2010

Surprise: An Alternative Qualitative Uncertainty Model

Zina Ibrahim
University of Windsor

Surprise: An Alternative Qualitative Uncertainty Model

by

Zina M. Ibrahim

A Dissertation

Submitted to the Faculty of Graduate Studies

through the School of Computer Science

in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy at the

University of Windsor

Windsor, Ontario, Canada

2010

Surprise: An Alternative Qualitative Uncertainty Model

by

Zina Ibrahim

APPROVED BY:

Dr. E. Neufeld, External Examiner

University of Saskatchewan

Dr. S. Ejaz Ahmed

Department of Mathematics & Statistics

Dr. Jessica Chen

School of Computer Science

Dr. Dan Wu

School of Computer Science

Dr. A. Tawfik, Advisor

School of Computer Science

Dr. A. Ngom, Advisor

School of Computer Science

Dr. J. Frank, Chair of Defense

Faculty of Graduate Studies

January 29, 2010

Declaration of Co-Authorship / Previous Publication

I. Co-Authorship Declaration

I hereby declare that this thesis incorporates material that is result of joint research, as follows:

This thesis also incorporates the outcome of a joint research undertaken under the supervision of Professors Ahmed Tawfik and Alioune Ngom. The collaboration is covered in Chapter 4, 5 and 6 of the dissertation. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of co-authors was primarily through the provision of advice when needed.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

II. Declaration of Previous Publication

This thesis includes six original papers that have been previously published/submitted for publication in peer reviewed conferences, as follows:

Thesis Chapter	Publication title/full citation	Publication status
4	Towards Qualitative Outlier Detection	Accepted
5	Surprise-Based Qualitative Probabilistic Networks. The 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2009) Pages 228-239. May 2009.	Published

4	A Surprise-based Qualitative Probability Calculus. Proceedings of the Twenty Second Conference of the Florida Artificial Intelligence Research Society (FLAIRS 2009). Pages 573-574, July 2009	Published
6	Qualitative Motif Detection in Gene Regulatory Networks. IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2009). Pages 124-129, November 2009	Published
6 & 7	Qualitative Hidden Markov Models for Classifying Gene Expression Data. Proceedings of AI-2008 Twenty-eigth SGAI International Conference on Artificial Intelligence (AI-2008). December 2008	Published
5	A Qualitative Hidden Markov Model for Spatio-temporal Reasoning. The European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007). Pages 707-718. November 2007.	Published

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyones copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis are fully acknowledged. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other Institution.

Abstract

This dissertation embodies a study of the concept of surprise as a base for constructing qualitative calculi for representing and reasoning about uncertain knowledge. This document presents two functions, κ^{++} and $\mathfrak{z}$, which construct qualitative ranks for events by obtaining the order of magnitude abstraction of the degree of surprise associated with them. The functions use natural numbers to classify events based their associated surprise and aim at providing a ranking that improves those provided by existing ranking functions. This in turn enables the use of such functions in an à la carte probabilistic system where one can choose the level of detail required to represent uncertain knowledge depending on the requirements of the application.

The proposed ranking functions are defined along with surprise-update models associated with them. The reasoning mechanisms associated with the functions are developed mathematically and graphically.

The advantages and expected limitations of both functions are compared with respect to each other and with existing ranking functions in the context of a bioinformatics application known as "reverse engineering of genetic regulatory networks" in which the relations among various genetic components are discovered through the examination of a large amount of collected data. The ranking functions are examined in this context via graphical models which are exclusively developed or this purpose and which utilize the developed functions to represent uncertain knowledge at various levels of details.

Dedication

To my parents,

for instilling in me what I needed for this journey...

To the love of my life,

for being himself...

Acknowledgements

This dissertation would not have taken its shape without a number of individuals who have seen me through it all; this page is theirs.

I must start by recognizing my parents, because their role extends beyond these sheets and is far greater than my Doctoral degree. These two individuals must be saluted for raising me as a strong, independent and a hopeful individual. I am indebted to them for always trusting my decisions and having faith in me; they showed me by example that will and faith are the essentials which can make the most difficult of tasks achievable. More importantly, I am thankful to them for teaching me how to do what I love and love what I do.

My advisors Dr. Ahmed Tawfik and Dr. Alioune Ngom, whom I dearly value and respect, will be individually addressed.

As I write this page on the morning of my 30^{th} birthday, I am remembering the eight years I spent having Dr. Ahmed Tawfik as my mentor. He grew from being my advisor to a counselor whose opinion I value and trust to a fatherly figure whom I care for and appreciate. As with my Master's work, this dissertation has started by him showing me a paper that he thought deserved a day worth of reading. I am indebted to Dr. Tawfik not only for guiding me through the most difficult times I have had with my work, but for shaping me as an academic and doing so with joy. Despite the difficult circumstances, he managed to leave me extremely happy every time we had a meeting. I cannot be thankful enough to him for committing to me and making

me a priority. More importantly, I cannot thank him enough for everything that he taught me for it is far greater than what I can write.

Dr. Alioune Ngom has a very special influence on my Doctoral work and on me as a person. His special talent of always putting himself in the other person's shoes has never ceased to amaze me. He has taught me by example how to always thrive to do what's right, be generous and smile. My Doctoral work would not have been the same without his influence as he helped me explore the previously unknown area of Bioinformatics. I value his priceless opinions, wonderful advice and constant encouragement, which were dearly needed under the circumstances. This is especially so because I would never have taken the bold step of being involved in Bioinformatics research without the knowledge that he is my guide. His "keep up the good work" has done miracles, I cannot thank him enough for it every time I hear it.

My other half, who continues to embody unconditional love, has been a wonderful part of this journey. Through it all, he has brought me nothing but smiles and has filled me with will and determination. He has been the rock I loved resting on and has single-handedly managed to restore my strength after a period of loss. Ahmed Kamil Al-Rikabi, thank you for being you.

My sister Dina has been an endless supply of positive energy and a tremendous source of wisdom. Getting angry at me for wasting time has definitely worked Dina. I love you and thank you for your objective opinions, your encouragement and for being proud of me.

Dr. Robert Kent, I will not say more than thank you for giving me valuable advice when most needed and always having a story whose moral made my days shinier and my heart lighter.

I loved having my lab mates and friends as a part of my life and continue to value their company every day. Samia Assoul, Hijaz Al-Ani, Marwa Khater, Elham Salehi, Amir Amintabar, Fadi Hanna, Natasha Kovacev, Adlane Habed, Saad Shaya, thank you for listening, caring and being a part of my life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

"I think you're begging the question," said Haydock, "and I can see looming ahead one of those terrible exercises in probability where six men have white hats and six men have black hats and you have to work it out by mathematics how likely it is that the hats will get mixed up and in what proportion. If you start thinking about things like that, you would go round the bend. Let me assure you of that!"

- Agatha Christie, *The Mirror Crack'd*

Probabilistic systems for reasoning about uncertain knowledge are well-studied in Artificial Intelligence. Graphical models have been developed to perform evidential reasoning by propagating the probabilities associated with domain variables. These models have in some respects revolutionized AI techniques in that they provide means to perform the kind of reasoning needed in expert systems much faster than when resorting to using the full joint probability distribution, thus enabling the use of probabilistic formalisms in AI applications.

As every good theory comes with shortcoming, so does the use of probabilities as measures of uncertainty in AI systems; it is constrained by the ability to obtain the probabilities required for the construction of the model (Parsons, 2001) and ensuring that the numbers acquired are precise enough to guarantee an acceptable behaviour

of the resulting system (Druzdzel and Van Der Gaag, 2000).

As a response to such difficulties, and motivated by the idea of incorporating reasoning about uncertainty in the set of capabilities to be modeled by a *common-sense reasoning* framework (McCarthy, 1959), qualitative methods that abstract probabilistic systems have been proposed. The aim is to perform uncertain reasoning in ways that cope with either partial numerical information or a complete lack of numerical information. These have been termed *qualitative probabilistic methods* in the sense that they do not provide a precise description of an uncertain belief state (as in probabilistic systems) but only provide a description that constrains the probability of the uncertain belief to belong to some set (Wellman, 1994). As a result, they do not require the specification of numerical probabilities.

At the time of writing, the literature contains a variety of such formalisms differing in the way probabilities are made into qualitative notions. Some of these formalisms provide an abstraction that captures how change in the various of probabilities affects the overall behaviour of the system by capturing how probabilities influence each other. Other formalisms captures notions such as the order of magnitude class under which a probability falls under instead of its exact numerical value. A natural consequence of this variety is that each qualitative formalism comes with its own set of features and faults (Parsons, 2001), which is the motivation behind this work.

At the heart of this dissertation lies the idea that a good way to speak about uncertainty without resorting numbers is done by making the notion of surprise the central concept behind it. More specifically, we investigate the ability of surprise to present better means for abstraction in formulating a qualitative uncertainty calculus in instances where qualitative probability fails in providing a good representation.

We leave the question of what is good and what is surprise vague for the moment but note that the notion of surprise we are interested in captures the relative unexpectedness of an event with respect to other events belonging to the same distribution and is therefore not merely that of the inverse probability. The usefulness of this measure stems from the fact that unlike probability, it provides for a relative measure and not an absolute one, which enables comparing the uncertainty attached with events belonging to different distributions and therefore achieves distribution

independence.

An important aspect of envisioning the ability of surprise in modeling uncertain knowledge is that of considering its capability of managing changing information. By envisioning, we mean the ability to hypothesize, to project into the future, or the past or present. This is an important aspect of our hypotheses as existing qualitative uncertainty calculi are not well-developed with respect to managing information that is changing through time.

Another important aspect of envisioning this capability is to see if the calculus is compatible with and is complementary to existing calculi in the sense that it can be used in an à la carte framework where one is able to choose the qualitative uncertainty calculus based on the level of abstraction required and the needs of the application. This aspect comes directly from the fact that different abstractions are tailored to meet different needs of the application domain as the literature review shows.

The last aspect of this vision is the examination of how well the theoretical benefits discovered perform in a real-world application where the parameters involved are complex and the uncertainty ubiquitous.

1.1 Motivation and Desiderata

This dissertation is motivated by the hypothesis that the concept of surprise is more appropriate than probability for being utilized as a base to abstract uncertain knowledge in a qualitative uncertainty formalism.

Hence, this dissertation is about presenting a common-sense notion of 'surprise' and incorporate it in a formalism that should:

1. Be capable of capturing the different epistemic states of events of being either believed, disbelieved or neither.
2. Present a way for propagating uncertain knowledge correctly and efficiently.
3. Not suffer from undesirable characteristics not attributed to an equivalent numerical representation of surprise.

4. Not introduce more undesirable characteristics that existing qualitative methods do not suffer from.

1.2 Objectives

It is now possible to state the objectives of this dissertation in a more detailed form. They are given as follows:

1. Present an alternative to existing qualitative probabilistic calculi that is specifically based on the idea of abstracting surprise rather than probability.
2. Demonstrate that the presented formalism is capable of representing uncertain knowledge as in qualitative probabilistic formalisms.
3. Demonstrate that the new calculus retains the inference methods found in the calculi that use a direct abstraction of probability theory.
4. Demonstrate that the presented calculus is epistemically better equipped to deal with complete or partial lack of information than one that abstracts probability.
5. Demonstrate the flexibility of the surprise-based qualitative formulation for being used in conjunction with existing calculi to offer different forms and levels of abstraction as needed by the application domain.
6. Demonstrate the capability of the formalism in being used to capture intricate forms of uncertainty in a real-world application.
7. Demonstrate the possibility of using the proposed formalism in a calculus to propagate uncertain knowledge through time.

1.3 Contributions and Expected Benefits

As noted earlier, the contribution of this work lies in being able to use the notion of whether or not events are considered surprising in a qualitative calculus to represent

uncertain knowledge. As the concept of surprise has not been previously explored with respect to qualitative uncertainty reasoning, the ideas contained between the folds of this dissertation represent a completely novel approach to qualitative probabilistic reasoning.

This alternative view is useful in several ways. Before this work, the way to go about formulating a qualitative uncertainty calculus has been to apply a chosen abstraction method to an existing quantitative uncertainty calculus. The motivation behind this is to have the resulting qualitative formalism retain the belief inference capabilities and some of the useful semantics of the original calculi. In this work, the surprise-based formalism is not directly based on abstracting probability but retains the mechanism of belief propagation through conditioning offered by probabilistic methods. What the alternative theory presented here offers is the elimination of the undesirable implications that follow from abstracting probability theory. One of these being the lack of *relativity*. For instance, let two events W_1 and W_2 be part of a set of N events and have probabilities $Pr(W_1) = 0.43$ and $Pr(W_2) = 0.47$. Although probability enables deducing that the event W_2 is more probable than the event W_1, whether or not these two events are likely or less likely is not directly obtainable from their numerical probabilities alone. In order to reach such knowledge, one must know all the probabilities of the N events. In addition, if W_1 and W_2 were drawn from two different distributions, then their probabilities cease to be representative of their relative likelihood. This undesirable feature transfers itself to any direct abstraction of numerical probability. A qualitative formalism that abstracts surprise instead will not suffer from this issue as the dissertation demonstrates. Some of the details of this calculus have been published in (Ibrahim et al., 2009c) and (Ibrahim et al., 2009b).

Another benefit lies in what this dissertation contributes to the application domain chosen as a test bed, which is the discovery of the various causal interactions among cellular components by the examination of their genetic profiles under different conditions and at different times. Apart from achieving a workable behaviour in this environment, the studies performed in this work shed a light on the importance of extracting the vast qualitative information embedded in the heaps of data collected by biologists. The study we present here (Ibrahim et al., 2009a) is unprecedented

in the sense that although there exist voices in the bioinformatics community that present concerns with respect to using purely quantitative methods for dealing with the large amount of data collected, our study goes a step further by shifting the focus on demonstrating the abundance of qualitative knowledge in the collected data that can be fully utilized with qualitative formalisms.

1.4 Dissertation Structure

The rest of the dissertation is structured as follows. The next two chapters present an overview of the literature with respect to the various forms under which probability is used as an uncertainty-handling framework in Artificial Intelligence. More specifically, Chapter 2 provides an introduction to probabilistic reasoning to handle uncertainty in Artificial Intelligence and the graphical representations used as tools to perform the various reasoning tasks associated with it. The aim of the chapter is not to merely provide an introduction to the field, but also to make visible some of the difficulties that face probabilistic methods. The discussions of Chapter 2 also present a motivation to research concerned with the various ways that probabilistic methods can be reformulated to avoid the difficulties faced by probabilistic reasoning, which is the topic of Chapter 3.

Chapter 3 is concerned with uncertainty-handling systems that are based on the principles and calculations of probability theory but abstract away from actually using numbers. These are systems that present an alternative where the precision required by numerical probabilistic systems is either unattainable or unnecessary. The chapter presents a taxonomy of the various calculi that differ by the form of abstraction applied to probability calculus. As done in Chapter 2, Chapter 3 presents a critical study of each formalism, outlining what it achieves over quantitative probabilistic systems and the pitfalls it suffers from and uses these to motivate alternative work.

After these chapters, the dissertation shifts to introduce the proposed formalisms of surprise-based qualitative uncertainty calculi in Chapter 4. The chapter proposes two formalisms based on the idea of having a qualitative uncertainty calculus that offers characteristics that are not available in the qualitative calculi that abstract

probability.

Studying the features resulting from the proposed formalisms is done by formulating a number of graphical models that use these formalisms to perform reasoning tasks in a complex, real-life environment. While Chapter 5 is concerned with the development of the graphical models that are used to test the newly-formulated qualitative calculi, Chapter 6 utilizes the models developed in a real-life application. The application chosen for the purpose is in a bioinformatics setting and consists of predicting the relations among genetic components using gene expression data. The chapter studies how the various graphical representations formulated in Chapter 5 are used to 1) study the advantages of the calculi proposed in Chapter 4 as improvements over existing qualitative probabilistic methods 2) to provide results that are comparable or better than ones available in the literature with respect to the problem at hand. Chapter 7 summarizes the findings of this work and outlines the various possible paths that future research can follow.

Chapter 2

Probabilistic Systems: A Quantitative Perspective

When it is not in our power to determine what it true, we ought to follow what is most probable

- René Descartes, *Discourse on the Method*

Probability theory has been used to represent uncertain knowledge for several hundreds of years, taking various forms. It provides a systematic way to determine the truth associated with the occurrence of events when prior knowledge and current observations are not sufficient to do so categorically (Pearl, 1988).

Probability theory quantifies variables (representing events) by associating a numerical value designating the degree of belief granted to a variable by some body of knowledge (Pearl, 1990). Essentially, for any variable W drawn from Ω, the set of all variables defined by the problem domain (Pearl, 1988), a probability distribution Pr assigns to each value w of the variable W a probability measure $Pr(W = w) \in [0, 1]$ that estimates the degree to which w is the current true value of W. Accordingly, a probability distribution Pr is a mapping from the set of all possible values of the

variables to the unit interval $[0, 1]$ and assigns to each value a probability $Pr(w)$ (for short). This probability value obeys a set of axioms that constrain its value and enable the formulation of well-defined rules to combine and manipulate probabilities of variables and deduce other useful probabilities.

Essentially, the probability distribution Pr is capable of answering many questions about the domain using rules to compute the probabilities of the conjunctions and disjunctions of variables as well as rules to compute the conditional probabilities of variables given other variables (Bayes' rule). In practice however, a joint probability distribution can become intractably large causing the computations needed to extract probability values increasingly intractable (Pearl, 1988; Parsons, 2001). As this presented a limitation to the use of probability theory in Artificial Intelligence (D'Ambrosio, 1999), research focused on creating formalisms that attempt to identify various independence relations among the domain variables to make the computations required more efficient.

This chapter introduces one such formalism, called Bayesian Networks (BNs). A BN is a graphical representation whose structure and semantics explicitly model the independence relations among the various variables in the domain in a way that reduces the amount of computations required for the various queries. As a result, BNs capture the uncertainty present in the domain in a more compact and efficient way (relative to dealing with the complete joint probability distributions) (Pearl, 1988; Russell and Norvig, 2003). The graphical representation is termed a Bayesian network as it is based on the computations performed using Bayes' rule and is introduced in section 2.1 below.

In the discussions that follow, variables are denoted by upper-case letters (W) and their values by lower-case letters (w). Bold-face upper-case letters (e.g. $\boldsymbol{W}$) denote a set of variables while a bold-face lower-case letter (e.g. $\boldsymbol{w}$) denotes a set of instantiations. Subscripts refer to the particular context that the variable refers to.

2.1 Bayesian Networks

A Bayesian network (BN) is a compact graphical representation of a joint probability distribution Pr (Pearl, 1988; Jensen, 2001) defined over a set of variables Ω. The network consists of a directed acyclic graph (DAG) (also termed *diagraph*, for short) where the nodes correspond to random variables and the arcs represent direct dependencies between the linked variables (Pearl, 1988). The structure of the graph describes the influences exerted among the variables and constitutes the qualitative aspect of the BN. To fully specify a BN, the conditional probabilities of each variable given its immediate predecessors are given, and form the quantitative part of the BN (Pearl, 1986). Together, the qualitative and quantitative parts of a BN uniquely define a joint probability distribution on the set of variables Ω under study.

Definition 1. Bayesian Networks:

Given a probability distribution Pr, a Bayesian network (BN) representation of Pr is a directed acyclic graph $G = (V(G), E(G))$, where $V(G)$ is the set of nodes capturing the variables of the domain and $E(G)$ is the set of arcs capturing the conditional independence among the variables in the following way:

1. *$\forall\ W_i, W_j \in V(G)$, if $(W_i, W_j) \in E(G)$ then there exists a direct probabilistic dependence relationship between W_i and W_j in which W_j is directly dependent on W_i.*

2. *$\forall\ W_i, W_j \in V(G)$, if $(W_i, W_j) \in E(G)$ then W_i is said to be an immediate predecessor, or parent of W_j, $W_i \in \pi(W_j)$, where $\pi(W_j)$ is the set of all parents of W_j.*

3. *$\forall\ W \in V(G)$, W is described in terms of a conditional probability distribution $Pr(W|\pi(W))$ defined on W, where $Pr(W|\pi(W))$ reduces to an unconditional distribution if $\pi(W) = \varnothing$.*

Example 1. *Figure 2.1 shows a network that describes the factors affecting the condition in which a civilian is found by rescue agents inspired by an example found in* (Korb and Nicholson, 2003). *The age of the civilian (modeled by the variable* A)

affects both the civilian's mental and physical health (*modeled by* MH *and* PH *respectively*). *An increase in the humidity* (*modeled by* HM) *and temperature* (*modeled by* T) *may cause the development of Hyperthermia* (*modeled by* H). *Moreover, the civilian's mental health, physical health, her diagnosis with hyperthermia as well as the efficiency of the rescue process* (*modeled by* E) *all influence the condition in which the civilian is found* (*modeled by* C).

