,x') = \lambda_X(x)\lambda_X(x') + (-1)^{x\neq x'} \frac{\lambda_X(x)(1-\lambda_X(x))}{s_{\lambda_X}+1}, \tag{15}$$

and α_b is a normalizing constant.

Finally, the opinion message sent to the children of X are

$$\pi_{X,Y_j}(x) = \alpha_\pi \prod_{i\neq j} \lambda_{Y_i,X}(x)\pi_X(x), \tag{16}$$

$$s_{\pi_{X,Y_j}} = \left(\frac{\pi_{X,Y_j}(\mathfrak{r})\pi_{X,Y_j}(\bar{\mathfrak{r}})}{\pi_X(\mathfrak{r})\pi_x(\bar{\mathfrak{r}})} \frac{1}{s_{\pi_X}+1} + \sum_{i\neq j} \frac{\pi_{X,Y_j}(\mathfrak{r})\pi_{X,Y_j}(\bar{\mathfrak{r}})}{\lambda_{Y_i,X}(\mathfrak{r})\lambda_{Y_i,X}(\bar{\mathfrak{r}})} \frac{1}{s_{\lambda_{Y_i,X}}+1} \right)^{-1} - 1,$$

where α_π is a normalizing constant.

The equations for the projected probability updates in SBP mirror the updated equations in standard belief propagation due to the first-order Taylor approximation. Actually, the normalizing constants α_λ and α_β are superfluous in standard belief propagation, but necessary in SBP so that the λ message are proper subjective opinions. In short, SBP provides the same answer as belief propagation in the mean value. The difference is that SBP also provides a quantification of the uncertainty through the Dirichlet strength. On a technical note, SBP will actually increase the Dirichlet strength as computed in the update equations to ensure that all belief values are non-negative. We refer the interested reader to [12, 13] for more details. Finally, the information flow in SBP is exactly the same as in belief propagation. For the sake of comparison, a node can send a message to one particular neighbor once it receives messages from all of its other neighbors.

4 Experimentation

4.1 Methodology

SBNs can learn a model of the domain with a very limited number of observations; however, the inferred opinions through such a network will become more certain as the number of observations increases. To measure how well these models can be learned with limited data and measure the uncertainty associated with the inferences, we build gold standard models, which are Bayesian networks that are generated using a much larger number of observations. The gold standard models are Bayesian networks with completely certain conditional probabilities that we treat as the ground truth.

For structure learning of the gold standard models, we used the well-known K2 algorithm [17]. The K2 algorithm is used to learn the best structure of a singly connected Bayesian network to represent the interactions between the random variables. The resulting network serves as a surrogate for a subject matter expert who would use their background knowledge to create the network structure, for example, via the conversational interface (see Figure 2). Further discussion on this topic is provided in the conclusion of the paper. Then, the conditional and marginal probabilities at each node of the network are calculated in the traditional manner using the entire available data.

We use real data to evaluate the quality of the uncertainty (or Dirichlet strength) in the subjective opinions inferred by SBP to represent the actual spread between the corresponding 'projected' and 'ground truth' probabilities that are well captured by the gold standard models. The full data is then divided into non-overlapping segments of N_{ins} instantiations (i.e., observations). Each segment represents the sparse data that would actually be available to train a SBN. A SBN is trained for each segment, and the set of exterior nodes, i.e., nodes with one single neighbour (either a parent or child), are considered to be observed. For each combination of possible values for these exterior nodes, the marginal opinions for the interior nodes are inferred by SBP. Likewise, to establish the ground truth, the marginal probabilities are inferred by standard belief propagation using the underlying gold standard Bayesian network for the same values of the observed exterior nodes. Then, the marginal opinions and ground truths for all interior nodes are determined over all combinations of observed values and non-overlapping segments. Finally, the uncertainty of the marginal opinions is evaluated.

To evaluate the quality of the derived uncertainty, the actual root mean squared error (RMSE) between the projected and ground truth probabilities is calculated. Next, the predicted RMSE is computed without knowledge of the ground truth, as the square root of the average variance predicted from the opinions via (3). The similarity between the actual and predicted RMSE is one way to establish the quality of the uncertainty in the subjective opinions that are to characterise the spread between the projected and actual probabilities.