In this network, $V(G) = \{$HM, A, T, MH, PH, H, E, C$\}$. $E(G)$ *reflects the dependence relationships among the nodes of the network. Moreover,* π(C) $= \{$E, H, MH, A, PH$\}$ *while* π(A) $= \phi$ *for instance.*

The conditional probabilities for the nodes are also shown in the figure. For the sake of simplicity, we assumed that all variables are binary in the example given in Figure 2.1. The values for each variable is given by High/Good (True) *and Low/Bad* (False). *Also, we only enlist the probabilities for the true case in the figure as the probability for the false case can be directly inferred from it. Moreover, we only enlist 16 out of the possible* $2^5 = 32$ *cases for the table showing* $Pr(C|E, H, MH, A, PH)$ *as it would be too large otherwise.*

2.2 Independence Assumptions in Bayesian Networks

Bayesian networks establish a clear correspondence between the topology of the network and the various types of independence relations that exist among the variables in the network. More specifically, the concept of conditional independence can be formalized via the topology of the BN. For any two nodes W_i and W_j, if they are separated by a subset of nodes $\boldsymbol{W}_k \subseteq \{V(G)/\{W_i \cup W_j\}\}$ in the network, then this implies the independence of W_i and W_j given $\boldsymbol{W}_k$. In other words, W_i and W_j are said to be conditionally independent given the set of nodes $\mathbf{W}_k$, given by: $Pr(W_i|W_j, \boldsymbol{W}_k) = Pr(W_i|\,\boldsymbol{W}_k)$. The idea of conditional independence is formalized by the *d-separation* criterion. In essence, two sets of nodes $\boldsymbol{W}_i$ and $\boldsymbol{W}_j$ are conditionally independent given set of nodes $\boldsymbol{W}_k$ if every path from any node in $\boldsymbol{W}_i$ to any

node in $\boldsymbol{W}_j$ is blocked by $\boldsymbol{W}_k$, where the notion of blocking is given in definition 2.

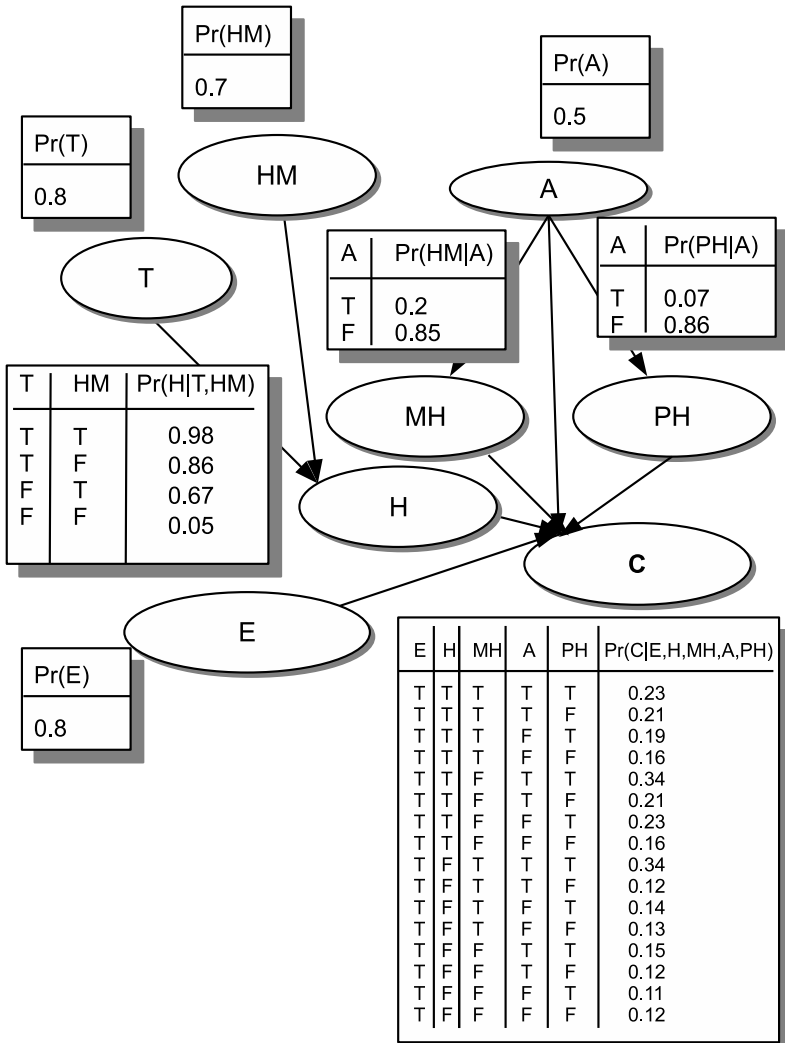


Figure 2.1: An example Bayesian Network DAG structure Inspired by one given in (Korb and Nicholson, 2003). In the network: HM: Humidity; T: Temperature; H: Hyperthermia; E: Rescue Efficiency; A: Age; MH: Mental Health; PH: Physical Health; C: Condition When Found.

Definition 2. Blocking (Pearl et al., 1989) :

Given a directed acyclic graph $G = (V(G), E(G))$, let p be a path of arcs in $E(G)$ connecting two nodes W_i and W_j, where $W_i, W_j \in V(G)$. Then p is said to be blocked

by the set of nodes $\mathbf{W}_k \subseteq V(G)$ *iff either* W_i *or* W_j *is included in* $\mathbf{W}_k$ *or* p *contains three consecutive nodes* P_1, P_2, P_3 *for which one of the following conditions holds:*

1. *The arcs* $P_1 \leftarrow P_2$ *and* $P_2 \rightarrow P_3$ *are on the path* p*, and* $P_2 \in \mathbf{W}_k$*. In this case,* P_2 *is said to be* tail-to-tail *with respect to the path from* P_1 *and* P_2*.*

2. *The arcs* $P_1 \rightarrow P_2$ *and* $P_2 \rightarrow P_3$ *are on the path* p*, and* $P_2 \in \mathbf{W}_k$*. In this case,* P_2 *is said to be* head-to-tail *with respect to the path from* P_1 *and* P_2*.*

3. *The arcs* $P_1 \rightarrow P_2$ *and* $P_2 \leftarrow P_3$ *are on the path* p*, and* $\sigma^*(P_2) \cap \mathbf{W}_k = \phi$*. In this case,* P_2 *is said to be* head-to-head *with respect to the path from* P_1 *and* P_2*. Where* $\sigma^*(P_2)$ *denotes the set of nodes composed of* P_2 *and all its descendants.*

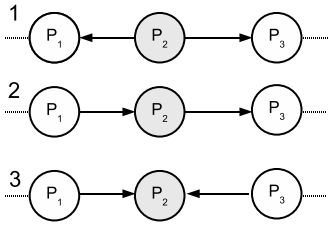


Figure 2.2: Blocking Conditions

The three conditions are illustrated in Figure 2.2. If one of the conditions is satisfied, then $\mathbf{W}_k$ is said to *block* the path p between nodes W_i and W_j, otherwise the path is said to be *active*.

Example 2. *In the network shown in Figure 2.1, examples for the three blocking conditions are:*

1. *Observing a value for* A *renders the nodes* MH *and* PH *independent. They are dependent however, if* A *has not been observed. This is because observing* A *blocks the tail-to-tail path between* MH *and* PH*.*

2. *Observing a value for* H *renders the nodes* T *and* C *independent, while they remain dependent if no observation has been made. This is because observing H blocks the head-to-tail path between* T *and* C*.*

3. *Given no observation, the nodes* T *and* HM *are independent since they both have no parent nodes. However, when observing* H *or any of its descendants, this independence disappears.*

Applying the concept of blocking to all the possible paths from W_i to W_j leads to the concept of *d-separation* (Verma and Pearl, 1992), given in definition 3 below.

Definition 3. d-separation (Coupé and Gaag, 2002):

Given a directed acyclic graph $G = (V(G), E(G))$*, let* $\mathbf{W}_i$*,* $\mathbf{W}_j$*,* $\mathbf{W}_k \subseteq V(G)$*. The set of nodes* $\mathbf{W}_k$ *is said to d-separate the sets of nodes* $\mathbf{W}_i$ *and* $\mathbf{W}_j$ *in* G*, denoted by* $\langle \mathbf{W}_i | \ \mathbf{W}_k | \mathbf{W}_j \rangle_G^d$*, if for each node* $W_i \in \mathbf{W}_i$ *and each node* $W_j \in \mathbf{W}_j$*, every path* p *from* W_i *to* W_j *in* G *is blocked by* $\mathbf{W}_k$*.*

D-separation in a BN can be discovered using a linear time, depth-first-search-like algorithm (Verma and Pearl, 1992; Pearl et al., 1989; Shachter, 1988).

Example 3. *In Figure 2.1, observing* H *blocks all the possible paths from* Temperature *to* C *and therefore renders the two d-separated. This is not the case when observing* MH*. Although it does block the path* {A-MH-C}*, it does not render* A *and* C *d-separated as the path* {A-PH-C} *is still active (even* PH *is observed), there remains the active path* {A-C}*.*

If every *d-separation* corresponds to a true independence in the probability distribution defined over the system, then the Bayesian Network is said to be an *I-map* (short for an independence map) (Ghahramani, 1998; Pearl, 1988), which is formally given in definition 4 below.

Definition 4. I-map (Coupé and Gaag, 2002):

Given a directed acyclic graph $G = (V(G), E(G))$ *with a probability distribution* Pr *defined on* $V(G)$*,* G *is called an I-map for* Pr *if for all sets of nodes* $\mathbf{W}_i$*,* $\mathbf{W}_j$*,* $\mathbf{W}_k \subseteq V(G)$*, we have:*

If $\langle \mathbf{W}_i | \ \mathbf{W}_k | \mathbf{W}_j \rangle_G^d$*, then* $\mathbf{W}_i$ *and* $\mathbf{W}_j$ *are conditionally independent given* $\mathbf{W}_k$ *in* Pr*.*

Hence, the independence relations of Pr are captured by the topology the graph G in the following ways (Pearl, 1988):

1. A node is conditionally independent of its non-descendants, given its parents.
2. A node is conditionally independent of all other nodes in the network, given its parents, children, and children's parents (a nodes parents, children and children's parents are termed the nodes Markov blanket).

Example 4. *In Figure 2.1, the conditional probabilities of the variable Mental Health* (MH) *are only calculated with respect to its parent Age* (A)*. The same applies for all the other variables. This reflects the idea of independence given the values of the parent nodes* [1]*.*

Given the conditional independence properties given above and the conditional probabilities, the conditional probability of a node given its parents and the prior probabilities of the root nodes are the only probabilities required to fully specify the joint distribution represented by the network. Using this, the chain rule can be used to recursively factorize the joint probability over all variables.

$$\begin{aligned} Pr(V(G)) &= Pr(W_1|W_2,...,W_n).Pr(W_2|W_3,...,W_n).....Pr(W_n) \\ &= \prod_{W_i \in V(G)} Pr(W_i|\pi(W_i)) \end{aligned}$$

Where the above computation defines the joint probability distribution Pr for a BN G whose set of nodes $V(G) = \{W_1, ..., W_n\}$ such that the variables are ordered so that no variable W_i follows its immediate predecessors $\pi(W_i)$ (Pearl, 1988), which is how the acyclic nature of the graph is ensured (Ghahramani, 1998).

[1] It is worth noting that the number of probabilities to be calculated for a node is exponential with respect to the number of parents it has. For example, when constructing the conditional probability table for Condition When Found, it becomes necessary to assign a total of $2^5 = 32$ for the complete description of the joint probabilities of the variable (Condition When Found (C) has five parents, all of which are binary variables).

Moreover, it should be visible from the above equation that there are many ways to factorize the variables in a joint distribution, which implies that more than one BN can be constructed in consistency with a joint. Here, the concept of the *I-map* (definition 4) ensures a correct structure for the generated Bayesian Network.

2.3 Inference in Bayesian Networks

The most common inference task in Bayesian networks is that of computing the posterior marginal probability, which is performed through *evidence propagation* (Castillo et al., 1996). This is the process of computing the posterior probabilities of a set of variables in the network when the specific values of some other variables in the network are observed. More formally, given a set of variables $\mathbf{Y}$ having been observed to possess values $\mathbf{y_0}$, the task is to compute the posterior probability of another set of (unobserved) variables $\mathbf{X}$, $Pr(\mathbf{X}|\mathbf{Y} = \mathbf{y_0})$. The literature contains many algorithms for evidence propagation and can be classified into algorithms that perform *exact inference*, providing the exact posterior probability of the query variable, and those that perform *approximate inference* by only producing a bound to which the correct solution is guaranteed to belong.

Exact inference exploits the independence structure of the network to efficiently propagate the evidence. There exist several algorithms for this task. One such algorithm is the *clique tree propagation* algorithm (CTP) (Lauritzen and Spiegelhalter, 1988; Shafer and Shenoy, 1990; Jensen et al., 1990) which transforms a BN into a secondary structure called the *clique tree* or *junction tree*. This secondary structure allows the computation of the answers to all possible queries having one query variable and a fixed set of observations (Jensen et al., 1990). Another algorithm is the *variable elimination* algorithm (Zhang and Pool, 1996), which makes use of the fact that only a subset of the variables present in the network is required for most queries and builds a query-oriented method that can carry out inference in large networks that the CPT algorithm cannot deal with. *Variable elimination* is related to a set of algorithms that focus the inference on a small subset of the variables when the query at hand does not requires the entire network (see for example (Li and D'ambriosio, 1994; Shachter,

1990; Jensen, 1995; Dechter, 1996; D'ambrosio, 1994)). Despite the various efforts to use the topology and independence assumptions present in the network however, all the algorithms that perform exact inference are known to be $\mathcal{NP}$-hard in the worst case (Cooper, 1990; Chickering et al., 2004; Jensen et al., 1990).

Approximate algorithms on the other hand do not provide an exact solution, but produce a bound to which the correct solution is guaranteed to belong to. Although approximate inference in BNs has also been proven to be $\mathcal{NP}$-hard in the worst case (Dagum and Luby, 1993), they are likely perform in polynomial time (Dagum and Luby, 1993; Haipeng Guo and William Hsu, 2002) and therefore constitute an interesting alternative (when possible) to exact inference algorithms in large and highly-connected networks because it becomes impossible to obtain results using exact inference algorithms with such networks (Castillo et al., 1996; Cooper, 1990).

Complexity results for inference in Bayesian Networks can also be categorized with respect to the topology of the Network. Essentially, BNs can be categorized as either *singly-connected* or *multiply-connected*. A BN is said to be singly-connected if for any node in the DAG, there exists at most one path to any other node in the network (Wu and Butz, 2005). Despite the previous general consensus that inference in singly-connected BNs is more efficient that in multiply-connected BNs (Kim and Pearl, 1983), (Shimony and Domshlak, 2003; Wu and Butz, 2005) demonstrate that evidence propagation is NP-hard even in multi-valued singly-connected BNs. Moreover, Wu and Butz (2005) show that the hardness of exact inference in BNs cannot be determined exclusively based on the topological structure of the DAG of a BN.

2.4 Constructing Bayesian Networks

A central issue to the development of systems that use Bayesian Networks is the construction of the network. Essentially, building a Bayesian Network for an application domain involves three tasks. The first is to identify the variables required to define the domain at hand, followed by identifying the relationships holding among the variables in order to define the structure of the network and finally to find the probabilities required to fully specify the quantitative part of the network (Druzdzel

and Van Der Gaag, 2000).

The required probabilities are mainly obtained from statistical data, the literature pertaining to the domain at hand and human experts (Mani et al., 2005). Using the data, BNs are constructed using one of two methods. The first is through knowledge engineering sessions that involve domain experts who identify the important variables along with the topology of the DAG and assess the prior and conditional probabilities for the various nodes of the network. The second method is that of automatically learning the structure and probabilities of the network from data (Buntine, 1994). Algorithms for learning the DAG structure from data can be generally classified into those that are based on constraint-based search (Pearl and Verma, 1991; Spirtes et al., 1993) or through a Bayesian search for a graph that produces the highest probability given the data (Cooper and Herskovits, 1992). Given the DAG structure, computing the probabilities becomes a straightforward task (Druzdzel and Van Der Gaag, 2000).

Constructing the qualitative part of the network using either method, despite the considerable efforts required, is considered achievable (Coupé and Gaag, 2002; Mani et al., 2005). When the structure of the network is complete and the data is fully available, then building the corresponding probability tables becomes an easy task (Druzdzel and Van Der Gaag, 2000). On the other hand, for most real-life application, the data is usually incomplete and constitutes only partial knowledge of the domain. When this is the case, domain experts assess the required numbers with the aid of the literature, available data and experience (see for example (Coopé et al., 1999) in which an example of a common method for constructing BNs in medical diagnosis is shown). This leads to an inevitable inaccuracy in the quantitative part of the generated network and deems the assessments made from the network inaccurate and possibly unreliable (Druzdzel and der Gaag, 1995). Algorithms for learning the structure and parameters of BNs given incomplete data are also available, but are also consequently imprecise as the result of the induced inaccuracy (Friedman, 1997; Wong and Guo, 2006; Heckerman, 1995; Getoor et al., 2002; Li et al., 2005).

As a result, there has been various studies to analyze Bayesian Networks for sensitivity to the inaccuracies induced by the construction process (Castillo et al., 1997; Laskey, 1995; Jensen et al., 1995; Kwisthout and Gaag, 2008; Coupé and Gaag, 2002;

Coopé et al., 1999). These studies have been generally termed *sensitivity-analysis* and are concerned with "understanding the relationship between local network parameters and global conclusions drawn based on the network" (Chan and Darwiche, 2004). As different parameters of the network require different levels of accuracy to provide an acceptable behavior from the resulting network (Coopé et al., 2000; Chan and Darwiche, 2002), sensitivity analysis is used to evaluate the level of accuracy that a parameter requires and subsequently fine-tune and then evaluate the resulting network. The problem has been given several formal definitions (see (Jensen et al., 1995; Coupé and Gaag, 2002; Chan and Darwiche, 2002; Jensen, 2001)). Here, we enlist the one given in (Chan and Darwiche, 2004) in Definition 5 below.

Definition 5. Sensitivity Analysis (Chan and Darwiche, 2004):

Given a Bayesian Network G and a subset of network parameters, sensitivity analysis on G is the process of identification of the possible changes to these parameters that can ensure the satisfaction of one or more of the following query constraints:

1. *Given an event w and an evidence e, $Pr(w|e) \geqslant p$, for some $p \in [0, 1]$*
2. *Given two events w_1 and w_2 and an evidence e, $Pr(w_1|e)/Pr(w_2|e) \geqslant k$, for some $k > 0$.*
3. *Given two events w_1 and w_2 and some evidence e, $Pr(w_1|e) - Pr(w_2|e) \geqslant k$, for some $k \in \mathbb{N}$.*

Example 5. *In the network given in figure 2.1, the network contends that being diagnosed with Hyperthermia given that it is both hot and humid is twice as likely as being diagnosed with Hyperthermia given that it is only humid, i.e. $Pr(HM = True, T = True|H = True)/Pr(HM = True, T = False|H = True) = 2$, while a domain expert may believe that the ratio should be at least 4. In this case, the question becomes: which network parameters should be changed to enable obtaining the correct ratio? and by how much?*

Central to the problem of addressing the constraints are the following questions:

1. What are the number of parameters to be systematically varied at a time for every test? (Coupé and Gaag, 2002; Chan and Darwiche, 2004).

2. What are the conditional probabilities that are uninfluential to the parameter (or set of parameters) under study? (Coupé and Gaag, 2002; Chan and Darwiche, 2002)

There exist algorithms that perform different forms of sensitivity analysis. Chan and Darwiche (2002) introduce a tool , called SAMIAM (Sensitivity Analysis, Modeling, Inference And More) that has many features, one of which is to perform sensitivity analysis on a given network. It has been used to obtain some bounds on the impact that the relevant parameters may have on the results of the queries (also studied in (Renooij and Gaag, 2004)). Coupé and Gaag (2002) conduct a study that is based on using the concept of conditional independence to reduce the number of parameters to be tested when performing sensitivity analysis by discovering the parameters that are irrelevant to a certain evidence. Despite the continuous efforts for more efficient ways to perform the analysis, such studies remain a burn with respect to the computational effort required (Kwisthout and Gaag, 2008; Coupé and Gaag, 2002; Chan and Darwiche, 2004). In fact, the task of sensitivity analysis has been shown to be $\mathcal{NP}$-hard (Kwisthout and Gaag, 2008).

Chapter 3

Qualitative Abstractions of Probabilistic Systems

Common sense is a wild thing, savage, and beyond rules

- G.K. Chesterton, *Charles Dickens: A Critical Study*

The previous chapter introduced the basic concepts of probabilistic reasoning and how it is used to deal with uncertainty by attaching probabilities to variables and using these probabilities as measures of belief or likelihood. Bayesian Networks were introduced as the main graphical representation used to perform probabilistic inference.

There are different opinions in regards to the use of numbers to represent uncertainty. On the protestor's side, the sources of the numbers assigned as probabilities are questioned (Druzdzel and Van Der Gaag, 2000). For instance, there exist concerns with respect to the ability of a domain expert to assess the probabilities because even domain experts may not have sufficient information to establish valid subjective probabilities (Parsons, 2001) or maybe reluctant to do so. This is because of the difference between the categorical reasoning offered by their developed intuitions and the precision required for the specification of the probabilities (Chard, 1991). For in-

stance, medical experts maybe able to tell which disease is more likely than the other given the symptoms and by how many times, but are not willing to say for instance that $Pr(disease1|symptoms) = .23$ and $Pr(disease2|symptoms) = 0.82$. Moreover, data for real-life applications in many domains tend to be incomplete, which implies that only partial knowledge is available. Hence, the assessments obtained through the different means are inevitably inaccurate (Druzdzel and der Gaag, 1995). Even the most enthusiastic probabilists such as Cheeseman (1988), who despite his avid defense of probabilities as being sufficient to reason about uncertainty when used correctly (Cheeseman, 1985), questions the sources of the numbers used as probabilities and their accuracy (Cheeseman, 1988). Moreover, as we have seen in Chapter 2, parameter tuning and analyzing BNs for sensitivity to noise in order to prevent inconsistent results is an $\mathcal{NP}$-hard task (Kwisthout and Gaag, 2008).

On the other hand, there exist domain-specific studies in which empirical evidence shows the robustness of Bayesian techniques in some applications despite the large amounts of random noise. For example, Henrion et al. (1996) demonstrate that Bayesian medical diagnosis is relatively insensitive to noise. Also, Pearl (1988) reasons that the numbers provided as probabilities are the ones most likely to be acquired through experience and are even possibly represented in cognitive structures. Nikovski (2000) has also proposed a method to obtain the necessary probabilities in the medical domain by having experts estimate both the sensitivity and specificity for pairs of findings and diseases, which has the advantage of physicians being more able to relate to these measures and more capable of providing good estimates (Nikovski, 2000). In addition, when the statistical data is available, systems that learn the probability values required for the construction of the conditional probabilities of the BNs are considered reliable as they have been heavily studied (see for example (Nachman et al., 2004; Barash and Friedman, 2002; Getoor et al., 2002, 1999)).

Apart from all of the above, it is important to note that these systems are only useful if either enough statistics exist for the values to be inferred with sufficient accuracy or that there are experts available in the specific domain. This luxury is not enjoyed by many applications for which a complete specification of the probability values of the events constituting the model is not achievable. An example from the

medical domain is that of skin diseases for children where no expert is currently available, which renders any numerical information unattainable (Beumer, 2006).

Although overcoming the lack of real data by means of approximation (Ramezani and Marcus, 2002) or synthesis (Hand, 2006) is possible for some applications, it is not an option for many types of applications. For example, epidemiological data describing factors influencing the occurrence of illnesses cannot be approximated or synthesized when not sufficient. Another example is the problem of predicting the topological structure of proteins, where the topology of very few proteins are currently known, and available data is in general incomplete and uncertain, and approaches using numerical probability have only been successfully used in the prediction of a special class of proteins called transmembrane proteins (Kahsay et al., 2005).

In other applications, the data is available in heaps but is not verifiable for correctness and tends to be filled with noise (Parsons and Mamdani, 1993). Such applications may benefit from the added robustness that qualitative formulations can offer. A (possibly) surprising domain that fits this criterium is bioinformatics, where the nature of and cost associated with current microarray technologies prevent the possibility of repeating the experiments that are used to obtain the data, which make statistical studies of a single measurement unattainable (Filkov et al., 2002). Therefore, the numbers usually represent outcomes of a single, non-repeated experiment. There exist several concerns with respect to this issue (D'hæseleer, 2000) whose treatment remains an unanswered question (Friedman, 2004). Moreover, some applications may not require the precise specification of probabilities. For example, for crime scene investigations and forensic applications, the objective is usually to provide a "justifiable indication of the difference in magnitude of support for one hypothesis over another, given the available evidence' (Keppens, 2007) and not the exact probability of the hypotheses.

Apart from the epistemic concerns, Bayesian Networks are known to have $\mathcal{NP}$-hard inference algorithms (Cooper, 1990; Chickering et al., 2004) as Chapter 2 has shown. More efficient alternatives for reasoning exist but these usually pose fundamental tradeoff between efficiency and accuracy (Diez, 1996).

Hence, it is easy to see that the above problems necessitate alternative approaches

for some application domains. This is especially so given that the most obvious approach of approximating probabilities is by allowing them to fall within intervals rather than be constrained to mere points merely describing second-order distributions that do not achieve anything that point probabilities cannot (as argued by several authors including (Kyberg, 1989))[1]. This chapter presents a number of systems that aim at somehow abstracting probabilistic systems by using no or little quantitative information. The systems presented in this chapter are divided into three classes (a taxonomy inspired by the one given in (Parsons, 2001)) and are: systems of strict abstractions (Section 3.1), systems of order of magnitude (Section 3.2) and symbolic abstractions. These systems have been proposed as stand alone systems as well as systems to be used in conjunction with quantitative probabilistic ones (Ng and Ong, 2000). A review of the literature with respect to systems of strict and order-of-magnitude abstractions is given. Symbolic systems extend to a number of formalisms from systems of argumentation (Benferhat et al., 1993; Darwiche, 1993; Fox et al., 1992) to systems for nonmonotonic reasoning (Neufeld, 1990). However, a deeper review of these systems is beyond the scope of this work [2].

3.1 Systems with Strict Abstractions

Systems of strict abstraction are based on the idea of replacing real number by three quantities: positive, negative and zero with an aim to do away with the quantitative details and perform more qualitative, large scale reasoning (Hayes, 1985a). This area of research was initiated by Hayes' naïve physics (Hayes, 1978, 1985a,b) that aimed at formulating a new school of thought, focusing on studying the large-scale aspects of a system and "put(ting) away childish things" (Hayes, 1985a) (in reference to abundant numerical details). The general aim was to formulate a qualitative theory governing

[1]This has not altered the popularity of interval probabilities. The interested reader may refer to (Choquet, 1953; Weichselberger and Pöhlmann, 1990; Dubois et al., 1992; Breeze and Fertig, 1991) for more details. As interval-based probabilities extend much further than the scope of our work, they will not be covered in this document.

[2]The interested reader may refer to (Parsons, 2001) for an elaborate discussion on symbolic approaches.

the behavior of liquids (Hayes, 1985a) based on the idea of describing the evolution of their various properties as increasing (modeled by having a positive value), decreasing (having a negative value) and non-changing (having a value of zero). Applying the same reasoning with probabilities is the motivation behind this class of abstractions of probability theory. The idea is to shift the attention from the numerical descriptions of the point probabilities and instead study how the probability value changes with respect to evidence by increasing, decreasing or remaining constant. Graphical models to serve this aim have been formulated similarly to Bayesian Networks but having the idea of strictly-qualitative change as a core. The two main graphical models are *Qualitative Probabilistic Networks* (QPNs) (Neufeld, 1989; Wellman, 1990a) and *Qualitative Certainty Networks* (QCNs) (Parsons, 2003) and are discussed in the rest of this section.

3.1.1 Qualitative Probabilistic Networks (QPNs)

Qualitative probabilistic networks (QPNs) (Neufeld, 1989; Wellman, 1990a; Renooij and Gaag, 2008) are graphical abstractions of Bayesian Networks (BNs) that comprise of a directed acyclic graph capturing, as in BNs, the independence relations that hold among the variables the network represents. Instead of the conditional probabilities that BNs use however, QPNs replace the numerical probability tables defined on the arcs of a BN by qualitative relations that describe how evidence given for one or more nodes influences other nodes in the network (Wellman, 1990a) without resorting to numerical probabilities. In other words, the abstraction that yields QPNs preserves the DAG topology modeling the causal relations that exist among the variable while replacing the numerical probabilities that regular BNs encode by qualitative relations.

The relations that the arcs of QPNs capture are qualitative in that they define a partial order over the conditional values that would make up the conditional probability table which is replaced by the relations. Essentially, the only information they capture is the direction of the influence exerted from a node (or a group of nodes) on another node (i.e. whether the evidence makes a node more or less likely) and is hence represented by its sign, being positive, negative, zero (constant) or unknown instead

of the numerical representation given in BNs (Wellman, 1990a; Parsons, 2001).

A QPN is given by a pair $G = (V(G),E(G))$, where $V(G)$ is the set of nodes capturing the variables of the domain being represented and $E(G)$ is the set of arcs capturing the conditional independence relations holding among the variables as in Bayesian Networks. For $V(G)$, a total order on the values of each node $W \in V(G)$ is assumed (Renooij and Gaag, 2008). Also, conditional independence is captured in the same way as in BNs via d-separation (Renooij and Gaag, 2008).

A QPN uses *hyperarcs* for the diagraph G that replace the conditional probability distribution and capture qualitative probabilistic relations among the variables (Renooij, 2001) by finding monotonic characteristics in the conditional probability distribution based on the idea of first-order stochastic dominance (Bawa, 1975) (given in definition 6 below) to establish order properties over the probabilities of events.

Definition 6. First-order Stochastic Dominance (Renooij, 2001):

Let Pr be a probability distribution defined over a set of variables $\{W_1, ...W_n\}$. For any variables W_i, W_j, $1 \leqslant i, j \leqslant n$ with values $< w_{i_1}, ..., w_{i_m} >, < w_{j_1}, ..., w_{j_k} >$ where m is the total number of values W_i can have and k is the total number of values W_j can have, then W_i exhibits a first-order stochastic dominance over W_j if for every two values $w_{i_{k1}}, w_{i_{k2}}$ of W_i with $w_{i_{k1}} > w_{i_{k2}}$, the cumulative conditional probability distribution of node W_j, $Pr(w_{j_1}|w_{i_{k1}}) \vee ... \vee Pr(W_{j_m}|W_{i_{k1}})$, lies, graphically speaking, below the cumulative conditional probability distribution of W_j given $w_{i_{k2}}$, i.e. $Pr(w_{j_1}|w_{i_{k2}}) \vee ... \vee Pr(w_{j_n}|w_{i_{k2}})$ or according to the inequality:

$$Pr(w_{j_1}|w_{i_{k1}}) \vee ... \vee Pr(w_{j_n}|w_{i_{k1}}) \leqslant Pr(w_{j_1}|w_{i_{k2}}) \vee ... \vee Pr(w_{j_n}|w_{i_{k2}})$$

The above concept is used to devise a number of qualitative relations. The most basic type of relations are qualitative influences (Wellman, 1990a); they exhibit properties that are used to define other types of qualitative relations in QPNs. For this, we list them first, along with their properties, then shift out attention to other types of qualitative relations.

Qualitative Influences

Influences describe how the change of the value for one variable affects that of another, and is the most important type of qualitative relations in QPNs as the definition of the other types of relations are dependent on that of qualitative influences. Qualitative influences can be of four types, positive, negative, constant or unknown (Wellman, 1990a).

A positive influence exists between two variable W_i and W_j (W_i is said to positively influence W_j, written as $I^+(W_i, W_j)$) if observing higher values for W_i makes higher values of W_j more probable, regardless of the value of any other direct ancestor of W_j which may directly influence its value (denoted by W and expressed by $W = \pi_G(W_j)/\{W_i\}$) as given in definition 7.

Definition 7. Positive Influence (Renooij, 2001):

Let $G = (V(G), E(G))$ be a directed acyclic graph and let Pr be a joint probability distribution on $V(G)$ such that G is the I-map for Pr. Let $W_i, W_j \in V(G)$ be nodes in G with $W_i \rightarrow W_j \in E(G)$. Let $W = \pi_G(W_j)/\{W_i\}$. Then node W_i positively influences node W_j, written as $I^+(W_i, W_j)$ iff for all values w_j of W_j and all values w_{i_1}, w_{i_2} of W_i, with $w_{i_1} > w_{i_2}$ the following inequality holds

$$I^+(W_i, W_j) \text{ iff } Pr(W_j > w_j | w_{i_1}, W) \geqslant Pr(W_j > w_j | w_{i_2}, W)$$

It is worth mentioning that the above inequality can be redefined for binary variables W_i and W_j by placing a partial order on their values such that for a variable W_i with two values w_i and $\neg w_i$, $w_i > \neg w_i$. Negative and constant QPN influences are similarly given in Definitions 8 and 9 below.

Definition 8. Negative Influence (Renooij, 2001):

Let $G = (V(G), E(G))$ be a directed acyclic graph and let Pr be a joint probability distribution on $V(G)$ such that G is the I-map for Pr. Let $W_i, W_j \in V(G)$ be nodes in G with $W_i \rightarrow W_j \in E(G)$. Let $W = \pi_G(W_j)/\{W_i\}$. Then node W_i negatively influences node W_j, written as $I^-(W_i, W_j)$ iff for all values w_j of W_j and all values w_{i_1}, w_{i_2} of W_i, with $w_{i_1} > w_{i_2}$ the following inequality holds

$$I^+(W_i, W_j) \text{ iff } Pr(W_j > w_j | w_{i_1}, W) \leqslant Pr(W_j > w_j | w_{i_2}, W)$$

Definition 9. Constant Influence (Renooij, 2001):

Let $G = (V(G), E(G))$ be a directed acyclic graph and let Pr be a joint probability distribution on $V(G)$ such that G is the I-map for Pr. Let $W_i, W_j \in V(G)$ be nodes in G with $W_i \rightarrow W_j \in E(G)$. Let $W = \pi_G(W_j)/\{W_i\}$. Then node W_i exerts a constant influences node W_j, written as $I^0(W_i, W_j)$ iff for all values w_j of W_j and all values w_{i_1}, w_{i_2} of W_i, with $w_{i_1} > w_{i_2}$ the following inequality holds

$$I^+(W_i, W_j) \text{ iff } Pr(W_j > w_j | w_{i_1}, W) = Pr(W_j > w_j | w_{i_2}, W)$$

Example 6. *Figure 3.1 shows the QPN equivalent of the Bayesian Network given in Figure 2.1 of Chapter 2. As mentioned above, the nodes are ordered by their values such that a* True *value is higher than a* False *value. G is defined by its set of nodes $V(G)$ =* {Age, Mental Health, Physical Health, Temperature, Humidity, Hyperthermia, Rescue Efficiency, Condition When Found} *and its set of edges $E(G)$ defined by the directed arcs shown in the graph. The only information encoded in the arcs are the signs of the influences from one node to another. For instance, the Figure shows that node* Temperature *positively influences node* Hyperthermia *as higher temperature increase the probability of causing hyperthermia, while the node* Age *has a negative influence on* Physical Health *as one becomes increasingly weaker as one's age increases.*

Properties of Qualitative Influences

QPN influences exhibit a number of properties that make possible their propagation along paths in QPNs.

1. **Symmetry:** In a QPN G where a node W_i exerts a qualitative influence on a node W_j, W_j exerts an influence of the same sign on W_i. In other words:

   $I^{\varrho}(W_i, W_j) \Leftrightarrow I^{\varrho}(W_j, W_i) \qquad \varrho \in \{+, -, 0, ?\}$.

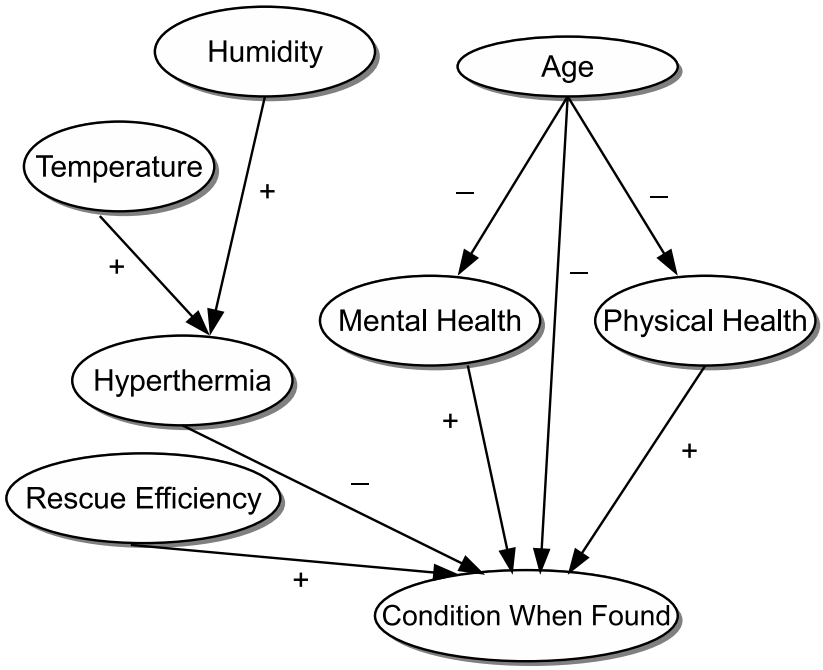


Figure 3.1: The QPN Equivalent to the Rescue Bayesian Network

It is important to note that the symmetry only applies to the signs of the influences and not on their corresponding strengths (which differ considerably in the two directions). The complete proof of the symmetry of qualitative influence can be found in (Wellman, 1990a).

2. **Transitivity:** The property of transitivity dictates that if a QPN given by the DAG G and containing three nodes W_i, W_j and W_k with $(W_i, W_j) \in E(G)$, $(W_j, W_k) \in E(G)$, and $I^{\varrho_1}(W_i, W_j)$ and $I^{\varrho_2}(W_j, W_k)$, then the sign of the influence of node W_i on node W_k is the 'product' of the signs of the two influences, that of W_i on W_j and that of W_j on W_k (Druzdzel and Henrion, 1993b). The 'product' of the signs is found via the $\bigotimes$ operator given in the left side of Table 3.1

   This property allows the construction of the sign of the net influence along a path of qualitative influences associated with the individual arcs.

3. **Composition:** The property of composition dictates that if a QPN given by

the DAG G and containing three nodes W_i, W_j and W_k with $(W_i, W_k) \in E(G)$, $(W_j, W_k) \in E(G)$, and $I^{\varrho_1}(W_i, W_k)$ and $I^{\varrho_2}(W_j, W_k)$, then the sign of the influence on node W_k is the 'sum' of the signs of the two influences, that of W_i on W_k and that of W_j on W_k (Druzdzel and Henrion, 1993b). The 'sum' of the signs is found via the $\bigoplus$ operator given in the right side of Table 3.1 and is used to evaluate the net influence of parallel connections.

Table 3.1: Sign multiplication ($\bigotimes$) and sign addition ($\bigoplus$) Operators (Wellman, 1990a)

$\bigotimes$	$+$	$-$	0	$?$
$+$	$+$	$-$	0	$?$
$-$	$-$	$+$	0	$?$
0	0	0	0	0
$?$	$?$	$?$	0	$?$

$\bigoplus$	$+$	$-$	0	$?$
$+$	$+$	$?$	$+$	$?$
$-$	$?$	$-$	$-$	$?$
0	$+$	$-$	0	$?$
$?$	$?$	$?$	$?$	$?$

Example 7. *In the rescue example given in Figure 3.1, the overall negative influence of node* Age *on node* Condition When Found *is obtained by evaluating the combined effect on* Condition When Found *via the three paths* Age → Physical Health → Condition When Found, Age → Condition When Found and Age → Mental Health → Condition When Found, *which is given by:*

$$
\begin{aligned}
I^-(\text{Age,Con. W. Found}) = & [I^+(\text{Age,Ph. Health}) \bigotimes I^-(\text{Ph. Health,Con. W. Found})\,] \\
& \bigoplus \\
& [I^-(\text{Age,Condition When Found})\,] \\
& \bigoplus \\
& [I^+(\text{Age,M. Health}) \bigotimes I^-(\text{M. Health,Con. W. Found})\,] \\
& = [- \bigotimes +] \bigoplus - \bigoplus [- \bigotimes +] \\
& = -
\end{aligned}
$$

Qualitative Synergies

Although qualitative influences define the basic interactions among variables, they are not always sufficient to capture all the interactions that exist in the network (Parsons,

2001). This is the case when it is necessary to identify the combined effect of a pair of variables in union on another variable. For this, the concept of qualitative synergies is created in order to model the interaction among the influences between three nodes in a network's diagraph (Renooij, 2001). Qualitative synergies are essentially of two classes depending on the type of interaction, mainly *additive* and *product* synergies, and can be positive, negative, constant or unknown as in the case for influences.

◇ **Additive Synergies**

Additive synergies express one way of how the value of two nodes jointly influence the probabilities of the values of a third node (Wellman, 1990a). Specifically, they describe the situations in which the combined influence of the parents on their common child is greater than the individual influence of each parent on the child. For example, a positive additive synergy of two nodes W_i and W_j on their common child W_k, written as $S^+(\{W_i, W_j\}, W_k)$, exists if the sum of their joint influence on W_K is greater than the sum of their separate influence regardless of the value of any direct ancestor W of W_k other than W_i and W_j (i.e. $W = \pi_G(W_k)/\{W_i, W_j\}$). Definition 10 formalizes the concept of positive additive synergy [3].

Definition 10. Positive Additive Synergy (Wellman, 1990a):
Let $G = (V(G), E(G))$ be a directed acyclic graph and let Pr be a joint probability distribution on $V(G)$ such that G is the I-map for Pr. Let $W_i, W_j, W_k \in V(G)$ be nodes in G with $W_i \rightarrow W_k, (W_j, W_k) \in E(G)$. Let $W = \pi_G(W_k)/\ \{W_i, W_j\}$. Then nodes W_i and W_j exhibit a positive additive synergy on node W_k, written as $S^+(\{W_i, W_j\}, W_k)$ iff *for any values w_i, w_j, w_k of W_i, W_j, W_k, respectively, we have:*

$$Pr(W_k \geqslant w_k | w_i, w_j, W) + Pr(W_k \geqslant w_k | \overline{w_i}, \overline{w_j}, W) \geqslant Pr(W_k \geqslant w_k | w_i, \overline{w_j}, W) + Pr(W_k \geqslant w_k | \overline{w_i}, w_j, W)$$

Example 8. *In the example given in Figure 3.1, both* Temperature *and* Humidity *exhibit an additive synergy on their common child* Hyperthermia (*written as*

[3] Definition 10 states positive synergies for binary variables and can be easily extended to infer positive synergies for multi-valued variables.

S^+({Temperature, Humidity}, Hyperthermia)) *as the presence of both parent makes it more likely to catch hyperthermia as opposed to the situations where the temperature is high without humidity or having a humid and a cold weather.*

Negative and constant additive synergies are analogously defined. Moreover, all nodes exhibit additive synergies on their common children in qualitative probabilistic networks (Wellman, 1990a).

◇ **Product Synergies**

In product synergies, one studies how the value of two parent nodes influence each other given as evidence the value of the two nodes' common child. The concept of product synergy was introduced in (Wellman and Henrion, 1991, 1993) while investigating the phenomenon *explaining away*, which was introduced by Pearl (1988) "as the kind of reasoning in which on observing an event, knowledge that makes one of its causes more likely makes another cause less likely" (Parsons, 2001). For example, in the network give in Figure 3.1, observing a value *False* for *Condition When Found* (in other words, observing that the civilian was in a bad state when rescued), then diagnosing the civilian with hyperthermia (observing a *True* value for *Hyperthermia*) makes one conclude that he/she was left in the debris for a relatively long time (i.e. a *False* value for *Efficiency of Rescue*).