An even more precise method to determine the quality of the uncertainty characterisation is to establish γ-confidence intervals from the opinions to capture the fraction of γ ground truths within these intervals. One then tabulates the fraction of times that the actual ground truth falls within the confidence interval. This is done for various values of $\gamma \in [0, 1]$, and the plot of the actual $\hat{\gamma}$ and the desired γ should follow a straight line as it should be the case that $\hat{\gamma} \approx \gamma$. A more detailed discussion can be found in [14].
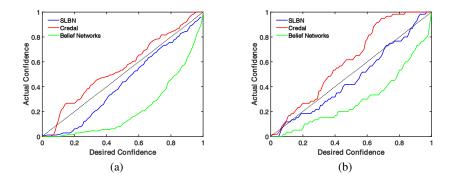
Fig. 4. Comparing SBN against Belief Networks and Credal with $N_{train} = 10$ (over 365) (a) and $N_{train} = 30$ (over 365) (b) for the German stock exchange data. Best closest to the diagonal.

The quality of the inferred subjective opinion ω_X should be judged on how well its expression of uncertainty captures the spread between its projected probability and the actual ground truth probability.

We compare the performance of SBP against previous methods for reasoning over uncertain probabilistic networks. Namely, we consider credal networks and belief networks, which are summarized below:

Credal Networks: A credal network over binary random variables extends a BN by replacing single probability values with closed intervals representing the possible range of probability values. The extension of Pearl's message-passing algorithm by the 2U algorithm for credal networks is described in [23]. This algorithm works by determining the maximum and minimum value (an interval) for each of the target probabilities based on the given input intervals. It turns out that these extreme values lie at the vertices of the polytope dictated by the extreme values of the input intervals. As a result, the computational complexity grows exponentially with respect to the number of parents nodes. For the sake of comparison, we assume that our subjective network elicited from the given data corresponds to a credal network in the following way: if $\omega_x = [b_x, b_{\bar{x}}, u_X]$ is a subjective opinion on the probability p_x, then we have $[b_x, b_x + u_X]$ as an interval corresponding to this probability in the credal network. It should be noted that this mapping from the Beta distribution to an interval is consistent with past studies of credal networks [15].

Belief Networks: In [20], Smets introduced a computationally efficient method to reason over networks via Dempster-Shafer theory. It is an approximation of a valuation-based system. Namely, a (conditional) subjective opinion $\omega_X = [b_x, b_{\bar{x}}, u_X]$ from our SBN obtained from data is converted to the following belief mass assignment: $m(x) = b_x$, $m(\bar{x}) = b_{\bar{x}}$ and $m(x \cup \bar{x}) = u_X$. (Note that in the binary case, the belief function overlaps with the belief mass assignment). The method exploits the disjunctive rule of combination (DRC) to compose beliefs conditioned on the Cartesian product space of the binary power sets. This enables both forward propagation and backward propagation after inverting the belief conditionals via the generalized Bayes' theorem (GBT). By operating in the Cartesian product space of the binary power sets, the computational

complexity grows exponentially with respect to the number of parents, similar to the 2U algorithm for credal sets and our SBP method.

4.2 German Stock Exchange Predictions

Let us consider the case where a machine learning system is used to mine data from the German Stock Market, Börse Frankfurt. To simplify the scenario, let us consider a binary variable per each company listed in Börse, where such a variable is $true$ if there is a significant increase (i.e. $+0.5\%$) in the company's stock value over a day, and $false$ otherwise. Let us then suppose that a well-known off-the-shelf algorithm for structure learning of dependencies among selected variables, such as K2 [17], has been used. Using such an algorithm, the dependency networks highlighted in Figures 1(a) and 1(b) are derived. Table 1 explains the variables used in the dependency networks.

Figure 1(a) shows how there is a dependency between Daimler stock variations and BMW; between BMW and Porsche; between Porche and Volkswagen (all automotive manufacturers); and between BMW and Continental, a tyre manufacturer. Similarly, Figure 1(b) depicts the dependencies between Bayer—a pharmaceutical company— and Henkel—a company producing a variety of chemical products including cosmetics ingredients; and between Henkel and Beiersdorf, cosmetic companies. Those dependencies are far from being a surprise, given that they are companies working in similar, or related, segments of the market. These two networks have then been merged to produce the single network given in Figure 2.