The above is an example of a negative product synergy. In general, stating that nodes W_i and W_j exhibit a negative product synergy with respect to value w_0 of their common child W_k expresses the notion of, given w_0, observing higher values of W_i makes higher values of W_j less likely regardless of any other influence W on W_k, and as a result explaining W_i as the most likely cause of the value w_0 of W_k. Hence, product synergies describe the inter-causal dependence between two causes W_i and W_j given an observation w_0 of their common effect W_k. Definition 11 below presents the notion of negative product synergy, which is the concept behind the phenomenon explaining away.

Definition 11. Negative Product Synergy (Wellman, 1990a):
Let $G = (V(G), E(G))$ be a directed acyclic graph and let Pr be a joint probability distribution on $V(G)$ such that G is the I-map for Pr. Let $W_i, W_j, W_k \in V(G)$ be

nodes in G with $W_i \rightarrow W_k$ *and* $W_j \rightarrow W_k \in E(G)$. *Let* $\boldsymbol{W} = \pi_G(W_k)/\{W_i, W_j\}$ *and let* $\boldsymbol{w}$ *be the combination of observed values for* $\boldsymbol{W}$. *Then nodes* W_i *and* W_j *exhibit a negative product synergy on node* W_k, *written as* $Y^-(\{W_i, W_j\}, W_k)$ iff *for value* w_0 *of* W_k *and any values* $w_i, w_j, \boldsymbol{w}$ *of* $W_i, W_j, \boldsymbol{W}$ *respectively, we have:*

$$Pr(w_0|w_i w_j \boldsymbol{w}) \,.\, Pr(w_0|\overline{w_i}\,\overline{w_j}\boldsymbol{w}) \leqslant Pr(w_0|w_i\overline{w_j}\boldsymbol{w}) \,.\, Pr(w_0|\overline{w_i}w_j\boldsymbol{w})$$

Note that product synergies are defined with respect to value and not their respective variables as it only affects W_i and W_j when the specific value of W_k, i.e. w_0 is known. When this happens, W_i and W_j cease to be d-separated. In the example given above, the relationship only holds knowing that *Condition When Found* has been initialized. Also, Definition 11 assumes that all the other ancestors of W_k have been instantiated (or $W = \phi$). The definition of product synergy where there are uninstantiated ancestor nodes can be found in (Druzdzel and Henrion, 1993a). In our example, node *Age* and its descendants should be initialized in order for the product synergy to hold.

Formal Definition of a QPN

As seen in the previous sections, a QPN is given by the graph it represents and a set of hyperarcs that identify the qualitative relations governing the interactions among its variables. As a result, we can now provide a formal definition of a QPN as in Definition 12 below. .

Definition 12. Qualitative Probabilistic Network (Renooij, 2001):

A qualitative probabilistic network is a tuple $A = (G, \Delta)$ *such that*

- $G = (V(G), E(G))$ *is a directed acyclic graph with nodes* $V(G)$ *and arcs* $E(G)$.
- $\Delta = I \cup S \cup Y$ *is a set of hyperarcs for the graph G where:*
  - *I is a set of qualitative influences for G such that*
    - *I includes a qualitative influence* $I^{\varrho}(W_i, Y)$ *for every two nodes* $W_i, W_j \in V(G)$ *with* $W_i \rightarrow W_j \in E(G)$, *where* $\varrho \in \{+, -, 0, ?\}$ *and*

  - $*$ *I is closed under the properties of symmetry, transitivity and composition.*

- *S is a set of additive synergies for G such that*
  - $*$ *S includes an additive synergy $S^{\varrho}(\{W_i, W_j\}, W_k)$ for every three nodes $W_i, W_j, W_k \in V(G)$ with $(W_i, W_j), (W_j, W_k) \in E(G)$, where $\varrho \in \{+, -, 0, ?\}$ and*
  - $*$ *S is closed under the properties of symmetry, transitivity and composition.*
- *Y is a set of product synergies for G such that*
  - $*$ *Y includes a product synergy $Y^{\varrho}(\{W_i, W_j\}, w_0)$ for every three nodes $W_i, W_j, W_k \in V(G)$ with $W_i \rightarrow W_k, W_j \rightarrow W_k \in E(G)$ and W_k is known to have w_0 as a value, where $\varrho \in \{+, -, 0, ?\}$ and*
  - $*$ *Y is closed under the properties of symmetry, transitivity and composition.*

Inference in QPNs

Observed evidence is propagated through the network via the qualitative operators given in Table 3.1, producing the net effect of nodes on other nodes depending on the topology of the nodes considered. The original QPN inference algorithm (Wellman, 1990a,b) is the qualitative equivalent of Shachter's reduction algorithm for inference in quantitative BNs (Shachter, 1986). It uses repetitive arc-reversal (using the property of symmetry of qualitative influences with respect to their signs discussed in section 3.1.1) and node-reduction operators until the graph is reduced to having a singly-directed link between the observed node and the one we want to study the effect of the observation on (Wellman, 1990b). The algorithm suffers from the fact that finding the optimal reduction sequence to minimize the ambiguity is of an unknown computational complexity (Wellman, 1990b).

With the aim of obtaining an improved performance, Henrion and Druzdzel (1991); Druzdzel (1993); Druzdzel and Henrion (1993c) present a sign-propagation algorithm which relies on passing messages containing signs through the graph instead of the

graph reduction algorithm presented above. The basic idea is to determine the effect of an observation on other nodes in the network by passing messages between neighboring nodes and using the properties of symmetry, transitivity and composition to recursively propagate the signs of influences between the observed nodes and all other nodes in the network. The algorithm results in the assignment of change signs that indicates the effect of the observed node's probability on all the nodes of the network (Druzdzel and Henrion, 1993c).

The algorithm takes as input a QPN G, a set of previously observed nodes O, a newly observed node o and a sign of the current observation of o. The sign is $+$ for an observation of $true$ or $-$ for $false$. Initially, the signs of all the nodes $W \in V(G)$ are set to zero. The algorithm proceeds by entering the appropriate sign for o which is then used to update the signs for all its neighbors and every variable on which it exerts an induced intercausal influence by passing them a message containing the $\bigotimes$ sign-multiplication of the newly observed sign of o and the sign associated with the arc traversed. When the node receives the message, it updates its sign with the $\bigoplus$ sign-addition operator of the sign it receives and its original sign. This procedure recursively propagates using the properties of symmetry, transitivity and composition of influences until the net effect of the evidence is observed on the required node or all the nodes are known to be visited twice by the algorithm (Druzdzel and Henrion, 1993c).

The efficient polynomial-time message-passing algorithm can be found in (Druzdzel and Henrion, 1993c) and has been extended in (Druzdzel, 1993; Renooij and Gaag, 2002) to determine the effect of multiple observations at once as opposed to a single-observation.

Issues in QPNs

Despite the efficiency of the polynomial time arc-based message-passing algorithm for inference with QPNs (Druzdzel and Henrion, 1993c) compared to the NP-hard reasoning in BNs (Cooper, 1990), QPNs may suffer from over-abstraction. This is because the reasoning mechanism QPNs use is only concerned with finding the effect

of new evidence on each node in terms of the sign of the change in belief (increase or decrease) and does not take into account the strength of the effect, which may lead to ambiguity that must be resolved to continue reasoning. For instance, when a node receives two influences of conflicting signs through two parallel paths, the $\bigoplus$ operator labels the resulting influence as ambiguous (?). The problem is that ambiguous signs quickly spread throughout the network as the $\bigoplus$ and $\bigotimes$ operators generate more ambiguous signs once they receive one as input. An example is given in Figure 3.1 where applying the $\bigoplus$ operators to obtain the net influence of nodes *Age*, *Hyperthermia* and *Rescue Efficiency* over node *Condition When Found* results in an unknown influence ? $(-\bigoplus -\bigoplus +)$, which then propagates through the network as the effect of W_k on W is computed via the $\bigotimes$ operator. The problem of ambiguous signs has been attributed to two main causes (Renooij, 2001). The first is the existence of *trade-offs* in which two nodes in the network are connected by multiple parallel paths and the signs of the influences along these paths are conflicting as in the example above, and is especially problematic in the case of inter-causal reasoning (Wellman and Henrion, 1993). The second cause of ambiguity is that when influences are *non-monotonic*, meaning that the sign of the influence from some node W_i to another node W_j can only be determined when the value of a third node W_k is known (Renooij and Gaag, 2000). The two problems are caused by the coarseness of the representation and have been addressed separately in efforts listed below.

◇ **Trade-off Resolution Mechanisms** A trade-off occurs when two influences of conflicting signs are exerted on one node as in the example given in the previous section, and is due to the absence of any notion of strength in QPNs as only the signs of influences are recorded (Renooij, 2001; Parsons, 2001). The literature contains several approaches to resolving trade-offs. We summarize them below:

(a) Renooij and Gaag (2008) distinguish between strong and weak influences (where a strong positive influence of W_i on W_j, termed $I^{++}(W_i, W_j)$, carries more weight than a weak one, termed $I^{+}(W_i, W_j)$ (with the same nomenclature used for negative, zero and unknown influences). Renooij and Gaag (2008) also provide a method for comparing indirect qualitative influences along different paths with respect to their strengths for trade-off

resolution by retaining the length of the paths over which influences have been multiplied. For this, every influence's sign is augmented by a superscript, called the signs multiplication index, and is used as an indicator of its strength. Higher values of multiplication indices indicate a longer path and as a result, a weaker influence. This enables generalizing the message-passing algorithm of (Druzdzel and Henrion, 1993c) by adapting the $\bigoplus$ and $\bigotimes$ operators to the different types of influences as given in Tables 3.2 and 3.3.

Table 3.2: Enhanced Sign Addition ($\bigoplus$) Operator (Renooij and Gaag, 2008)

$\bigoplus$	$++^j$	$+^j$	0	$-^j$	$--^j$	?
$++^i$	$++^{ij}$	$++^i$	$++^i$	a)	?	?
$+^i$	$++^j$	$+^{i,j}$	$+^i$	?	d)	?
0	$++^j$	$+^j$	0	$-^j$	$--^j$	?
$-^i$	b)	?	$-^i$	$-^{i,j}$	$--^{i,j}$	?
$--^i$	?	c)	$--^i$	$--^i$	$--^{i,j}$	?
?	?	?	?	?	?	?

b) $++^{-i,j}$, $if j \leqslant i$; ?, otherwise
c) $--^{i,-j}$, $if i \leqslant j$; ?, otherwise
c) $--^{i,-j}$, $if j \leqslant i$; ?, otherwise

Table 3.3: Enhanced Sign multiplication ($\bigotimes$) Operator (Renooij and Gaag, 2008)

$\bigotimes$	$++^j$	$+^j$	0	$-^j$	$--^j$	?
$++^i$	$++^{i+j}$	$+^j$	0	$-j$	$--^{i+j}$	?
$+^i$	$+^i$	$+^{i+j}$	0	$-^{i+j}$	$-^i$	?
0	0	0	0	0	0	0
$-^i$	$-^i$	$-^{i+j}$	0	$+^{i+j}$	$+^i$	?
$--^i$	$--^{i+j}$	$-^j$	0	$+^j$	$++^{i+j}$	?
?	?	?	0	?	?	?

(b) Another approach is to concentrate on identifying nodes at which trade-offs occur (Renooij et al., 2000), making it possible to establish automatically that if the influence of W_i on W_k is greater than that of W_j on W_k then the combined influence on W_i and W_j on W_k is $I^+(W_i \wedge W_j, W_k)$. This is done by identifying a pivot and zooming to the part of the network where it resides and subsequently identifying the information needed to resolve the trade-off. A similar approach can be found in (Liu and Wellman, 1998).

(c) Renooij et al. (2003) propose the use of a ranking function to rank influences in terms of strengths and use the resulting rankings to resolve the conflicts when trade-offs occur. The ranking function is the κ function (to be discussed in section 3.2), which utilizes natural numbers to rank beliefs such that the greater the number associated with a belief state, the less believable (or more surprising) it is. The κ values associated with the influences of a QPN are then used to assign a value to the influence instead of ? by adding strength factors to the $\bigoplus$ and $\bigotimes$ operators given in Tables 3.4 and 3.5. In the tables, each influence is associated with a minimum and a maximum strength factor, which are updated as influences are combined.

The approach retains the efficiency of arc-based reasoning of QPNs while reducing the unwanted coarseness in the representation by using κ values as measures of strength of QPN influences and resorting to them for trade-off resolution.

Table 3.4: Sign Addition ($\bigoplus$) Operator for Combining Signs and Strength Factors

$\bigoplus$	$+[r,s]$	$-[r,s]$	0	?
$+[p,q]$	$+[p+r+1, q+s]$	$-[p+r+1, q+s]$	0	?
$-[p,q]$	$-[p+r+1, q+s]$	$+[p+r+1, q+s]$	0	?
0	0	0	0	?
?	?	?	0	?

◇ **Non-monotonic Influences**

Table 3.5: Sign Multiplication ($\bigotimes$) Operator for Combining Signs and Strength Factors

$\bigotimes$	$+[r,s]$	$-[r,s]$	0	$?$
$+[p,q]$	$+[u,v]$	$a)$	$+[p,q]$	$?$
$-[p,q]$	$b)$	$-[u,v]$	$-[p,q]$	$?$
0	$+[r,s]$	$-[r,s]$	0	$?$
$?$	$?$	$?$	$?$	$?$

$[u,v] =$ $[\min\{p,r\}, \min\{q,s\}]$

a) $+[p,q], if p+1 < s;$
$+[\infty,q], if p < s;$
$-[r,s], if r+1 < q;$
$-[\infty,s], if r < q$
?, otherwise

b) see a) with + and − reversed

In the case of non-monotonic influences, the ambiguity of the influence is due to the inherent ambiguity of the interactions among the variables and not due to ignorance. In other words, the sign of the influence is not independent of the variables of the network other than the two for which the influence exists (Renooij et al., 2002).

Definition 13. Non-monotonic Influence (Renooij et al., 2002):
Let $G = (V(G), E(G))$ be a directed acyclic graph and let Pr be a joint probability distribution on $V(G)$ such that G is the I-map for Pr. Let $W_i, W_j \in V(G)$ be nodes in G with $W_i \rightarrow W_j \in E(G)$. Let $W = \pi_G(W_j)/\{W_i\}$. The ambiguous influence of node W_i on node W_j, written as $I^?(W_i, W_j)$ is a non-monotonic influence iff *for all values w_j of W_j and all values w_{i_1}, w_{i_2} of W_i, with $w_{i_1} > w_{i_2}$, the following inequality yields contradictory signs for different combinations of the value of W.*

$$Pr(W_j > w_j | w_{i_1}, W) - Pr(W_j > w_j | w_{i_2}, W)$$

Given the specific value of W however, i.e. a specific state of the network, the influence ceases to be ambiguous and has a specific sign of $+$, $-$ or 0 (Bolt et al., 2003a; Renooij and Gaag, 2000). Taking this thought further, Bolt et al. (2003a,b) introduce the concept of a *situational influence* that provides the nature of the influence given a specific state of the network. Definition 14 below introduces the concept of a positive situational influence. Negative, zero and unknown situational influences are analogously defined.

Definition 14. Situational Sign (Renooij and Gaag, 2000):
Let $G = (V(G), E(G))$ be a directed acyclic graph and let Pr be a joint probability distribution on $V(G)$ such that G is the I-map for Pr. Let $W_i, W_j \in V(G)$ be nodes in G with $W_i \rightarrow W_j \in E(G)$. Let $W = \pi_G(W_j)/\{W_i\}$. Let $\mathcal{G}$ be the present state of G in which $W = w$. Node W_i exhibits a positive situational influence on node W_j, written as $I^{?(+)}(W_i, W_j)$ iff *for all values w_j of W_j and all values w_{i_1}, w_{i_2} of W_i, with $w_{i_1} > w_{i_2}$ and for value w of W, the following inequality holds.*

$$I^{?(+)}(W_i, W_j) \text{ iff } Pr(W_j > w_j | w_{i_1}, w) - Pr(W_j > w_j | w_{i_2}, w)$$

Hence, while influences and synergies exhibit a validity that is general to the network, the signs of situational influences hold for a specific state of the network and depend on Pr. Bolt et al. (2003a) present an *adapt sign-propagation algorithm* which is a modification of the elegant algorithm given in section 3.1.1 to incorporate situational signs in the original algorithm.

3.1.2 Qualitative Certainty Networks (QCNs)

The popularity of QPNs motivated work to extend the qualitative propagation mechanism to other uncertainty formalisms (Parsons, 2003) in an aim to use the approach from QPNs to not only propagate qualitative probability, but also possibility (Zadeh, 1978; Dubois and Prade, 1988) and evidence theory (Shafer, 1988) in a uniform way. The idea is to redefine the concept of qualitative influences so that they can be used to abstraction probability, possibility and belief functions. As this document is mainly concerned with abstracting probabilistic methods, the discussion of QCNs is

constricted to its probabilistic interpretation. The reader may refer to (Parsons and Mamdani, 1993) for the alternative interpretations.

Definition of a QCN

As in QPNs, a QCN is a DAG $G = (V(G), E(G))$ in which $V(G)$ is the set of nodes and $E(G)$ is the set of arcs representing the dependencies among the nodes. The set of influences defined over $E(G)$ however, differ from those of QPNs in that they express the change of probabilities in terms of the derivatives that relate the different values of the variables together. More specifically, a positive qualitative derivative relating two variables is given in Definition 15 below.

Definition 15. Qualitative Derivative (Parsons, 2003):

Let Pr be a probability distribution defined over a set of variables $\{W_1, ...W_N\}$. For any variables W_i, W_j, $1 \leqslant i, j \leqslant N$ each having multiple values, the qualitative derivative $[\frac{\partial Pr(w_{i_1})}{\partial Pr(w_{j_1})}]$ relating the probability of W_i taking value w_{i_1} to the probability of W_j taking value w_{j_1} has the value $[+]$, iff, *for all other values w_{j_2} of W_j and any other variable W:*

$$Pr(w_{i_1}|w_{j_1}, W) \geqslant Pr(w_{i_1}|w_{j_2}, W)$$

The square brackets surrounding the value of the derivative is used in (Parsons and Mamdani, 1993; Parsons, 2003) to denote that it is the qualitative value we are interested in. Derivatives having values $[-]$ and $[0]$ are defined by replacing the $\geqslant$ in the definition by $\leqslant$ and $=$ respectively. Moreover, a derivative is given a value $[?]$ if it cannot be determined to be $[+]$, $[-]$ or $[0]$.

The immediately noticed difference between the qualitative derivative and an influence is that the former relates two values of two variables, whereas an influence describes the general relation describing the behavior of the two variables. As a result, while only one type of influence is present between a parent and a child in a QPN, a set of qualitative derivatives relates the two in a QCN. However, it is worth noting that when dealing with binary variables, QCN qualitative derivatives and QPN influences become equivalent (Parsons and Mamdani, 1993; Parsons, 2001).

Properties of QCN Derivatives

As in QPN influences, QCN derivatives exhibit a number of properties that make possible their propagation along the networks various paths.

1. **Symmetry:** The qualitative derivative $[\frac{\partial Pr(w_{i_1})}{\partial Pr(w_{j_1})}]$ is said to be symmetric, i.e. $[\frac{\partial Pr(w_{i_1})}{\partial Pr(w_{j_1})}] = [\frac{\partial Pr(w_{j_1})}{\partial Pr(w_{i_1})}]$ if:

   a) $[\frac{\partial Pr(w_{i_1})}{\partial Pr(w_{j_1})}] = [+]$ or $[-]$, or

   b) $[\frac{\partial Pr(w_{i_1})}{\partial Pr(w_{j_1})}] = [0]$ and $[\frac{\partial Pr(w_{i_k})}{\partial Pr(w_{j_1})}] = [0]$ for all $k \neq 1$. (Parsons and Mamdani, 1993).

2. **Transitivity:** The property of transitivity dictates that the qualitative value of QCN derivatives along multiple paths can be combined using the $\bigotimes$ operator defined Table 3.1, similarly to QPN influences (Parsons, 2001).

3. **Composition:** The property of composition dictates that the qualitative value of QCN derivatives along parallel paths can be combined using the $\bigoplus$ operator defined Table 3.1, similarly to QPN influences (Parsons, 2001).

QCNs Versus QPNs

Given the properties of QCN derivatives in the previous section, they can be propagated along the different paths of a given network in the same way that QPN influences are. In fact, the same arc-reversal algorithm (Druzdzel and Henrion, 1993c) can be used to propagate the qualitative signs of QCN derivatives.

Despite this, the discussion of QCNs has shown two fundamental differences with QPNs summarized here. Apart from the fact that QCNs is general enough to admit alternative formalizations to abstract other uncertainty formalisms, they define the relations among the instantiations of variables and not among the variables themselves. This is different from QPN influences which define a general relation between the variables (with non-monotonic influences being an exception).

As a result, QCNs are less ambiguity-prone than QPNs. However, they still suffer from over-abstraction, for which resolution is achieved by introducing strengths over the derivatives' qualitative values (Parsons, 2001).

3.2 Order of Magnitude Abstractions

The previous section introduced QPNs, which abstract regular BNs by capturing the sign of change in probability given the evidence instead of the probability value. The coarseness of the representation identified with QPNs has inspired a different class of abstraction of quantitative systems (the coarseness experienced in QPNs is general to systems of strict abstractions, including ones that do not aim at abstracting numerical uncertainty calculi, see (Raiman, 1986) for an example in mechanics). Instead, these systems deal with orders of magnitude or probability instead of point probability and form calculi with built-in reasoning mechanisms.

This class of calculi aims at reducing the task of having to specify point probabilities by providing a belief measure that is more abstract and intuitive than point probability (Darwiche and Goldzmidt, 1994). The main idea is to abstract probability theory by devising tentative rankings for beliefs consistent with their corresponding probabilities. As a result, they escape having to assign the precise numerical value of the corresponding probabilities. This type of abstractions has resulted in a series of work in both philosophy and artificial intelligence (Huber, 2006; Spohn, 1988a; Goldszmidt and Pearl, 1996; Shenoy, 1990).

This body of work is based on the definition of a function which maps the belief state of a propositions to a natural number that can be interpreted as the order-of-magnitude of the inverse of subjective probability (Goldszmidt and Pearl, 1996). This number assigned by the function is descriptive of the epistemic state of a variable in that the higher the number assigned, the less believable or more surprising the corresponding variable is (Shenoy, 1990). Such function has been termed an ordinal conditional function, a natural conditional function (Spohn, 1988b) and a disbelief function (Shenoy, 1990) (the latter term is the one we are to use for the rest of this review as it intuitively describes the semantics of the corresponding calculus).

With the general aim of providing an intuitive *ranking* of propositions and beliefs, the paradigm of ranking functions has been thoroughly studied and implemented in applications such as diagnosis (Darwiche and Goldzmidt, 1994) and as complete decision theories (Pearl, 1993; Wilson, 1995).

3.2.1 Ranking Theory

Given a universe of events Ω, with each event represented by a variable denoted by $W \in \Omega$, a disbelief function ϱ (Spohn, 1988a,b; Shenoy, 1990) is defined as the mapping from Ω to the set of non-negative integers $\mathbb{N}^+$:

$$\varrho : \Omega \rightarrow \mathbb{N}^+$$

The mapping is defined in a way that ensures that ϱ satisfies the following axiom (Spohn, 1988a):

$$\min_{W \in \Omega} \varrho(w) = 0$$

Where w is a value of W. In other words, at least one out of all the variables $W \in \Omega$ must have $\varrho(w) = 0$ [4] (with this being the smallest possible assignment since ϱ has the set of non-negative integers as a range).

The disbelief function defined above is a complete representation of the epistemic state of a domain that also includes degrees of belief and disbelief regarding the variables of the domain; it achieves this via a set of properties that dictate the assignment of an epistemic state to a variable (or a set of variables) as given below (properties collected form (Spohn, 1988a,b; Shenoy, 1990)):

1. $\varrho(w) \in \mathbb{N}^+$ for any value w of variable $W \in \Omega$;

[4] In this discussion, we assume all the variables are binary with each variable W as having two possible values, w and $\neg w$. The reason behind this is not for simplicity, but due to the fact that ranking functions were created to be a part of a logical language. i.e. every variable is representative of a variable.

2. The higher $\varrho(w)$, the more disbelieved the statement $W = w$ is;
3. There exists a configuration $w \in \Omega$ for which $\varrho(w) = 0$;
4. $\varrho(true) = 0$
5. $\varrho(\phi) = +\infty$
6. $\varrho(\bigvee_{W \in \Omega}) = \min_{W \in \Omega} \varrho(w)$
7. For each $W \in \Omega$, either $\varrho(w) = 0$ or $\varrho(\neg w) = 0$

Hence, we can deduce that if $\varrho(w_1) > 0$, then $\varrho(w_1)$ can be interpreted as the degree of disbelief in variable W_1 having w_1 as its truth value (Spohn, 1988a). This is stated in the second property above which describes the ranking semantics of ϱ in that for two truth values w_1 and w_2, w_1 is more disbelieved than w_2 if $\varrho(w_1) > \varrho(w_2) > 0$. Also, if $\varrho(\neg w) > 0$ then $\varrho(\neg w)$ can be interpreted as the degree of belief for w, i.e. w_1 is more believed than w_2 if $\varrho(\neg w_1) > \varrho(\neg w_2) > 0$.

Property 3 above not only defines 0 as the minimum value $\varrho(.)$ can have, but also mandates that at least one configuration in Ω must have a ranking of 0. Since $\varrho(.)$ is interpreted as the degree of disbelief in a variable, having a value of 0 indicates that the variable is the least disbelieved, or the most believed i.e. having a true value, which is stated in the property 4. $\varrho(.)$ takes this notion a step further and uses this to ensure deductive closure in properties 3 and 6 which mandate that the minimum $\varrho(.)$ value of a domain defined by Ω must always be zero (or in other words, the disjunction of the $\varrho(.)$ values of a domain produces zero, because disjunctions of $\varrho(.)$ values are obtained through the minimum operator as per property 5) (Shenoy, 1990).

It is important to note that ϱ can also be extended to any non-empty subset $\mathbf{W}$ of Ω as given below:

$$\varrho(\mathbf{W}) = \min_{W \in \mathbf{W}} \varrho(w) \;\; \forall \mathbf{W} \subseteq \Omega$$

3.2.2 Probabilistic Interpretation, the κ Calculus

Despite the fact that ranking functions were proposed as a (alternative) theory of change, there exists a strong relation between them and standard probability (Spohn, 1988a,b). Work on this relation resulted in the inception of a mapping that enables the interpretation of $\varrho(.)$ in terms of probabilities (Spohn, 1988b; Goldszmidt and Pearl, 1996; Giang and Shenoy, 1999) and was used to create a calculus enabled with probability-like reasoning on ranking functions (Goldszmidt and Pearl, 1996).

The ranking provided by such functions is interpreted as the order-of-magnitude of the inverse of probabilities (Huber, 2006) and can be imagined by projecting probability measures onto a quantized logarithmic scale and then treating beliefs that map onto different quanta as being of different orders of magnitude (Goldszmidt and Pearl, 1996) as shown in figure 3.2. This is achieved via a procedure which begins by considering a probability function Pr defined over a set Ω of possible variables (or states of the world) such that $Pr : \Omega \rightarrow [0, 1]$ and representing the probability of each value w of variable $W \in \Omega$, $Pr(w)$, by a polynomial function of one unknown, ϵ, an infinitesimally small positive number ($0 < \epsilon < 1$). For instance, possible assignments are $Pr(w_1) = \alpha$, $Pr(w_2) = \beta\epsilon$, $Pr(w_3) = \gamma\epsilon^2$ (Parsons, 2001). The resulting rank of a variable W, termed $\kappa(w)$ (read *kappa*[5]), is represented by the power of the most significant ϵ-term in the polynomial representing $Pr(W)$ (the lowest power of ϵ in the polynomial). In other words, $\kappa(w) = n$ such that n is the smallest integer that insures that $\dfrac{Pr(w)}{\epsilon^n}$ is finite but not infinitesimal for infinitesimal ϵ (i.e. $\lim\limits_{\epsilon \rightarrow 0} Pr(w)/\epsilon^n$ is nonzero (Goldszmidt, 1995)), or $\kappa(w) = n$ if and only if $Pr(w)$ is of the same order of magnitude as ϵ^n . Accordingly, the relation between the probability $Pr(w)$ and κ values $\kappa(w)$ can be described as given in the equation below (Goldszmidt and Pearl, 1996):

$$\epsilon < \frac{Pr(w)}{\epsilon^n} \leqslant 1 \text{ or equivalently: } \epsilon^{n+1} < Pr(w) \leqslant \epsilon^n$$

Where ϵ^n is the most significant ϵ-term of the polynomial representing $Pr(w)$.

[5]While authors of (Goldszmidt and Pearl, 1996) and their collaborators use the term κ, the group of Shenoy and others insists on the original ϱ in their work.

Another definition that explicitly specifies how n is obtained is:

$$\kappa(w) = \begin{cases} \min\{n \, such \, that \, \lim_{\epsilon \to \infty} \frac{Pr(w)}{\epsilon^n} \neq 0\} & iff \;\; Pr(w) > 0 \\ \infty & iff \;\; Pr(w) = 0 \end{cases}$$

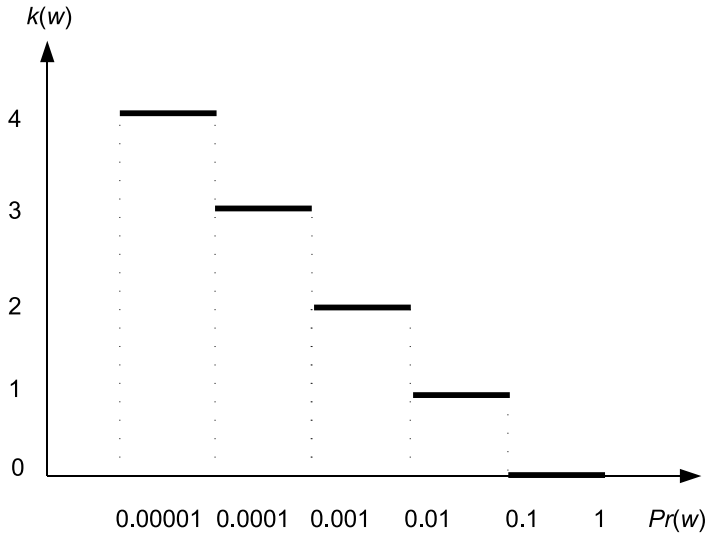


Figure 3.2: Mapping from Probabilities to Kappas Using $\epsilon = 0.1$ (Parsons, 2001)

A direct consequence of how $\kappa(w)$ is obtained is that since the most significant term is that with the smallest n, it corresponds to the inverse of the likelihood of w, and is therefore representative of the degree of surprise associated with believing w, or the degree of incremental surprise or abnormality associated with finding w to be true (Goldszmidt and Pearl, 1996). The value of $\kappa(w)$ is assigned so that probabilities having the same order of magnitude belong to the same κ class, and that $\kappa(w)$ grows inversely to the order of magnitude of the probability value $p(w)$ as seen in figure 3.2.

Moreover, the $\kappa(.)$ function supports classifying propositions into believed, disbelieved and uncommitted, which is different from classical probability where propositions are merely graded by their probability value (Darwiche and Goldzmidt, 1994).

This can be seen in a more intuitive manner in table 3.6 (obtained from (Goldszmidt and Pearl, 1996)), which shows an example of how kappas can be mapped to linguistic quantifiers of beliefs.

Table 3.6: Mapping κ Values to Linguistic Quantifiers

.	.	.
.	.	.
$p(w) = \epsilon^0$	w and $\neg w$ are possible	$\kappa(w) = 0$
$p(w) = \epsilon^1$	$\neg w$ is believed	$\kappa(w) = 1$
$p(w) = \epsilon^2$	$\neg w$ is strongly believed	$\kappa(w) = 2$
.	.	.
.	.	.

Apart from the above mapping, it is important to note that Spohnian ranking functions can be mapped to other quantitative formulations for reasoning about uncertainty. For instance, (Dubois and Prade, 1991) shows that the basic disbelief function presented in (Spohn, 1988a) can be interpreted as the negative of the logarithm of a possibility function (Zadeh, 1978). Hence, ranking functions in general (and the κ function in specific) are considered "well-positioned in the web of quantitative approaches to represent and reason about uncertain beliefs" (Giang and Shenoy, 1999). The reasons behind the extensive work on mapping the κ to probabilistic measures will be discussed in section 3.2.3.

Reasoning in the κ Calculus

The mapping given above has been used to justify the properties of the κ calculus using order-of-magnitude operations without appealing to probabilistic interpretation (Spohn, 1988a). These properties are accordingly used for belief revision and update as part of the κ reasoning system, which yields an integer-based calculus that enables combining κ's via rules that are derived from those of probability theory by replacing

multiplication by addition and addition by minimum (Spohn, 1988a; Goldszmidt and Pearl, 1996). The resulting properties are given below.

1. $\kappa(\Omega) = \min_{W \models \Omega} \kappa(w)$
2. $\kappa(true) = 0$
3. $\kappa(w) \vee \kappa(\neg w) = 0$
4. $\kappa(w_1 \vee w_2) = \min(\kappa(w_1), \kappa(w_2))$
5. $\kappa(w_1 | w_2) = \kappa(w_1 \wedge w_2) - \kappa(w_2)$

The third property is to ensure deductive closure and is synonymous to having the sum of the probabilities of the two values of the variable adding to one.

Reasoning with κ's can then be performed using the above properties in one of the following two ways:

1. Constructing a graphical causal model quantified with order-of-magnitude probabilities instead of point probabilities and using the κ properties for updating beliefs according to evidence across the networks. The networks constructed, termed *Kappa Networks*(Darwiche, 1992) are populated using κ values. The κ values can either be obtained from their respective probabilities using algorithm 1 below, or estimated when the probabilities are unknown or unavailable.

   The case for using Kappa Networks when probabilities are unavailable should now be clear to the reader. On the other hand, converting probability values to their κ equivalents as per algorithm 1 has been justified as leading to a more robust inference results (Darwiche and Goldzmidt, 1994; Darwiche, 1992) (when used correctly, as will be discussed in section 3.2.4) and computational simplicity from a human perspective (Giang and Shenoy, 1999) as results given in κ's are easier to assess by human experts than numerical probabilities.

   In terms of efficiency, (Goldszmidt, 1995) presents an algorithm for performing prediction tasks across Kappa Networks. The algorithm *Predict* uses the structure

Algorithm 1 An Algorithm for Finding a Solution to the Equation $\epsilon^{\kappa+1} < Pr \leqslant \epsilon^{\kappa}$ to Translate a Probability Pr to a κ, taken from (Darwiche and Goldzmidt, 1994).

1. If $Pr = 0$, then return ∞
2. $\kappa \leftarrow 0$
3. $Pr \leftarrow Pr/\epsilon$
4. If $Pr > 1$, then return κ, otherwise $\kappa \leftarrow \kappa + 1$
5. Go to 3

and quantification of the network to perform its prediction tasks and is of polynomial asymptotic complexity (Goldszmidt, 1995), similarly to the *polytree* algorithm given in (Pearl, 1988). The algorithm however, is sound but not complete. It is sound in that a believed value produced by an algorithm will always match the results obtained through ranking function manipulations. However, it may sometimes fail to recognize a believed result for a variable (Goldszmidt, 1995) and is hence incomplete.

Ignoring the incompleteness of (Goldszmidt, 1995)'s algorithm, having a fast algorithm for computing with κ 's has also been used to map existing probabilities to kappas and using them for fast belief update via the κ computation instead of the probabilistic one (Goldszmidt and Pearl, 1992). This can be useful despite the fact that the algorithm for exact inference with Kappa networks is in fact NP-hard (Darwiche, 1992) similarly to exact inference in standard BNs (Cooper, 1990).

2. The feature of deductive closure combined with the ability to prioritizing defaults has enabled using the κ calculus in a framework for reasoning about defeasible beliefs in which κ's serve the role of default priorities (Hunter, 1990, 1991) and provide probabilistic semantics to the if-then rules used in monotonic reasoning as done in (Goldszmidt and Pearl, 1996; Goldszmidt, 1992; Boutilier, 1997).

3.2.3 Using Rank-based Calculi

κ calculus provides an abstraction which only requires specifying the κ values of propositions, which is an easier task than specifying the exact probabilities associ-

ated with the specific value of the variable. This has made it an attractive choice for representing and reasoning about uncertain knowledge in lieu of its quantitative equivalents when they are not unobtainable or simply unknown.

In addition, when it is possible to obtain the probabilities required for an application domain, the κ calculus remains tremendously useful. The results obtained for queries conducted using standard BNs can be abstracted into well-defined ranks for the purpose of being viewed by experts that may be reluctant to evaluate exact point probabilities (Adams, 1975). This can serve a great role in evaluating the inference mechanisms of the quantitative system with respect to specific applications (Darwiche and Goldzmidt, 1994).

It is also worth noting that κ's have been used not only as a stand-alone system, but also in conjunction with other qualitative measures. For instance, (Renooij et al., 2003) uses κ's to capture the strength of qualitative influences in QPNs and hence reducing the chance of having unknown influence by resolving tradeoffs across the network. Moreover, (Tamma and Parsons, 2001) uses κ's in conjunction with a symbolic qualitative system (not covered in this review) to create a platform in which arguments are ranked according to κ values associated with them.

3.2.4 Issues with Ranking Functions

Despite the above uses, an emphasis should be made with respect to the relationship between κ's probabilities in that it rests on the assumption that the ϵ used is infinitesimal (Spohn, 1988a; Goldszmidt and Pearl, 1996). According to (Darwiche and Goldzmidt, 1994), it is only when an infinitesimal ϵ is chosen that the following computations are guaranteed to produce the same results (points below are taken from (Darwiche and Goldzmidt, 1994)):

1. "Computing posterior probabilities using probability and then abstracting them into κ rankings" (Darwiche and Goldzmidt, 1994)

2. "Abstracting probabilities into κ rankings and then computing posterior κ rankings using the κ calculus" (Darwiche and Goldzmidt, 1994).

The issue is that for infinitesimal ϵ values, the above two computations produce a trivial value of κ (the actual output is zero) unless the equivalent probabilities are arbitrarily high or low (for example, above 0.95 or below 0.05), in which case meaningful κ values can be obtained.

Therefore, one is forced to use non-infinitesimal values of ϵ in practice for the purpose of computations. This, however, presents a continuous task of monitoring of how close ϵ is to zero and studying its effect on the above two computations as practical studies of a car troubleshooting problem conducted by Darwiche and Goldzmidt (1994); Henrion et al. (1994) show that the farther from zero the value of ϵ is, the less dependable the mapping becomes as the above two computations will cease producing identical results (Darwiche and Goldzmidt, 1994). This consequently leads to discrepancies between the rankings produced by κ and the orderings implied by quantitative systems, despite the fact that the results may be well-justified from a defeasible reasoning perspective (Giang and Shenoy, 1999). To illustrate the problem, consider the following example taken from (Giang and Shenoy, 1999).

Example 9. *let* $\Omega = \{W_1, W_2, W_3, W_4\}$ *be governed by a probability distribution as given in table 3.7. And let* $\epsilon = 0.2$ *be the value used to extract the* κ *values of the corresponding probabilities.*

According to Giang and Shenoy (1999), letting $A = \{W_2\}$ *and* $B = \{W_3, W_4\}$ *entails that* $Pr(A) < Pr(B)$*, which disagrees with the results obtained after the transformation with* $\epsilon = 0.2$ *as it gives* $\kappa(A) = 0 < \kappa(B) = 1$*.*

Table 3.7: Example Showing the Discrepancies between κ and Pr

W	$Pr(W = w)$	$\kappa_{\epsilon=0.2}$
W_1	0.5185	0
W_2	0.2308	0
W_3	0.1538	1
W_4	0.0969	1

Motivated by the fact that the experiments performed in (Pearl, 1993; Henrion et al., 1994) do not provide an ambiguous answer in regards to the value of ϵ that resolves the tension between having a trivial κ and a consistent result, (Giang and Shenoy, 1999) defines a transformation function T between κ 's and probability values and subsequently finds the most appropriate ϵ value for defining κ's. T is given in algorithm 2 for comparison with the initial transformation given in algorithm 1.

The procedure founded by (Giang and Shenoy, 1999) adds assertive power to the computational simplicity offered by the κ function (which has also been supported by the empirical results found in (Kahneman et al., 1982) with respect to decision making using ranks). Moreover, it offers a bridge between reasoning using plain beliefs and rational behavior, which can be considered as a response to theses stating that rational behavior is based on probabilistic measures (See for example (Savage, 1972; Neumann and Morgenstern, 1953)).

Algorithm 2 An Algorithm Transforming Probability Pr to a κ, taken from (Giang and Shenoy, 1999).

Input: A sequence of probabilities $(Pr_1, Pr_2, ..., Pr_n)$.
Output A sequence of disbelief degrees $(\kappa_1, \kappa_2, ..., \kappa_n)$.
$r = 0$ r is a disbelief counter, initially 0
$M = 1$ M is remaining mass, initially 1

for $i = 1$ to n
$\kappa_i = r$ κ_i is disbelief degree of w_i
$M = M - Pr_i$ Pr_i is probability of w_i
if $Pr_i > M$ then $r = r + 1$
end

3.3 A Note on Qualitative Approaches

The previous sections have outlined the major schools of thought with respect to qualitative approaches for abstracting probabilistic reasoning. It was found that such

approaches can be of tremendous aid whether used independently (e.g. using QPNs in medical diagnosis (Beumer, 2006) or the κ calculus in fault diagnosis (Darwiche and Goldzmidt, 1994)) or in conjunction with quantitative probabilistic systems. However, this does not mean that these calculi aim at completely replacing standard quantitative approaches as they can be very well used in conjunction with quantitative probability. For example, (Lucas, 2005) uses QPNs for aiding the process of constructing standard BNs.

The above uses for qualitative approaches are in addition to providing different perspectives on the domain at hand. As we have shown, having a transformation that enables shifting from probabilistic to qualitative abstractions (and other measures of uncertainty) can be tremendously useful for the development of systems in which experts may be reluctant to commit to numerical probabilities. This is especially evident in the mapping between the κ values and possibility measures (Dubois and Prade, 1991) which consequently leads to a mapping between the latter and probabilistic measures, enabling an "à la carte" (Giang and Shenoy, 1999) system that can be used by experts depending on the needs of the specific application. This capability also enables developing systems that are capable of performing inference using the qualitative and numerical information combined or separately and interpreting the results in the manner the user is most comfortable with (Shenoy, 1998; Giang and Shenoy, 1999).

Chapter 4

Surprise-based Qualitative Uncertainty Calculi [1]

I know too well that these arguments from probabilities are imposters, and unless great caution is observed in the use of them, they are apt to be deceptive.