	$N_{train} = 10$ (over 365)			$N_{train} = 30$ (over 365)		
	SBN	**Credal**	**Belief Net**	**SBN**	**Credal**	**Belief Net**
Actual RMSE	**0.124**	0.198	0.176	**0.047**	0.062	0.075
Predicted RMSE	0.101	0.187	0.132	0.049	0.089	0.061

Table 2. Error for the German stock exchange dataset. Gold standard trained with $N_{train} = 365$. Best results in bold.

The gold standard Bayesian Network is obtained by using all available data for (365 days) to determine the conditional probabilities. Then N_{train} days were used to generate $floor(365/N_{train})$ SBNs. Binary values were generated for the three nodes that have one edge, and the marginal probabilities (ground truth) and marginal opinions were generated via belief propagation and subjective belief propagation over the Bayesian and SBNs, respectively. Table 2 lists actual and predicted RMSE for the different approaches using different amounts of observations. It indicates that SBN achieves pretty good error rate even with 10 days of observations (sample size 2.74%) and the error decreases to 0.05 when 30 days of data is used (sample size 8.21%). Figure 4 shows the ratio of the times the ground truth falls within the bounds—set at various significance levels—when building SBNs over 10 and 30 days. Our results indicate that SBN

can capture the uncertainty more accurately than Credal networks and Belief Networks. Especially, when $N_{train} = 30$, confidence level of the SBN is around the desired one, i.e., diagonal on the figures. Moreover, Table 2 lists actual and predicted RMSE for our approach and the benchmark approaches when different amounts of observations are used. SBN is consistently able to predict an accurate RMSE.
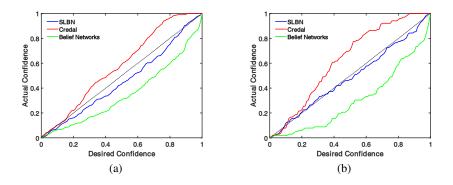

Fig. 5. Comparing SBN against Belief Networks and Credal with $N_{train} = 10$ (over 536) (a) and $N_{train} = 30$ (over 536) (b) for the Istanbul stock market data. Best closest to the diagonal.

4.3 Istanbul Stock Market Predictions

We also considered the dataset first derived in [1],[10] which considers stock exchange returns for several indexes, including those listed in Table 3. It is quite straightforward to derive a dependency network such as the one given in Figure 6 between those indexes.

Standard & Poor's 500 index includes leading US companies and captures approximately 80% of available US market capitalisation. Those companies are trading heavily with the rest of the world, including Asia, and notably Japan; and with South America, notably Brazil. Moreover, Brazil's economy heavily affects the MSCI Emerging Markets Index. According to the Foreign Trade figures from the United States Census Bureau, within Europe, the US has a strong commercial partnership with Germany,[11] much stronger than with the second strongest commercial ally, namely the UK.[12] Therefore, it is straightforward to see how the return for Standard & Poor's has a significant statistical dependence with the German Stock Market. Moreover, with 15% of the imports coming from Germany, the UK economy is also significantly dependent on the German market[13] (instead Germany imports mostly from the Netherlands and exports

---

[10] https://goo.gl/XzAZUX (on 4th May 2017)
[11] https://goo.gl/8PdBll (on 4th May 2017)
[12] https://goo.gl/n2V89z (on 4th May 2017)
[13] https://goo.gl/v1tXD4 (on 4th May 2017)

mostly to the US).[14] Finally, the MSCI European Index return is heavily affected by Germany, the first economy in the European Union.
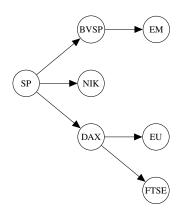

Fig. 6. Istanbul Stock Exchange Data Set [1] dependency network.

	Comment
SP	Standard & Poor's 500 Index Return.
DAX	Germany Stock Market Return
FTSE	UK Stock Market Return
NIK	Japan Stock Market Return
BVSP	Brazil Stock Market Return
EU	MSCI European Index Return
EM	MSCI Emerging Markets Index Return

Table 3. Indexes considered from the Istanbul Stock Exchange Data Set [1] in Figure 6.