- Plato, *Phaedo*

Semi-qualitative ordinal ranking functions exemplified by the κ function (Spohn, 1988a; Goldszmidt and Pearl, 1996) were introduced among the different qualitative frameworks presented in Chapter 3. These functions use natural numbers to measure the degrees of disbelief associated with events by capturing the order-of-magnitude of the reverse of probabilities (Goldszmidt and Pearl, 1996; Spohn, 1988b; Darwiche, 1992; Huber, 2006) of the occurrence of the event. The κ function (Goldszmidt and Pearl, 1996) provides a ranking that can be imagined by projecting the inverse of the probability measures onto a quantized logarithmic scale and then treating beliefs

[1] This chapter incorporates the outcome of a joint research undertaken under the supervision of Professors Ahmed Tawfik and Alioune Ngom. The key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of co-authors was primarily through the provision of advice when needed.

that map onto different quanta as being of different orders of magnitude (Goldszmidt and Pearl, 1996). The power underlying the κ calculus stems from the ease of assignments of the ranking values to events as opposed to assigning probabilities and the availability of robust mechanisms for belief updates via conditionality (Spohn, 1988b; Darwiche and Goldzmidt, 1994) and has resulted its repeated use in Artificial Intelligence (Tamma and Parsons, 2001; Renooij et al., 2003).

Despite the features that κ offers, Chapter 3 showed that the consistency between the ranking that it provides and that of probabilities is not guaranteed as it depends on the choice of parameters involved in the abstraction. As a result, assigning κ's to events is always associated with careful tweaking and reevaluation. This calls for alternative ranking functions, which is the motivation behind this chapter.

Hence, this chapter is concerned with formulating alternative order of magnitude abstractions of probabilistic systems that do not fall into the same pitfalls as existing frameworks, mainly the κ framework. It begins with outlining the concerns with κ in Section 4.1 followed by the ideas behind the proposed frameworks in Sections 4.2 and 4.3. In sections 4.4, 4.5 and 4.6, two alternative ranking calculi are presented: the κ^{++} and $\mathfrak{z}$ systems. After the formulation of the proposed functions, a critical evaluation of the ranking power of the new ranking systems with respect to each other and the known κ calculus are given. The chapter ends with a brief summary in Section 4.7

4.1 Epistemic Concerns with the κ Calculus

The κ function uses unsigned integers as an indication of the disbelief associated with the occurrence of an event by assigning events that are likely to occur a rank of zero, and giving a higher rank (higher integer value) to less likely ones. These semantics are enforced in the κ calculus, not only for modeling purposes but also to ensure the soundness of the calculus by establishing deductive closure (Goldszmidt and Pearl, 1996). $\kappa(w_1 \vee \neg w_1) = 0$ formalizes such semantics for any binary variable W having two values w and $\neg w$. The following are consequences of this semantics:

1. Although κ provides a rich range of possible ranks for surprising events ($\kappa(w) \geqslant 1$), all non-surprising (expected) events are given a rank of zero. For instance, if $\kappa(w_1) = 4$ we know that w_1 is surprising and is therefore relatively disbelieved, which renders $\neg w_1$ to be more believable than not. However, the κ calculus does not provide any means to obtaining a meaningful value for $\kappa(\neg w_1)$ as it will automatically assign it a rank of zero. Moreover, this semantics does not seem to distinguish between non-surprising events and expected ones, i.e. the average from the anticipated, which is well-defined in numerical uncertainty formulations including probability theory. On the level of reasoning, this makes belief propagation difficult with such calculus and is especially noticeable if two propositions w_1 and w_2 are considered, with $\kappa(w_2) = 1$. In this case, both $\kappa(\neg w_1) = 0$ and $\kappa(\neg w_2) = 0$. Although if taking $\neg w_1$ and $\neg w_2$ individually this would be a reasonable conclusion to reach, because the fact that w_1 was labeled as surprising deems $\neg w_1$ non-surprising and hence, having a κ value of zero (with a similar reasoning drawn for w_2 and $\neg w_2$). It would seem awkward assigning both $\neg w_1$ and $\neg w_2$ the same rank and having them to be equally normal despite the fact that the κ values of w_1 and w_2 indicate that w_1 is much more surprising than w_2. This type of comparisons lead to unwarranted ignorance that a richer representation would not suffer from.

2. This awkwardness is not only on the conceptual level, but also propagates to the rules governing belief updates because beliefs and disbeliefs are not semantically comparable because disbeliefs are given a much richer semantics than beliefs. Spohn (1988a) deals with this problem by defining a *belief* function to complement the disbelief function by assigning $-\kappa(w_1)$ as the belief rank of $\neg w_1$ if $\kappa(w_1) > 0$. This, however, takes the value of the belief of $\neg w_1$ outside the range of the ranking function, and therefore, incorporating it into the reasoning system entails the use of mechanisms outside the conditional propagation, which is the main source of the power of the calculus. Because of this, Goldszmidt and Pearl (1996) did not incorporate the belief function in their version of the κ calculus.

3. A direct consequence of the fact that κ abstracts the inverse of the probability is that this order-of-magnitude measure suffers from the same problems as the numerical inverse. More specifically, taking the inverse of the probability of an

event as an indication of how 'unlikely' (Tawfik, 1997) it is does not make a good indication because, for instance, one of equally-likely n events maybe very surprising (or not) depending on n. However, in most situations, such occurrence should not be surprising (Tawfik and Neufeld, 1996). In other words, the 'unlikelihood' captured by the probability inverse gives an absolute measure that does not take into account the probabilities of the other events in the distribution.

4. How 'good' the ranking offered by κ rests on the assumption that ϵ is infinitesimal (Spohn, 1988a; Goldszmidt and Pearl, 1996). As seen in the literature review, it is only when an infinitesimal ϵ is chosen that the ranking offered by κ seems to agree with the rankings provided by numerical probabilities (Darwiche and Goldzmidt, 1994). The problem is however, that when an non-infinitesimal ϵ is used, the generated value for κ becomes mainly trivial (a κ value of 0 is produced more often than desired) unless the equivalent probabilities are arbitrarily high or low (for example, above 0.95 or below 0.05), in which case meaningful κ values can be obtained. This has been demonstrated in Example 9 of Chapter 3.

   As a result, one is forced to use an non-infinitesimal ϵ in practice, which presents the continuous task of monitoring of how close ϵ is to one and studying its effect on the resulting value of κ and any discrepancies between the ranking it offers and that provided by probabilities, or how trivial it becomes.

As a result, κ is considered of a purely of an ordinal nature and not rich enough to be used for decision making (Huber, 2006).

4.2 From Disbelief to Surprise

The epistemic problems identified in κ, along with the observation that using a notion other than the inverse of probability is a better way to state how unlikely an event is motivate this work.

More specifically, this dissertation investigates the possibility of creating ranking functions that capture *surprise* instead of *disbelief*. Surprise is distinguished from

pure disbelief here in that it provides a measure of the unexpectedness of the event relative to others in the domain while pure disbelief is merely measured as the inverse of the probability. This notion of surprise is expected to eliminate at least some of the undesirable characteristics exhibited by κ listed in the previous section. Hence, the objectives of this dissertation can be stated as:

To formulate alternative ranking functions that are based on the concept of abstracting surprise, and to study their capabilities in eliminating at least some of the less desirable properties of existing functions by:

1. *Providing comparable semantical richness for the three epistemic states of events of being 1) believed 2) disbelieved 3) neither believed or disbelieved.*
2. *Eliminating the mere ordinal nature of the ranking provided by κ and have a function whose ranks are semantically indicative of the relative strength or weakness with respect to other events.*
3. *Assigning more meaningful ranks and improving the process of abstraction by not being as sensitive to the value of auxiliary variables used to formulate the ranking function (e.g. ϵ).*
4. *Offering a better ranking and minimizing the discrepancies between probability rankings and their own.*

4.3 Approach

The approach to be followed is that of investigating the creation of qualitative ranking functions that are based on abstracting surprise measures instead of probability inverse and examining the characteristics exhibited by the resulting ranking functions theoretically and empirically. For each of the measures, the following procedures are followed:

1. Derive the qualitative function.

2. Extract the semantics embedded within the function.
3. Devise conditioning-based surprise propagation rules.
4. Examine the advantages and limitations induced by the function.
5. Empirically examine the features exhibited by the function.

4.4 Surprise Measures

Upon investigating several forms of surprise measures, two have prevailed: the Weaver index (Weaver, 1948) and the relative predictive surprise measure (RPS) (Bayarri and Morales, 2003; Bayarri and Berger, 1998). These will be the bases for the functions devised in this dissertation and are introduced below.

4.4.1 The Weaver Index

This index defines surprise as the ratio of the expected value of the probability to the probability of the event that actually occurs. This notion presents a measure that takes into account the distribution to which the event belongs and is termed the Weaver surprise index (Weaver, 1948) denoted by $\mathcal{W}(w)$ for an event w and is given below.

$$\mathcal{W}(w) = \frac{\sum_{i=1}^{I} p(w_i)^2}{p(w)} \tag{4.1}$$

The values given by $\mathcal{W}(w)$ range from zero to infinity, with a value between zero and one corresponding to a likely outcome (i.e. no surprise) while values greater than one indicate a surprise, with the larger the index the more astonishing the event is (Tawfik and Neufeld, 1996). As a result, the Weaver index offers a range of rankings for both surprising and non-surprising events.

4.4.2 The Relative Predictive Surprise Measure

This index is tailored to measure whether or not an observed event is unusual compared to the most likely value (Bayarri and Morales, 2003) and is calculated by:

$$\mathcal{RPS} = \frac{p(w)}{p(w_{max})} \tag{4.2}$$

Where w_{max} is the probability of the most likely outcome for W.

Small values ($\mathcal{RPS} < 1$) indicate a surprise while large values ($\mathcal{RPS} > 1$) indicate an anticipated value.

4.5 Abstracting the Weaver Index: the κ^{++} Ranking Function

Given a universe of events Ω, with each event $\in \Omega$ modeled by some variable W having a well-defined set of values, we define κ^{++}, a function that captures the order of magnitude abstraction of the relative numerical surprise $\mathcal{W}$ associated with W having some value w, written as $\kappa^{++}(w)$.

For a variable W with I possible values, $\kappa^{++}(W = w)$ is defined as the lowest integer k such that $\lim_{\epsilon \to 0} \mathcal{W}(w)/\epsilon^k$ is nonzero, where ϵ is a small positive number less than one. This makes $\kappa^{++}(w) = k$ of the same order of magnitude as $\mathcal{W}(w)$, where $\mathcal{W}(w)$ is the Weaver index of (Weaver, 1948) given in Equation 4.1.

As a result, an order of magnitude abstraction of $\mathcal{W}(w)$, namely $\kappa^{++}(w)$, constrains $\mathcal{W}(w)$ as follows:

$$\epsilon < \frac{\mathcal{W}(w)}{\epsilon^{\kappa^{++}(w)}} \leqslant 1 \tag{4.3}$$

4.5.1 Deriving κ^{++}

The derivation of κ^{++} begins by presenting all the probabilities of the variables by polynomials of ϵ's. In this sense, let χ_w^n be the polynomial representation of $p(w)$, and for every other value w_i of w, let $\chi_{w_i}^{\beta_i}$ denote the polynomial corresponding to $p(w_i)$, with n and β_i being the minimum powers of ϵ in the polynomials respectively. According to Equation (4.1), $\mathcal{W}(w)$ becomes:

$$\mathcal{W}(w) = \frac{\sum_{i=1}^{I} p(w_i)^2}{p(w)} = \frac{\sum_{i=1}^{I} (\chi_{w_i}^{\beta_i})^2}{\chi_w^n}$$

Since all the polynomials are to the base ϵ, it is possible to add the terms that have equal exponents. This makes the above summation:

$$\mathcal{W}(w) = \frac{\alpha_1\epsilon^{2\beta_1} + ... + \alpha_I\epsilon^{2\beta_I} + \alpha_{I+1}\epsilon^{2\phi_1} + ... + \alpha_l\epsilon^{2\phi_k}}{\chi_w^n}$$

$\forall\beta_i, 1 \leqslant i \leqslant I, \alpha_i\epsilon^{2\beta_i}$ is a term whose power is a candidate to be the minimum power of the polynomial representing $\sum_{i=1}^{I}(\chi_{w_i}^{\beta_i})$ (i.e. the most significant term) as each $2\beta_i$ is the minimum power of $(\chi_{w_i}^{\beta_i})^2$. The ϕ terms in the equation above are non-minimum terms and therefore, their number $(l-(I+1))$ and values are irrelevant for our purpose.

Let m be such term, i.e. $m = \beta_i$ is the minimum of the minimum powers of the polynomials $2\beta_i$. $\mathcal{W}(w)$ can now be represented only in terms of polynomials as:

$$\mathcal{W}(w) = \frac{\chi_{w_i}^{2m}}{\chi_w^n}$$

According to Equation (4.3), $\kappa^{++}(w) = \lim_{\epsilon\to 0}\mathcal{W}(w)$, which implies that:

$$\kappa^{++}(w) = 2m - n \tag{4.4}$$

Where m is the minimum of all minimum powers in the polynomial $p(w_i)$, $1 \leqslant i \leqslant I$, and n is the minimum power in $p(w)$.

Corollary 1. $\kappa^{++}(w) \in \mathbb{Z} \cup \{+\infty\}$ *for all values* w *of a variable* W*.*

Proof.

- **Case 1:** $Pr(w) = 0$*. In this case:*

  $\mathcal{W}(w) = +\infty$*, which implies that* $\kappa^{++}(w) = \lim_{\epsilon \to 0} \mathcal{W}(w) = +\infty$

- **Case 2:** w *is such that* $Pr(w) \neq 0$*. In this case* $\kappa^{++}(w) = 2m - n$ *according to Equation 4.4. Since both* m *and* $n \in \mathbb{N}$*,* $2m - n \in \mathbb{Z}$*.*

□

The ranking function is called κ^{++} as it is based on a concept similar to that of κ but differs in that it abstracts surprise rather than the inverse of probability in order to allow the explicit modeling of the degree of incremental surprise of both variables and their complements and incorporate complements in conditional belief propagation. κ^{++} explicitly models relative surprise, which is the notion we believe is necessary for a richer semantics for the order-of-magnitude abstraction. We describe κ^{++} as a *general ranking function* to distinguish it from the regular *ordinal ranking* functions proposed so far in (Shenoy, 1990; Goldszmidt and Pearl, 1996; Huber, 2006) as its semantics go beyond an ordinal description of suprrise.

4.5.2 Semantics of κ^{++}

In a problem domain defined the universe of events Ω, let $\mathbf{W}$ be a non-empty subset of Ω and let $\mathbf{w}$ be the set of all possible values of the variables in $\mathbf{W}$. The elements of $\mathbf{w}$ are called the configurations of $\mathbf{W}$.

The general ranking function κ^{++} for a set of variables $\mathbf{W}$ is defined as a mapping from the set of configurations $\mathbf{w}$ of $\mathbf{W}$ to the set of signed integers extended by infinity $\kappa^{++} : 2^{\mathbf{w}} \to \mathbb{Z} \cup \{+\infty\}$ as a function that assigns an integer value such that this value is representative of the epistemic state of $\mathbf{W}$ as:

Definition 16. General Form of κ^{++}

$$\kappa^{++}(\boldsymbol{W}) = \begin{cases} +\infty & \text{, iff } \boldsymbol{W} = \phi; \\ \min\limits_{\boldsymbol{w} \in \boldsymbol{W}} \{\kappa^{++}(\boldsymbol{w})\} & \text{, otherwise.} \end{cases}$$

κ^{++} provides a comprehensive definition of the epistemic state of $\mathbf{W}$ that ensures its consistency. An epistemic state of some set of variables $\mathbf{W}$ is defined by the truth assignments of its constituent variables and is said to be consistent if all of the variables it represents can be either believed, disbelieved or neither.

Interpreting κ^{++}

κ^{++} as given in Equation 4.4 implies that large (positive) values indicate a greater difference between $2m$ and n, and as a result, a greater surprise associated with the event. Similarly, smaller the (negative) value of a κ^{++}, the larger the difference between $2m$ and n (with $n > 2m$ in this case), and as a result, the more possible the event is compared to other events in the distribution.

Hence, the signed integer representation of $\kappa^{++}(.)$ enables the function of providing a continuous measure of the epistemic state that proposition falls under. The larger the value of $\kappa^{++}(w)$, the greater the difference between its constituent quantities ($2m$ and n), and as a result the more surprising the event in question, w, is. Therefore, the signed integer produced by κ^{++} carries the semantics defined by three possible classes for its value.

Positive: $(\kappa^{++}(w) = 2m - n) > 0$ implies that the event w is a lot less likely than the other events g_i $(1 \leqslant i \leqslant I)$ of the distribution, i.e. $2m > n$. Hence, the occurrence of w indicates a surprise. Moreover, the larger the value of $\kappa^{++}(w)$ (the greater the difference is between $2m$ and n), the more surprising the event w is.

Zero: $\kappa^{++}(w) = 0$ represents the normal world where both w and $\neg w$ are likely to occur as the order of magnitude of the probability of the variable w is comparable to that of the distribution, i.e. $2m = n$.

Essentially, $\kappa^{++}(w) = 0$ separates surprising events from expected ones. a variable $W \in \Omega$ is believed if its κ^{++} value falls under 0, is disbelieved if its κ^{++} value falls

above 0 and is neither believed or disbelieved if its κ^{++} value is zero. **Negative:** $\kappa^{++}(w) < 0$ refers to the case in which having the event w to be false is surprising as w becomes more likely than unlikely compared to other events in the distribution (because $n > 2m$), which implies that $\neg w$ is unlikely and its $\kappa^{++}(.)$ should indicate a surprise. In this case, the smaller the value of $\kappa^{++}(.)$, the more surprising $\neg w$ is.

Corollary 2. *For any value w of W, $\forall\ W \in \Omega$:*

1. $\kappa^{++}(w) > 0 \rightarrow$ *w is surprising and hence more disbelieved than believed.*
2. $\kappa^{++}(w) < 0 \rightarrow$ *w is anticipated and hence more believed than disbelieved.*
3. $\kappa^{++}(w) = 0 \rightarrow$ *w is normal: neither surprising nor anticipated.*

Proof. By contradiction:

Let $\kappa^{++}(w) > 0$ with w not being surprising.

$$\begin{aligned}\kappa^{++}(w) > 0 &\Rightarrow 2m - n > 0 \\ &\Rightarrow 2m > n\end{aligned}$$

Which implies that w is surprising (contradiction). Therefore, w is surprising.

The same reasoning can be used to prove the cases where w is anticipated and normal. □

κ^{++} Complements

The κ^{++} value of events and their complements are related in the κ^{++} calculus. Although this relation is not as strong as the one provided by probability theory, it provides an indication to the values in concern. This relation is stated in Theorem 1 below.

Theorem 1. : Negation of κ^{++} Values

$\exists \alpha : \forall w \in \Omega : \kappa^{++}(w) + \kappa^{++}(\neg w) = \alpha.$

Proof.

The above rule can be more easily understood if one examines the probabilistic interpretation of κ^{++} as follows.

Let $p(w) = \epsilon^{n_1}$ and $p(\neg w) = \epsilon^{n_2}$. In this case:

$\mathcal{W}(w) = \dfrac{\epsilon^{2n_1} + \epsilon^{2n_2}}{\epsilon^{n_1}}$ and $\mathcal{W}(\neg w) = \dfrac{\epsilon^{2n_1} + \epsilon^{2n_2}}{\epsilon^{n_2}}$

When calculating $\kappa^{++}(w) + \kappa^{++}(\neg w)$, the following cases arise:

- **Case 1:** $n_1 < n_2$

  $\kappa^{++}(w) = 2n_1 - n_1 = n_1$ and $\kappa^{++}(\neg w) = 2n_1 - n_2$.

  $\Rightarrow \kappa^{++}(w) + \kappa^{++}(\neg w) = 3n_1 - n_2$:

- **Case 2:** $n_1 > n_2$

  $\kappa^{++}(w) = 2n_2 - n_1$ and $\kappa^{++}(\neg w) = n_2$:

  $\Rightarrow \kappa^{++}(w) + \kappa^{++}(\neg w) = 3n_2 - n_1$

- **Case 3:** $n_1 = n_2 = n$

  $\kappa^{++}(w) = \kappa^{++}(\neg w) = n$

  $\Rightarrow \kappa^{++}(w) + \kappa^{++}(\neg w) = 2n = 3n - n$:

In all of the above cases, $\kappa^{++}(w) + \kappa^{++}(\neg w) = 3\alpha_1 - \alpha_2 = \alpha$, where α_1 denotes the larger exponent and α_2 the lesser one.

□

The α value enables having an explicit description of the value of $\kappa^{++}(\neg w)$ based on that of $\kappa^{++}(w)$ as given in Definition 17 below.

Definition 17. Complements

$$\kappa^{++}(\neg w) = \begin{cases} \min\limits_{w_i \in \Omega} \kappa^{++}(w_i) & , \textit{iff } \kappa^{++}(w) = +\infty; \\ \alpha - \kappa^{++}(w) & , \textit{otherwise.} \end{cases}$$

4.5.3 Reasoning with κ^{++}

Because the $\kappa^{++}(.)$ can be mapped to the order of magnitude abstraction of a probabilistic measure, the operations used for propagating its values can be defined by replacing numerical operators by order of magnitude equivalents as done in the κ calculus. Hence multiplication is replaced by addition and addition by minimum (Goldszmidt and Pearl, 1996). As a result, conditioning can be performed with the aid of the following rules:

Rule Set 1. : κ^{++} Propagation Rules

R1: $\kappa^{++}(w_1 \wedge w_2) = \kappa^{++}(w_1) + \kappa^{++}(w_2)$ *If* w_1 *and* w_2 *are independent variables.*

Theorem 2. (Conjunction) *Given* N *independent variables* $W_1, ..., W_N$ *drawn from some distribution* ζ*, the probability of the conjunction of the* N *variables is equivalent to the sum of the degrees of surprise,* κ^{++}*, associated with the variables.*

Proof.

We aim at demonstrating that the process of converting the probability of the conjunction of the variables $W_1, ..., W_N$ into a κ^{++} value replaces the product resulting from the conjunction by a summation. For this purpose, we define a mapping $\xi : Pr \rightarrow \kappa^{++}$ between the probability values and their κ^{++} equivalents. Hence, the task is to prove that:

$$\xi[Pr(\bigwedge_{c=1}^{N} W_c)] = \sum_{c=1}^{N} \kappa^{++}(W_c)$$

Since the N are independent, the probability of their conjunction is equivalent to the product of their corresponding probabilities as given below:

$$Pr(\bigwedge_{c=1}^{N} W_c) = \prod_{c=1}^{N} Pr(W_c)$$

As a result, the problem reduces to finding the mapping between the product of the probabilities of the variables to the sum of their corresponding κ^{++} values.

Hence, the first step is to obtain an expression resulting form converting the product of the probabilities of the variables to the product of the numerical surprise $\mathcal{W}(.)$ associated with each variable, which are subsequently readily available due from equation 4.1. The resulting expression is given by:

$$\prod_{c=1}^{N} \mathcal{W}(W_c) = \prod_{c=1}^{N} \frac{\sum_{i=1}^{I_c} Pr(W_{c_i})^2}{Pr(W_c)} = \prod_{c=1}^{N} \frac{\chi_{W_{c_i}}^{2m_c}}{\chi_{W_c}^{n_c}}$$

Where I_c is the number of possible values each variable W_c $(1 \leqslant c \leqslant N)$ has. The next step is to obtain the expression in terms of the κ^{++}'s associated with the variables. According to Definition 4.4, the desired expression is obtained by finding the limit of the expression given above as ϵ reaches zero.

$$\begin{aligned}
\lim_{\epsilon \to 0} \prod_{c=1}^{N} \mathcal{W}(W_c) &= \sum_{c=1}^{N} \lim_{\epsilon \to 0} \frac{\chi_{W_{c_i}}^{2m_c}}{\chi_{W_c}^{n_c}} \\
&= \sum_{c=1}^{N} 2m_c - n_c \\
&= \sum_{c=1}^{N} \kappa^{++}(W_c)
\end{aligned}$$

□

R2: $\kappa^{++}(w_1 \vee w_2) = \min(\kappa^{++}(g), \kappa^{++}(\neg g))$

Proof.

The proof to the rule **R2** is given in the theorem below. □

Theorem 3. (Disjunction) *Given N independent and mutually exclusive variables $W_1, ..., W_N$ drawn from some distribution ζ, the probability of the disjunction of the variables equivalent to obtaining the minimum of degree of surprise, κ^{++}, associated with the variables.*

Proof.

We aim at demonstrating that the process of converting the probability of the disjunction of the variables $W_1, ..., W_N$ into a κ^{++} value replaces the summation resulting from the disjunction by a minimum. For this purpose, we define a mapping $\xi : Pr \rightarrow \kappa^{++}$ between the probability values and their κ^{++} equivalents. Hence, the task is to prove that:

$$\xi[Pr(\bigvee_{c=1}^{N} W_c)] = \min_{c=1}^{N} \kappa^{++}(W_c)$$

Since the N are independent, the probability of their disjunction is equivalent to the summation of their corresponding probabilities as given below:

$$Pr(\bigvee_{c=1}^{N} W_c) = \sum_{c=1}^{N} Pr(W_c)$$

As a result, the problem reduces to finding the mapping between the sum of the probabilities of the variables to the corresponding minimum κ^{++} value.

As done in Theorem 2, we begin the process of conversion by obtaining a corresponding expression for the $\mathcal{W}$ values of the N variables as given below.

$$\sum_{c=1}^{N} (W_c) = \sum_{c=1}^{N} \frac{\sum_{i=1}^{I_c} Pr(W_{c_i})^2}{Pr(W_c)} = \sum_{c=1}^{N} \frac{\chi_{W_{c_i}}^{2m_c}}{\chi_{W_c}^{n_c}}$$

The corresponding κ^{++} value of the expression is that of the logarithm of the summation, which is given by:

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \prod_{c=1}^{N} \mathcal{W}(W_c) &= \min_{c=1}^{N} \lim_{\epsilon \rightarrow 0} \frac{\chi_{W_{c_i}}^{2m_c}}{\chi_{W_c}^{n_c}} \\ &= \min_{c=1}^{N} 2m_c - n_c \\ &= \min_{c=1}^{N} \kappa^{++}(W_c) \end{aligned}$$

□

R3: $\kappa^{++}(w_1|w_2) = \kappa^{++}(w_1 \wedge w_2) - \kappa^{++}(w_2)$

Proof.

We aim at demonstrating that the process of converting the conditional probability of W_1 given W_2 into a κ^{++} results in the desired expression. For this purpose, a mapping similar to the ones of Theorems 2 and 3 is defined, $\xi : Pr \rightarrow \kappa^{++}$. As a result, the task becomes proving the following:

$$\xi[Pr(w_1|w_2)] = \kappa^{++}(w_1 \wedge w_2) - \kappa^{++}(w_2)$$

As done before, the process of conversion begins by obtaining the corresponding $\mathcal{W}$ index for the conditional probability $Pr(w_1|w_2)$.

$$\begin{aligned}
\mathcal{W}(w_1|w_2) &= \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} Pr(w_i|w_j)^2}{Pr(w_1|w_2)} \\
&= \frac{\sum\sum Pr(w_i \wedge w_j)^2 / \sum Pr(w_j)^2}{Pr(w_1 \wedge w_2)/Pr(w_2)} \\
&= \frac{\sum\sum Pr(w_i \wedge w_j)^2}{\sum Pr(w_j)^2} \times \frac{Pr(w_2)}{Pr(w_1 \wedge w_2)} \\
&= \frac{\mathcal{W}(w_1 \wedge w_2)}{\mathcal{W}(w_2)}
\end{aligned}$$

As before, the κ^{++} equivalent of the conditional probability can be obtained finding the limit of the above expression as ϵ approaches zero.

$$\begin{aligned}
\kappa^{++}(w_1|w_2) &= \lim_{\epsilon \rightarrow 0} \frac{\mathcal{W}(w_1 \wedge w_2)}{\mathcal{W}(w_2)} \\
&= \lim_{\epsilon \rightarrow 0} \mathcal{W}(w_1 \wedge w_2) - \lim_{\epsilon \rightarrow 0} \mathcal{W}(w_2) \\
&= \kappa^{++}(w_1 \wedge w_2) - \kappa^{++}(w_2)
\end{aligned}$$

□

Example 10. *To illustrate reasoning with $\kappa^{++}(.)$, we constructed the network given in Figure 4.1 (inspired by one given in* (D'Ambrosio, 1999)*) and which describes the elements involved in sneezing during Spring. The network advises that the events of*

the patient having cold (O) *and that of the patient having an allergic reaction* (R) *both increase the chances that she is sneezing* (S). *Moreover, the knowledge that there is a cat in the house* (C) *effects the likelihood of* (R). *The figure also contains tables illustrating the* κ^{++} *values associated with each node being true.*

Using the surprise-propagation rules of $\kappa^{++}(.)$, *we are able explain away the symptoms by answering queries such as "given that the patient is sneezing, how surprising is it that she is suffering from a cold?", which is achieved through obtaining a value for* $\kappa^{++}(\mathrm{O}|\mathrm{S})$ *using the conditional propagation rule* **R3** *to compute the conditional surprise as shown below.*

$$\begin{aligned}
\kappa^{++}(O|S) &= \kappa^{++}(\mathrm{O} \wedge \mathrm{S}) - \kappa^{++}(\mathrm{S}) \\
&= \kappa^{++}(\mathrm{S}|\mathrm{O}) + \kappa^{++}(\mathrm{O}) - \kappa^{++}(\mathrm{S}) \\
&= \min[\kappa^{++}(\mathrm{S}|\mathrm{O},\mathrm{R}), \kappa^{++}(\mathrm{S}|\mathrm{O},\neg\mathrm{R})] + \kappa^{++}(\mathrm{O}) - \kappa^{++}(\mathrm{S}) \\
&= \min[\kappa^{++}(\mathrm{S}|\mathrm{O},\mathrm{R}), \kappa^{++}(\mathrm{S}|\mathrm{O},\neg\mathrm{R})] + \kappa^{++}(\mathrm{O}) - \\
&\quad \min[\kappa^{++}(\mathrm{S}|\mathrm{O},\mathrm{R}), \kappa^{++}(\mathrm{S}|\mathrm{O},\neg\mathrm{R}), \kappa^{++}(\mathrm{S}|\neg\mathrm{O},\mathrm{R}), \kappa^{++}(\mathrm{S}|\neg\mathrm{O},\neg\mathrm{R})] \\
&= \min[-6,-5] + 3 - \min[-6,-5,-3,+5] \\
&= +3
\end{aligned}$$

It is important to note that the result $\kappa^{++}(\mathrm{O}|\mathrm{S}) = +3$, *which indicates a surprise, is compatible with the probability* $Pr(\mathrm{O}|\mathrm{S}) = 0.0325$ *in terms of magnitude and sign.*

4.5.4 Benefits of κ^{++}

The previous sections introduced κ^{++} as a function that captures the notion of surprise in a qualitative manner. The framework developed has several characteristics which define semantics that eliminate some of the problems that κ suffers from. In this section, we summarize how κ^{++} enables solving these problems.

Belief from Disbelief

The relation between $\kappa^{++}(w)$ and $\kappa^{++}(\neg w)$ provides a direct way of assigning different degrees of expectedness to complements of surprising events. This is in contrast to

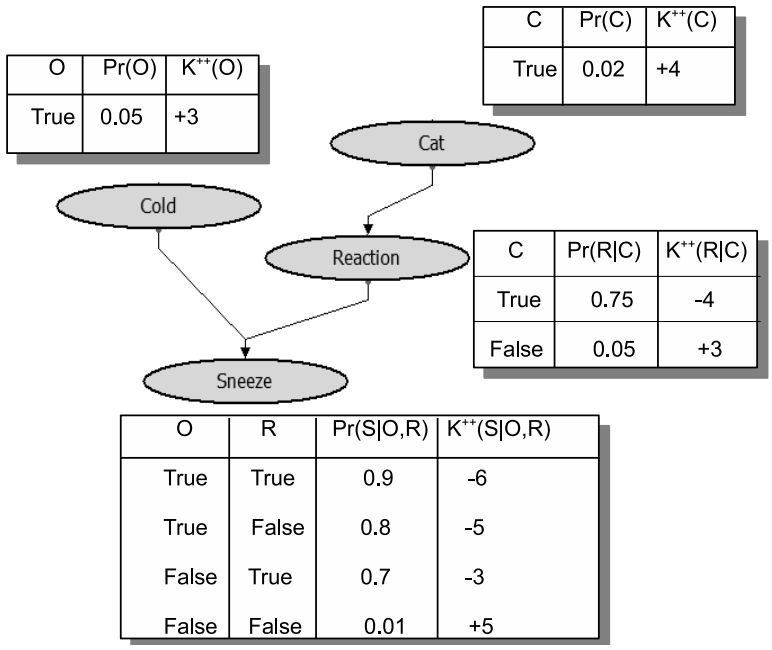


Figure 4.1: κ^{++} Network Representing Patients Information

the κ function where $\kappa(\neg w)$ is always given a value of 0 given that $\kappa(w) \neq +\infty$ which makes it unable to measure the degree of belief. In other words, Theorem 1 enables modeling opposite states of belief numerically and providing meaningful ranks for events and their negations, which is not possible in the κ calculus. Hence, although $\kappa^{++}(w)$ provides a measure of the degree of disbelief in the occurrence of w as the higher value of $\kappa^{++}(w)$ is, the less believable w becomes, it can also be used to obtain an idea regarding the degree to which $\neg w$ is believed.

Also, κ^{++} can be used to compare the *believability* of two propositions as given in Example 11.

Example 11. *Consider two variables* $W_1, W_2 \in \Omega$. *Let* $\kappa^{++}(w_1) = +2$ *and* $\kappa^{++}(w_2) = +5$, *and let* $\alpha = 0$ *for* W_1 *and* $\alpha = 1$ *for* W_2. *Given this information, it is possible to induce that* $\neg w_2$ *is* more believable *than* $\neg w_1$ *by noting that the function ranks* $\neg w_1$ *and* $\neg w_2$ *by finding* $\kappa^{++}(\neg w_1) = \alpha - 2 = -2$ *and* $\kappa^{++}(\neg w_2) = \alpha - 5 = -4$. *Notice that the* κ *calculus would have assigned a value of zero for both* $\kappa(\neg w_1)$ *and* $\kappa(\neg w_2)$.

The advantage of the above is being able to propagate $\kappa^{++}(\neg w_1)$ and $\kappa^{++}(\neg w_2)$ in addition to $\kappa^{++}(w_1)$ and $\kappa^{++}(w_2)$.

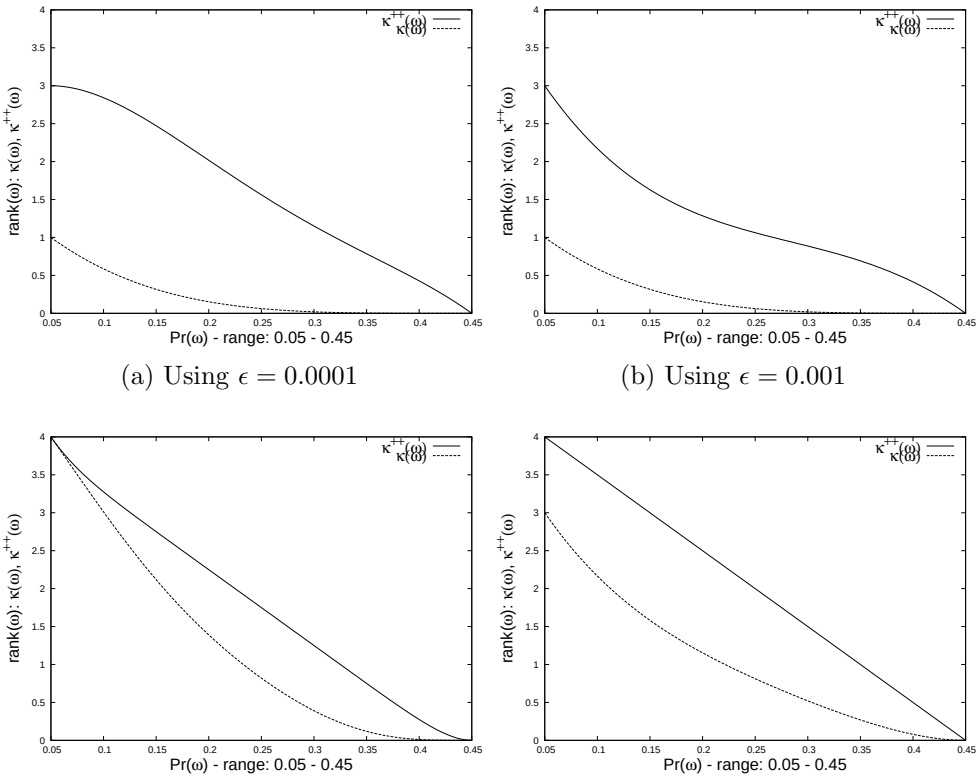
Better Ranking

As the literature review has demonstrated, κ assumes an infinitesimal ϵ for its abstraction. When a non-infinitesimal ϵ is used, the rankings given by κ tend to deviate from those given by probability. Moreover, an infinitesimal κ gives trivial rankings and is therefore not usable.

We investigated how good the ranking provided by κ^{++} compared that provided by κ with respect to the ϵ. The results are shown in Figure 4.2. As the figure shows, decreasing the value of ϵ causes the rank provided by κ to quickly degenerate to the trivial case (zero), while the slope of the conversion to zero is less in the case of κ^{++}. Moreover, the larger the value of ϵ, the less meaningful the values of κ compared to those of κ^{++} as they are closer to the trivial case (zero). As a result, κ^{++} provides a more useful and less trivial ranking than κ.

Domain Independence

One important feature of the κ^{++} calculus is that the number it assigns is computed while taking into account the expected value of probability of the distribution (as it is part of computing the Weaver index). As a result, the rank assigned to the event is not designated with respect to the absolute 1 but instead with respect to the average probability expected for the domain. This entails that the κ^{++} rank assigned to the event takes into account the relative distribution of surprise and non-surprise in the domain, which makes its value of use other than in the ordinal sense if looked at from outside the domain. This is obviously not the case for κ whose probabilistic interpretation does not take into account the way the probabilities are distributed among the events under consideration, causing its values to be of a mere ordinal nature.



(a) Using $\epsilon = 0.0001$ (b) Using $\epsilon = 0.001$

(c) Using $\epsilon = 0.01$ (d) Using $\epsilon = 0.1$

Figure 4.2: κ^{++}'s sensitivity to ϵ values compared to κ using different values of ϵ. The Comparison is made for extreme and non-extreme values of probability ranging between 0.05 and 0.45. The range is chosen so that only positive κ^{++} values are generated in order to compare with the non-signed κ values. The curves are Bezier interpolation of discrete values performed to demonstrate the difference between the values of κ and κ^{++} in terms of 1) the slope of the curve, which indicate the speed of degeneration to the trivial (zero) value 2) the range assumed by the function.

4.5.5 Limitations of the κ^{++} Calculus

Although in general, κ^{++} enables obtaining an idea regarding the degree of expectedness of $\kappa^{++}(\neg w)$ from $\kappa^{++}(w)$, obtaining the exact value of the ranking of the complement is not a straightforward process as it depends on the value of a third variable α as Theorem 1 shows. More specifically, the value of α is the same for a

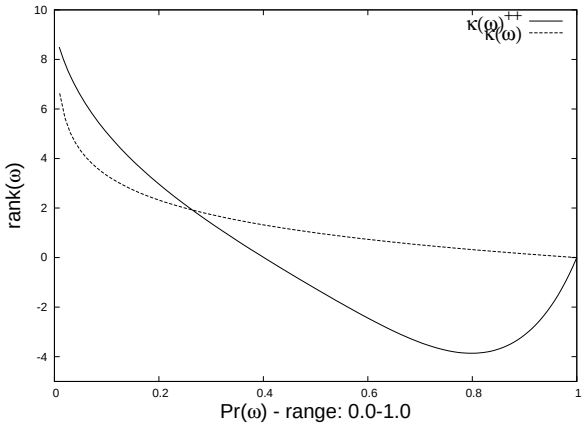


Figure 4.3: Problems in κ^{++}: 1) Range Bias: Positive κ^{++} values approach +8 while -4 is the minimum possible negative κ^{++} 2) Non-monotonic Section: From $Pr(w) = 0.84$ to $Pr(w) = 1$, $\kappa^{++}(w)$ is increasing.

universe Ω of events but is unique to each pair of events as the proof of the theorem indicates. Moreover, obtaining the value of α for each pair of events requires details that mandate referring to the numerical probabilities associated with the events. Needless to say, this associated cost represents hurdles that can render the added semantical richness offered by κ^{++} less attractive.

Moreover, although κ^{++} provides a richer semantics than κ, still does not provide equally-rich semantics for surprising and expected events. This problem is illustrated in Figure 4.3. In the figure, one can see that the range offered for surprising events (the positive portion of the κ^{++} values) is much wider than that offered for expected ones (the negative portion of the κ^{++} values). Added to this is the non-monotonic behavior exhibited by the function for a small (yet existing) portion of its negative range which can also be noted in the figure. In this portion (from $Pr(w) = 0.84$ to $Pr(w) = 1$), the κ^{++} rankings are consistently disagreeing with probabilistic rankings. Although this can be avoided algorithmically, it represents an inherent problem in the semantics of κ^{++}.

4.6 Abstracting the Relative Predictive Surprise Index: System $\mathfrak{z}$

The findings with respect to the limitations of the κ^{++} calculus created a motivation to investigate alternative surprise indices as bases for ranking functions. This led to the inception of the $\mathfrak{z}$ (read: zed) calculus presented in this section and which aims at achieving some of the advantages that κ^{++} has over κ without the added disadvantages.

Given a universe of events Ω, with each event modeled by a variable $W \in \Omega$ having a well-defined set of values, we define $\mathfrak{z}$, a function that captures the order of magnitude abstraction of the $\mathcal{RPS}$ measure given in Equation 4.2 associated with W having some value w, written as $\mathfrak{z}(w)$ and given in equation 4.5 below.

Definition 18.

$$\epsilon < \frac{\mathcal{RPS}(w)}{\epsilon^{\mathfrak{z}(w)}} \leqslant 1 \tag{4.5}$$

Which defines $\mathfrak{z}(W = w)$ as the lowest integer k such that $\lim_{\epsilon \to 0} \mathcal{RPS}(w)/\epsilon^k$ is nonzero, where ϵ is a small positive number less than one. This makes $\mathfrak{z}(w) = k$ of the same order of magnitude as $\mathcal{RPS}(w)$ given in Equation 4.2.

4.6.1 Deriving $\mathfrak{z}$

As done with κ^{++}, finding the range to which the values of $\mathfrak{z}$ belong to is done by representing all the probabilities $p(w)$ of the variables by polynomials of ϵ's ($0 < \epsilon < 1$), $p(w) = \chi_w^m$ where m is the minimum power of ϵ in the polynomial. The results obtained regarding $\mathfrak{z}$'s range are given in Theorem 4 below.