We also used this dataset of 536 entries to evaluate our approach using different amounts of observed data. Table 4 lists actual and predicted RMSE for our approach and the benchmark approaches when different amount of observations are used. It shows that SBN consistently predicts the error when trained either over 10 or 30 days, unlike the two other methods.

Figure 5 demonstrates our results in terms of γ-confidence intervals. Even for data of 10 days, the confidence for inferences with SBN only slightly diverges from the

---

[14] https://goo.gl/ZPJLdR (on 4th May 2017)

	$N_{train} = 10$ (over 536)			$N_{train} = 30$ (over 536)		
	SBN	**Credal**	**Belief Net**	**SBN**	**Credal**	**Belief Net**
Actual RMSE	**0.131**	0.170	0.172	**0.088**	0.089	0.104
Predicted RMSE	0.146	0.223	0.124	0.093	0.140	0.068

Table 4. Error for the Istanbul stock exchange dataset. Gold standard trained with $N_{train} = 536$. Best results in bold.

desired confidence levels. When training data is increased to 30 days, the confidence interval for SBN approximate the desired one very closely. Again, in this dataset, the best performance belongs to SBN in terms of γ-confidence intervals.

5 Empirical Evaluation of User Co-Design of Subjective Bayesian Networks

We selected three people among staff members and students at Cardiff University on the basis of their expertise with computational models of uncertainty. The first person should be considered an expert in probabilistic methods of inference; the second is a mature PhD student who spent many years in software companies and with a basic understanding of probabilistic methods of inference; the third is a sociologist with little or no experience with probabilistic methods of inference. We prepared then a written briefing—see Appendix A—and an example of the mockup—analogous to Figure 2. We then asked them to consider six cases, each of which refers to a specific command for an hypothetical conversational interface, namely:

1. explain, i.e. summarise the results;
2. explain ‹ company ›, i.e. describe the dependencies for a specific company, e.g. Henkel;
3. explain in detail ‹ company ›, i.e. describe *in detail*, i.e. with information on the probabilistic model—thus oriented to a more specialist audience—the dependencies for a specific company;
4. what happens if both ‹ company1 › and ‹ company2 › stock prices change?, i.e. exploring what-if scenarios;
5. what happens in detail if both ‹ company1 › and ‹ company2 › stock prices change?, i.e. exploring what-if scenarios for a more specialist audience;
6. ‹ company1 › depends on ‹ company2 ›, i.e. add a new dependency between two companies (cf. Figure 2).

For each of those cases, we ask the participants to answer the following questions from the Subjective Usability Scale (SUS) questionnaire [4]. The numeric responses ranged between 1: Strongly Disagree, and 5: Strongly Agree. We asked the following questions:

Q1: I think that I would like to use this command frequently
Q2: I found the answer unnecessarily complex
Q3: I thought that the interaction was quite natural
Q4: I think that I would need the support of a technical person to be able to understand this interaction

At the end of the experiment, using the same numeric scale, participants were asked to answer additional questions from the SUS questionnaire:

SQ1: I think that I would like to use the conversational interface frequently
SQ2: I found the application unnecessarily complex
SQ3: I thought that overall the interaction was quite natural
SQ4: I think that overall I would need the support of a technical person to be able to understand this interaction
SQ5: I found the various commands well integrated
SQ6: I thought there was too much inconsistency in the application
SQ7: I would imagine that motivated users would learn to use the application very quickly
SQ8: I found the commands very cumbersome
SQ9: I think I would need to learn a lot of things before I could get going with the application

At any time, participants could write notes and comments in natural language to share with us. Moreover, the participants were reminded that their participation was entirely voluntary and that they were free to abandon the evaluation at any time with no need for explanations. This project received ethical approval from the School of Computer Science & Informatics Research Ethics Group.

5.1 Results

Explain Figure 7 shows the mockup of interaction for the explain command and Figure 12 depicts the results of the evaluation using error bars.
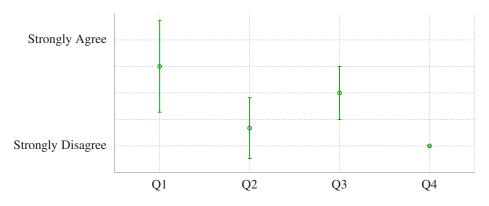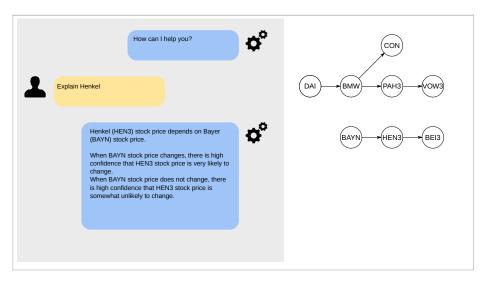

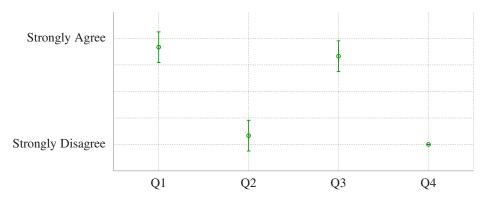**Fig. 7.** Mockup for the explain command.

Fig. 8. Results of the evaluation for the explain command: average and standard deviation of the aggregated received answers. Collected data are not normally distributed. Q1: I think that I would like to use this command frequently. Q2: I found the answer unnecessarily complex. Q3: I thought that the interaction was quite natural. Q4: I think that I would need the support of a technical person to be able to understand this interaction

Explain ‹ Company › Figure 9 shows the mockup of interaction for the explain ‹ company › command and Figure 10 depicts the results of the evaluation using error bars.

Fig. 9. Mockup for the explain ‹ company › command.

Fig. 10. Results of the evaluation for the explain ‹ company › command: average and standard deviation of the aggregated received answers. Collected data are not normally distributed. Q1: I think that I would like to use this command frequently. Q2: I found the answer unnecessarily complex. Q3: I thought that the interaction was quite natural. Q4: I think that I would need the support of a technical person to be able to understand this interaction

Explain in detail ‹ Company › Figure 11 shows the mockup of interaction for the explain in detail ‹ company › command and Figure 12 depicts the results of the evaluation using error bars.
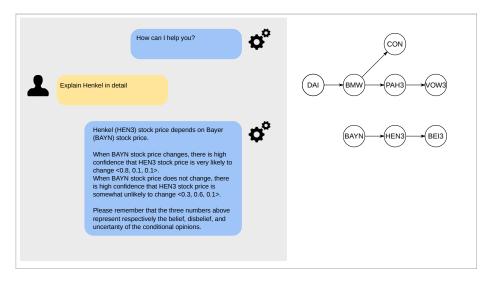

Fig. 11. Mockup for the explain in detail ‹ company › command.
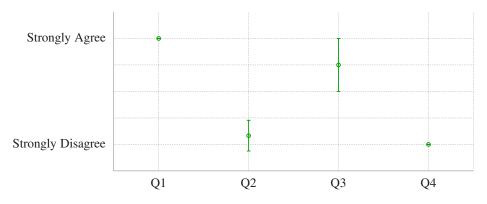

Fig. 12. Results of the evaluation for the explain in detail ‹ company › command: average and standard deviation of the aggregated received answers. Collected data are not normally distributed. Q1: I think that I would like to use this command frequently. Q2: I found the answer unnecessarily complex. Q3: I thought that the interaction was quite natural. Q4: I think that I would need the support of a technical person to be able to understand this interaction

What happens if . . .   Figure 13 shows the mockup of interaction for the what happens if . . . command and Figure 14 depicts the results of the evaluation using error bars.
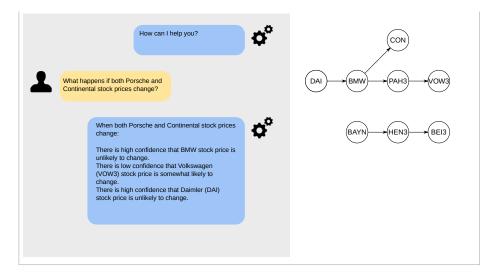

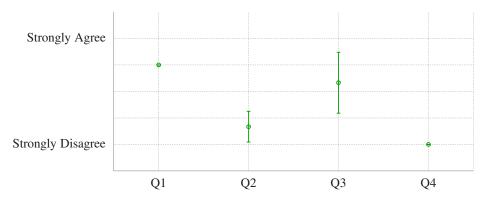**Fig. 13.** Mockup for the what happens if . . . command.

Fig. 14. Results of the evaluation for the what happens if . . . command: average and standard deviation of the aggregated received answers. Collected data are not normally distributed. Q1: I think that I would like to use this command frequently. Q2: I found the answer unnecessarily complex. Q3: I thought that the interaction was quite natural. Q4: I think that I would need the support of a technical person to be able to understand this interaction

What happens in detail if ... Figure 15 shows the mockup of interaction for the what happens in detail if ... command and Figure 16 depicts the results of the evaluation using error bars.
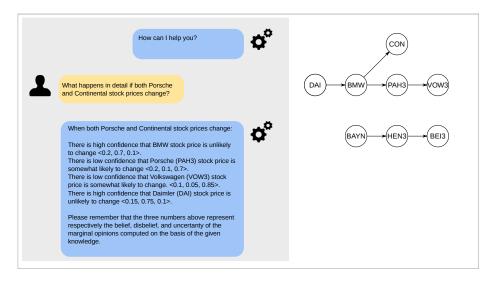

Fig. 15. Mockup for the what happens in detail if ... command.
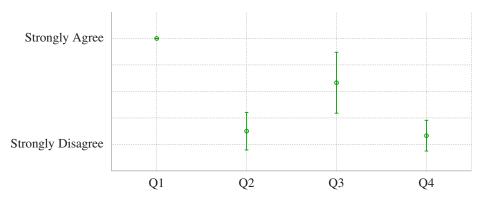

Fig. 16. Results of the evaluation for the what happens in detail if ... command: average and standard deviation of the aggregated received answers. Collected data are not normally distributed. Q1: I think that I would like to use this command frequently. Q2: I found the answer unnecessarily complex. Q3: I thought that the interaction was quite natural. Q4: I think that I would need the support of a technical person to be able to understand this interaction

‹ **company1** › **depends on** ‹ **company2** › Figure 2 shows the mockup of interaction for the ‹ company1 › depends on ‹ company2 › and Figure 17 depicts the results of the evaluation using error bars.

Fig. 17. Results of the evaluation for the ‹ company1 › depends on ‹ company2 › command: average and standard deviation of the aggregated received answers. Collected data are not normally distributed. Q1: I think that I would like to use this command frequently. Q2: I found the answer unnecessarily complex. Q3: I thought that the interaction was quite natural. Q4: I think that I would need the support of a technical person to be able to understand this interaction.

Overall evaluation Figure 18 depicts the results of the overall evaluation using error bars.

5.2 Summary of Evaluation

Considering the results summarised in Figures 8–18, we can conclude that in general our participants appreciated the idea of having a conversational interface for situational understanding. Indeed, the three participants mostly agreed with the positive statements and mostly disagreed with the negative statements in the questionnaire. There are, however, avenues for improvement.

As per the explain command, it is likely that users will use it only once, as their continuous interaction with the interface will lead them to familiarise with the graph-based interface. On this note, it would be useful to have multi-modal interaction, with also conditional probabilities tables associated with the graph with natural language labels or also subjective logic opinions. Also, looking at the collected data, it seems that providing the participants with additional information in textual format lead to a less "natural" interaction with the application. This might be correlated to the chosen scenario, or also to possible difficulties to understand SBNs: we will continue our investigation in the future taking these options into due consideration.

Moreover, there is a missing command in the list, namely a help, which is probably the first command a fresh user will ask. Finally, it is unclear whether the supplied com-
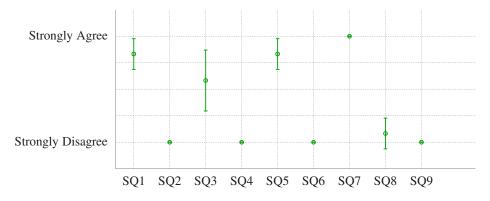
Fig. 18. Results of the evaluation for the what happens in detail if ... command: average and standard deviation of the aggregated received answers. Collected data are not normally distributed. SQ1: I think that I would like to use the conversational interface frequently. SQ2: I found the application unnecessarily complex. SQ3: I thought that overall the interaction was quite natural. SQ4: I think that overall I would need the support of a technical person to be able to understand this interaction. SQ5: I found the various commands well integrated. SQ6: I thought there was too much inconsistency in the application. SQ7: I would imagine that motivated users would learn to use the application very quickly. SQ8: I found the commands very cumbersome. SQ9: I think I would need to learn a lot of things before I could get going with the application.

mands are sufficient for supporting all the tasks of an analyst: this specific point will require further analysis with realistic scenarios.

6 Conclusion

In this paper we presented a methodology to exploit human-machine coalitions for situational understanding, i.e., the ability to relate relevant information with dependencies and form logical conclusions as well as identifying gaps in information. This process requires the ability to reason inductively, for which one must exploit the machines' ability to learn from data, although important phenomena are often rare in occurrence, severely limiting the availability of instance data and hence the applicability of many machine learning approaches.

To this end, we discussed at length the benefits of SBNs, especially when training with sparse data, and in Section 4 we showed that they are superior to previous methods to reason over uncertain probabilistic networks, Credal networks and Belief Networks. In the future, we plan to compare SBN to other probabilistic models for dependencies such as Maximum Likelihood Estimation of an Alternating Renewal Process [18, 21]. We considered two different datasets both related to the financial domain, but clearly SBNs can directly be applied to other datasets. We are working towards inference over general directed acyclic graphs as they characterise any joint probability distribution.

We also discussed the role that would be played by humans in situational understanding. Differently from other approaches aimed at explaining high-dimensional,

multivariate feature spaces and dependencies to humans, e.g. [22], we believe a conversational interface like the one depicted in Figure 2 can provide the right level of interactivity in the coalition of humans and machines for situational understanding. We are currently developing the first prototype of this conversational interface, and we are focusing on three major capabilities: (1) the ability to explain the dependencies (e.g., "When Bayer stock price changes, it is likely that..."); (2) the ability of what-if reasoning (e.g., "If Bayer stock price changes, then..."); and (3) as shown in Figure 2, the ability to modify the dependency network. The preliminary evaluation we discussed in Section 5 suggests that conversational interfaces are a positive way to interact with complex decision making systems such as drawing inferences using Bayesian Networks. However, additional interfaces, including extending the graphical representation with conditional probabilities tables and enable their manipulation, need to be studied as they might suit some of the potential users. Moreover, evaluating the conversational interface in different case-studies might show which commands are mostly used, and eventually which ones need to implemented.

This opens a large spectrum of future work, including the ability to evaluate the human expertise and the quality of data. If a human user adds a dependency that is not supported by available data, it might suggest that the user has knowledge that is "out of scope" in the data and/or model. However, such an assertion may simply be erroneous or could indicate data quality issues such as data that are incomplete, biased, or corrupted.

References

1. Akbilgic, O., Bozdogan, H., Balaban, M.E.: A novel hybrid RBF neural networks model as a forecaster. Statistics and Computing 24(3), 365–375 (2014)
2. Braines, D., Thomas, A., Kaplan, L., Sensoy, M., Ivanovska, M., Preece, A.D., Cerutti, F.: Human-in-the-loop situational understanding via subjective Bayesian networks. In: The 5th International Workshop on Graph Structures for Knowledge Representation and Reasoning (GKR 2017) (2017)
3. Brannon, N., Seiffertt, J., Draelos, T., II, D.W.: Coordinated machine learning and decision support for situation awareness. Neural Networks 22(3), 316 – 325 (2009)
4. Brooke, J., et al.: Sus-a quick and dirty usability scale. Usability evaluation in industry 189(194), 4–7 (1996)
5. Chawla, N.V., Japkowicz, N., Kotcz, A.: Special issue on learning from imbalanced data sets. ACM Sigkdd Explorations Newsletter 6(1), 1–6 (2004)
6. Dhami, M.K., Mandel, D.R., Mellers, B.A., Tetlock, P.E.: Improving intelligence analysis with decision science. Perspectives on Psychological Science 10(6), 753–757 (2015)
7. Dostal, B.C.: Enhancing situational understanding through employment of unmanned aerial vehicle. Army Transformation Taking Shape: Interim Brigade Combat Team Newsletter 01-18 (2007)
8. Helbing, D.: Globally networked risks and how to respond. Nature 497(7447), 51–59 (2013)
9. Ivanovska, M., Jøsang, A., Kaplan, L., Sambo, F.: Subjective networks: Perspectives and challenges. In: Proc. of the 4th International Workshop on Graph Structures for Knowledge Representation and Reasoning. pp. 107–124. Buenos Aires, Argentina (2015)
10. Jøsang, A.: Subjective Logic: A Formalism for Reasoning Under Uncertainty. Springer (2016)
11. Kahneman, D.: Thinking, fast and slow. Macmillan (2011)

12. Kaplan, L., Ivanovska, M.: Efficient subjective Bayesian network belief propagation for trees. In: Intl. Conf. on Information Fusion (FUSION). pp. 1300–1307 (2016)
13. Kaplan, L., Ivanovska, M.: Efficient subjective bayesian network belief propagation for singly-connected graphs (2017), submitted to *International Journal of Approximate Reasoning*
14. Kaplan, L., Şensoy, M., Chakraborty, S., de Mel, G.: Partial observable update for subjective logic and its application for trust estimation. Information Fusion 26, 66–83 (2015)
15. Karlsson, A., Johansson, R., Andler, S.F.: An empirical comparison of Bayesian and credal networks for dependable high-level information fusion. In: Intl. Conf. on Information Fusion (FUSION). pp. 1–8 (2008)
16. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (2015)
17. Lerner, B., Malka, R.: Investigation of the K2 algorithm in learning bayesian network classifiers. Applied Artificial Intelligence 25(1), 74–96 (2011)
18. Lin, X., Moussawi, A., Korniss, G., Bakdash, J.Z., Szymanski, B.K.: Limits of risk predictability in a cascading alternating renewal process model. Scientific Reports 7(1), 6699 (2017)
19. Pearl, J.: Fusion, propagation, and structuring in belief networks. Artificial Intelligence 29(3), 241–288 (1986)
20. Smets, P.: Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. International Journal of Approximate Reasoning 9, 1– 35 (1993)
21. Szymanski, B.K., Lin, X., Asztalos, A., Sreenivasan, S.: Failure dynamics of the global risk network. Scientific reports 5, 10998 (2015)
22. Timmer, S.T., Meyer, J.J.C., Prakken, H., Renooij, S., Verheij, B.: A two-phase method for extracting explanatory arguments from Bayesian networks. International Journal of Approximate Reasoning 80(C), 475–494 (2017)
23. Zaffalon, M., Fagiuoli, E.: 2U: An exact interval propagation algorithm for polytrees with binary variables. Artificial Intelligence 106(1), 77–107 (1998)
24. Zhou, Z.H., Sun, Y.Y., Li, Y.F.: Multi-instance learning by treating instances as non-iid samples. In: Proceedings of the 26th Annual International Conference on Machine Learning. pp. 1249–1256. ACM (2009)

A Briefing Received by the Participants

A computer analysed the data of the German Stock Market Börse Frankfurt related to nine companies:

- Bayer, a pharmaceutical company
- Beiersdorf, a cosmetic company
- Henkel, a cosmetic company
- BMW, an automotive manufacturer
- Daimler, an automotive manufacturer
- Porsche, an automotive manufacturer
- Volkswagen, an automotive manufacturer
- Continental, a tyre manufacturer

In particular, the computer was programmed only to consider whether the closing value of a stock price was significantly different from same stock price at the closing time of the day before ($\pm 0.5\%$). And then the computer automatically derived possible dependencies between stocks.

Example The Bayer stock value at the closing time on 7th December 2016 was 90.10; at the closing time on 8th December 2017 it was 93.17, thus with a significant change of 3.4%.

Similarly, the computer also analyses the changes of all the other companies considered in this study, thus producing a large table like the following:

Company	07/12/16	08/12/16	09/12/16	. . .
Bayer	Stable	Changed	Changed	. . .
Beiersdorf	Stable	Changed	Stable	. . .
Henkel	Stable	Stable	Stable	. . .
. . .	. . .	. . .	. . .	. . .

On the basis of such a large table, and by employing *Machine Learning* procedures, the computer identifies dependencies between companies' stock values. An example of such a dependencies can be:

> When Bayer stock price changes, there is low confidence that Henkel stock price is unlikely to change.