Theorem 4. $\mathfrak{z} : \Omega \to \mathbb{Z}^+ \cup 0$

Proof. According to Equation 4.2, $\mathcal{RPS}(w)$ can be written as:

$$\mathcal{RPS}(w) = \frac{p(w)}{p(w_{max})} = \frac{\chi_w^m}{\chi_{w_{max}}^n} \tag{4.6}$$

Where χ_w^m is the polynomial of ϵ representing $p(w)$ and having m as the power of the dominant ϵ term, and $\chi_{w_{max}}^n$ is the polynomial ϵ representing $p(w_{max})$ and having n as the power of the dominant term.

According to Definition 18, $\mathfrak{z}(w)$ is the order of magnitude abstraction of $\mathcal{RPS}(w)$ and can therefore be written as:

$$\begin{aligned}\mathfrak{z}(w) &= \lim_{\epsilon \to 0} \mathcal{RPS}(w) \\ &= \lim_{\epsilon \to 0} \frac{\chi_w^m}{\chi_{w_{max}}^n} \\ &= m - n\end{aligned}$$

Since n represents the minimum power of ϵ in the polynomial representing $p(w_{max})$, it is guaranteed that: $m >= n$ will always hold. As a result, $\mathfrak{z}(w)$ is guaranteed not to assume a negative value, which is further formalized in Corollary 3 below.

□

Corollary 3. $m - n \in \mathbb{Z}^+ \cup 0$

Proof.

By Contradiction:

Let $(w) \in \mathbb{Z}^-$, i.e. $m - n < 0$

$\Rightarrow m < n$

$\Rightarrow p(w) > p(w_{max})$

Which is impossible because $p(w_{max})$ defines the maximum probability associated with any value of W. □

4.6.2 $\mathfrak{z}$ Semantics

The findings of the previous section entail that the difference between $\mathfrak{z}$'s two components give the ranking function the following semantics:

Positive: $(\mathfrak{z}(w) = m - n) > 0$ implies that the event w is less likely than the most likely event w_{max} of the distribution, i.e. $m > n$. Hence, the occurrence of w indicates a surprise. Moreover, the larger the value of $\mathfrak{z}(w)$ (the greater the difference is between m and n), the more surprising the event w is.

Zero: $\mathfrak{z}(w) = 0$ indicates no surprise. This happens when $m = n$ or $p(w) = p(w_{max})$.

4.6.3 Reasoning with $\mathfrak{z}$

As done with κ^{++}, this section will show that propagation of $\mathfrak{z}$ values is done using rules similar to those of probability by replacing multiplication by addition and addition by minimum. This will be illustrated by deriving the rules of conjunction, disjunction and conditionality for $\mathfrak{z}$ below.

Rule Set 2. : $\mathfrak{z}$ Combination Rules

R4: $\mathfrak{z}(w_1 \wedge w_2) = \mathfrak{z}(w_1) + \mathfrak{z}(w_2)$ *If* w_1 *and* w_2 *are independent variables.*

R5: $\mathfrak{z}(w_1 \vee w_2) = \min(\mathfrak{z}(w), \mathfrak{z}(\neg w))$

Proof. The proof to rules **R4** and **R5** are given by Theorems 5 and 6 respectively. □

Theorem 5. ($\mathfrak{z}$ Conjunction) *Given* N *independent variables* $W_1, ..., W_N$ *drawn from some distribution* ζ*, the probability of the conjunction of the* N *variables is equivalent to the sum of the degrees of surprise,* $\mathfrak{z}$*, associated with the variables.*

Proof.

As with κ^{++}, we aim at demonstrating that the process of converting the probability of the conjunction of the variables $W_1, ..., W_N$ into a $\mathfrak{z}$ value replaces the product

resulting from the conjunction by a summation. For this purpose, we define a mapping $\xi : Pr \rightarrow \mathfrak{z}$ between the probability values and their $\mathfrak{z}$ equivalents. Hence, the task is to prove that:

$$\xi[Pr(\bigwedge_{c=1}^{N} W_c)] = \sum_{c=1}^{N} \mathfrak{z}(W_c)$$

Since the N are independent, the probability of their conjunction is equivalent to the product of their corresponding probabilities, which reduces the problem to finding the mapping between the product of the probabilities of the variables to the sum of their corresponding $\mathfrak{z}$ values.

This, however, is readily available through Equation 4.6 and is find as:

$$\prod_{c=1}^{N} \mathcal{RPS}(W_c) = \prod_{c=1}^{N} \frac{Pr(W_c)}{Pr(W_{max})} = \prod_{c=1}^{N} \frac{\chi_{W_c}^{m_c}}{\chi_{W_{max}}^{n}}$$

The next step is to obtain the expression in terms of the $\mathfrak{z}$'s associated with the variables. According to Definition 18, the desired expression is obtained by finding the limit of the expression given above as ϵ approaches zero, or:

$$\begin{aligned}\lim_{\epsilon \rightarrow 0} \prod_{c=1}^{N} \mathcal{RPS}(W_c) &= \sum_{c=1}^{N} \lim_{\epsilon \rightarrow 0} \frac{\chi_{W_c}^{m_c}}{\chi_{W_{max}}^{n}} \\ &= \sum_{c=1}^{N} m_c - n \\ &= \sum_{c=1}^{N} \mathfrak{z}(W_c)\end{aligned}$$

□

Theorem 6. ($\mathfrak{z}$ Disjunction) *Given N independent and mutually exclusive variables $W_1, ..., W_N$ drawn from some distribution ζ, the probability of the disjunction of the variables equivalent to obtaining the minimum of degree of surprise, $\mathfrak{z}$, associated with the variables.*

Proof.

We aim at demonstrating that the process of converting the probability of the disjunction of the variables $W_1, ..., W_N$ into a $\mathfrak{z}$ value replaces the summation resulting from the disjunction by a minimum. For this purpose, we define a mapping $\xi : Pr \rightarrow \mathfrak{z}$ between the probability values and their $\mathfrak{z}$ equivalents. Hence, the task is to prove that:

$$\xi[Pr(\bigvee_{c=1}^{N} W_c)] = \min_{c=1}^{N} \mathfrak{z}(W_c)$$

Since the N are independent, the probability of their disjunction is equivalent to the summation of their corresponding probabilities as given below:

$$Pr(\bigvee_{c=1}^{N} W_c) = \sum_{c=1}^{N} Pr(W_c)$$

As a result, the problem reduces to finding the mapping between the sum of the probabilities of the variables to the corresponding minimum $\mathfrak{z}$ value.

As done in Theorem 5, we begin the process of conversion by obtaining a corresponding expression for the $\mathcal{W}$ values of the N variables as given below.

$$\sum_{c=1}^{N} (W_c) = \sum_{c=1}^{N} \frac{Pr(W_c)}{Pr(W_{max})} = \sum_{c=1}^{N} \frac{\chi_{W_c}^{m_c}}{\chi_{W_{max}}^{n}}$$

The corresponding $\mathfrak{z}$ value of the expression is that of the logarithm of the summation, which is given by:

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \prod_{c=1}^{N} \mathcal{RPS}(W_c) &= \min_{c=1}^{N} \lim_{\epsilon \rightarrow 0} \frac{\chi_{W_c}^{m_c}}{\chi_{W_{max}}^{n}} \\ &= \min_{c=1}^{N} m_c - n \\ &= \min_{c=1}^{N} \mathfrak{z}(W_c) \end{aligned}$$

□

Rule Set 3. : $\mathfrak{z}$ Propagation Rule

R6: $\mathfrak{z}(w_1|w_2) = \mathfrak{z}(w_1 \wedge w_2) - \mathfrak{z}(w_2)$

Proof.

We aim at demonstrating that the process of converting the conditional probability of W_1 given W_2 into a $\mathfrak{z}$ results in the desired expression. For this purpose, a mapping similar to the ones of Theorems 5 and 6 is defined, $\xi : Pr \rightarrow \mathfrak{z}$. As a result, the task becomes proving the following:

$$\xi[Pr(w_1|w_2)] = \mathfrak{z}(w_1 \wedge w_2) - \mathfrak{z}(w_2)$$

As done before, the process of conversion begins by obtaining the corresponding $\mathcal{W}$ index for the conditional probability $Pr(w_1|w_2)$.

$$\begin{aligned}\mathcal{RPS}(w_1|w_2) &= \frac{Pr(w_1|w_2)}{Pr(w_{1max}|w_{2max})} \\ &= \frac{Pr(w_1 \wedge w_2)/Pr(w_2)}{Pr(w_{1max} \wedge w_{2max})/Pr(w_{2max})} \\ &= \frac{\mathcal{RPS}(w_1 \wedge w_2)}{\mathcal{RPS}(w_2)}\end{aligned}$$

As before, the $\mathfrak{z}$ equivalent of the conditional probability can be obtained by taking the limit of the above expression as ϵ approaches zero.

$$\begin{aligned}\mathfrak{z}(w_1|w_2) &= \lim_{\epsilon \rightarrow 0} \frac{\mathcal{RPS}(w_1 \wedge w_2)}{\mathcal{RPS}(w_2)} \\ &= \lim_{\epsilon \rightarrow 0} \mathcal{RPS}(w_1 \wedge w_2) - \lim_{\epsilon \rightarrow 0} \mathcal{RPS}(w_2) \\ &= \mathfrak{z}(w_1 \wedge w_2) - \mathfrak{z}(w_2)\end{aligned}$$

□

Example 12. *Figure 4.4 shows the $\mathfrak{z}$ calculus equivalent of the sneezing network inspired by* (D'Ambrosio, 1999)*. The tables provided show the probabilities along with the $\mathfrak{z}$ values associated with them.*

The query chosen for this example is "given that the patient is sneezing, how surprising is it that she is suffering from an allergy?". Finding an answer is achieved though obtaining a value for $\mathfrak{z}(\mathrm{R}|\mathrm{S})$ *using the conditional propagation rule* **R6** *to compute the conditional surprise as shown below.*

$$\begin{aligned}
\mathfrak{z}(R|S) &= \mathfrak{z}(\mathrm{R} \wedge \mathrm{S}) - \mathfrak{z}(\mathrm{S}) \\
&= \mathfrak{z}(\mathrm{S}|\mathrm{R}) + \mathfrak{z}(\mathrm{R}) - \mathfrak{z}(\mathrm{S}) \\
&= \min[\mathfrak{z}(\mathrm{S}|\mathrm{R},\mathrm{O}), \mathfrak{z}(\mathrm{S}|\mathrm{R},\neg\mathrm{O})] + \min[\mathfrak{z}(\mathrm{R}|\mathrm{C}), \mathfrak{z}(\mathrm{R}|\neg\mathrm{C})] - \mathfrak{z}(\mathrm{S}) \\
&= \min[\mathfrak{z}(\mathrm{S}|\mathrm{R},\mathrm{O}), \mathfrak{z}(\mathrm{S}|\mathrm{R},\neg\mathrm{O})] + \min[\mathfrak{z}(\mathrm{R}|\mathrm{C}), \mathfrak{z}(\mathrm{R}|\neg\mathrm{C})] - \\
&\quad \min[\mathfrak{z}(\mathrm{S}|\mathrm{O},\mathrm{R}), \mathfrak{z}(\mathrm{S}|\mathrm{O},\neg\mathrm{R}), \mathfrak{z}(\mathrm{S}|\neg\mathrm{O},\mathrm{R}), \mathfrak{z}(\mathrm{S}|\neg\mathrm{O},\neg\mathrm{R})] \\
&= \min[0,3] + \min[4,8] - \min[0,1,3,10] \\
&= 4
\end{aligned}$$

It is important to note that the result $\mathfrak{z}(\mathrm{R}|\mathrm{S}) = 4$ *is compatible with the probability* $Pr(\mathrm{R}|\mathrm{S}) = 0.53$.

4.6.4 Features of the $\mathfrak{z}$ System

Although $\mathfrak{z}$ does not incorporate the richer range of values that κ^{++} enjoys, it does possess a number of interesting features that render it favorable as a ranking function. The fact that the rank assigned by $\mathfrak{z}$ takes into account the maximum probability value available entails that the rank assigned is done so with respect to the relative distribution of surprise and non-surprise in the domain. This makes the values of $\mathfrak{z}$, similarly to the ranks of κ^{++}, domain independent and useful in ways that are greater than in a purely ordinal sense (which is the case of κ). This idea can also be utilized for comparing events belonging to different distributions by examining their $\mathfrak{z}$ rankings. Example 13 below further illustrates these ideas.

Example 13. *Given two universes* Ω_1, Ω_2 *each containing a set of events, let* $W_1 \in \Omega_1$ *and* $W_2 \in \Omega_2$ *be two events. Let* $\kappa(W_1) = 2$, $\mathfrak{z}(W_1) = 0$, $\kappa(W_2) = 2$ *and* $\mathfrak{z}(W_2) = 1$.

Examining the κ *values associated with the events* W_1 *and* W_2 *does not provide an indication of their relative surprise relative to each other because* $\kappa(W_1)$ *is only*

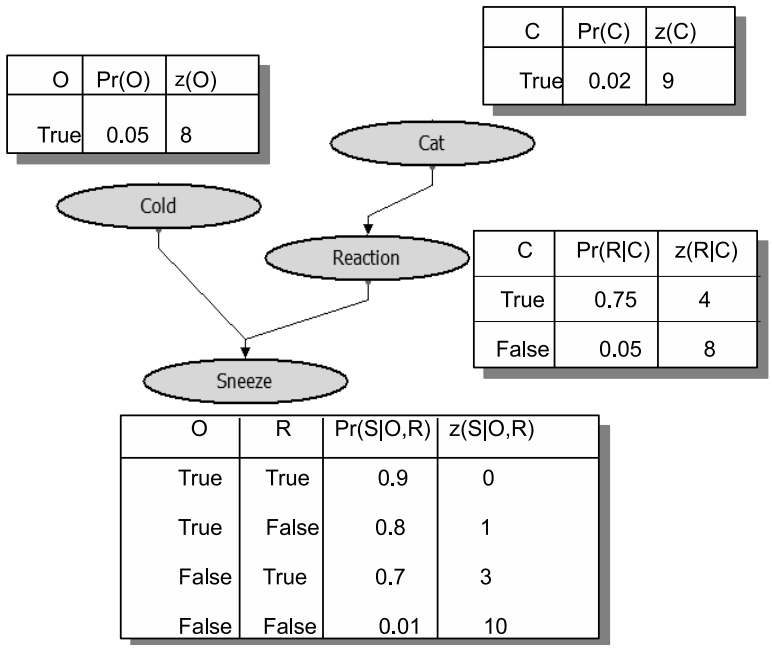


Figure 4.4: $\mathfrak{z}$ Network Representing Patients Information

meaningful within Ω_1 *and* $\kappa(W_2)$ *is only meaningful in* Ω_2*. Hence, despite the fact that these two events have equal* κ *values, no conclusions can be made about whether or not they are equally surprising in their respective universes.*

In contrast, one is able to deduce that within their respective distributions, W_2 *is more surprising than* W_1 *by examining their* $\mathfrak{z}$ *values as they describe the surprise associated with the events relative to the the minimum surprise associated with the domain (or the maximum probability).*

The above example demonstrates another useful facet of $\mathfrak{z}$, which is its ability to transform κ from a mere ordinal function to having a significant, domain-independent meaning to its ranks. This is done by finding the κ value which corresponds to $\mathfrak{z} = 0$ for the domain and restating the κ values relative to the one found.

In addition to the above feature of $\mathfrak{z}$, experimental studies concerning the values of $\mathfrak{z}$ show that the range its values assume is bigger than that of κ. As a result, the relative number of non-trivial rankings is bigger than that of κ, which makes $\mathfrak{z}$

inevitably more useful than κ. It is important to note however that this behavior is dependant on the maximum probability value in the domain. The closer this value is to one, the closer the behavior of $\mathfrak{z}$ is to that of κ. Figure 4.5 shows the range of $\mathfrak{z}$ and κ found for a domain containing probabilities ranging from 0.05 to 0.9 calculating using the value $\epsilon = 0.23$.

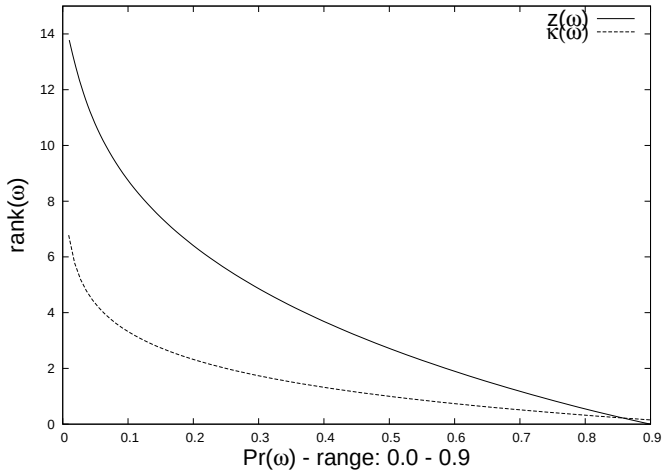


Figure 4.5: The Range of $\mathfrak{z}$ Compared to κ.

What makes $\mathfrak{z}$ an attractive function is that it lies between κ and κ^{++} in usefulness by being more expressive than κ while maintaining the monotonic behavior of surprise not present in κ++.

4.7 Summary

This chapter presented two formulations of surprise-based ranking functions as alternatives to the ordinal ranking function κ, mainly κ^{++} and the $\mathfrak{z}$ calculus. The aim of devising these calculi was to eliminate some of the disadvantages κ suffers from as a result of being a direct order of magnitude abstraction of the reverse of the probability.

For each proposed qualitative ranking function, this chapter presented its derivation from a numerical surprise measure, a thorough discussion of the resulting seman-

tics and the formulation of its propagation rules through conditioning.

An analysis of the newly-devised qualitative functions showed that the ranking offered by both is less sensitive to the infinitesimal nature of the ϵ used in the abstraction than in the κ function. As a result, the resulting functions are less prone to degenerating to useless trivial values than κ, which in turn reduces the likelihood of having to resort to values of ϵ that produce rankings that are inconsistent with those of probability.

However, the chapter also discovers that the sought-after semantical richness is not readily achievable at the level initially anticipated without hurdles. Although the signed integers of κ^{++} calculus offer this semantical richness by establishing a negative rank that is indicative of the anticipation associated with an event, this range of possible negative values is limited compared to the range of positive values that indicate a surprise and has some semantical issues. These problems are avoided in the $\mathfrak{z}$ calculus which although does not offer the same versatility of values as κ^{++}, offers a better ranking than κ in that it is less sensitive to the infinitesimal nature of the ϵ used and offers a ranking which is more consistent with the one given by probability than the ranking κ offers.

Chapter 5

Graphical Models for Surprise [1]

From the perspective of mere representation, the external world always remains only a phenomenon.

- Wilhelm Dilthey

Having established the theoretical base for the proposed ranking functions, the remainder of the dissertation is concerned with examining their performance empirically. The approach followed here is to involve the ranking functions proposed in a number of graphical models that abstract probabilistic system and perform a comparative study for each with respect to a real-life application. This chapter is concerned with the creation of the graphical models that will be used as test beds for the comparative studies.

The graphical models developed here represent different levels of abstractions but all use the ranking functions proposed in Chapter 4 to either 1) perform the reasoning required by the graphical model 2) provide conflict resolution to more abstract forms

[1] This chapter incorporates the outcome of a joint research undertaken under the supervision of Professors Ahmed Tawfik and Alioune Ngom. The key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of co-authors was primarily through the provision of advice when needed.

of qualitative uncertainty propagation (e.g. QPNs). The probabilistic models chosen to be used as bases for the abstraction are:

1. Bayesian Networks (BNs)
2. Dynamic Bayesian Networks (DBNs)
3. Hidden Markov Models (HMMs)

These models were chosen because of their avid use in Artificial Intelligence as well as the importance they have been shown to have with respect to the application domain Chapter 6 is concerned with. More specifically, each has a number of model-specific properties that make it an attractive choice with respect to the application discussed in Chapter 6 and which will be discussed separately for each model.

The remainder of the chapter introduces the three qualitative graphical representations of the models chosen. In Section 5.1, a Qualitative Probabilistic Network that uses κ^{++}'s used as strength factors is introduced. The idea is extended to formulating a qualitative equivalent of Dynamic Bayesian Networks in Section 5.2. The last model introduced in Section 5.3 is a Qualitative HMM uses ranking functions instead of numerical probabilities for the specification of its various parameters. The chapter ends with a summary in Section 5.4.

5.1 Surprise-based Qualitative Probabilistic Networks

In this section, an examination of the power of the ranking functions with respect to reasoning similar to that performed in Bayesian Networks is introduced. The model presented here is a Qualitative Probabilistic Network (Section 3.1.1) that uses κ^{++}'s to define its influences and uses the resulting rankings as strength factors for conflict resolution. Before the model is introduced, several points pertaining to the choice of the model are given below.

1. The ranking function is not used directly to perform conditional propagation the way probabilities are used in Bayesian Networks but are instead used as a base to define qualitative influence relations among the networks variables similarly to Qualitative Probabilistic Networks. This is done because a network that performs node-based reasoning by replacing probabilities with either κ^{++}'s or $\mathfrak{z}$'s will not enjoy the added efficiency of arc-based reasoning of the other forms of abstractions of Bayesian Networks (i.e. QPNs). This is evident from the analysis of the complexity of reasoning of κ-based networks presented in (Darwiche, 1992) which shows that its evidence propagation algorithms are $\mathcal{NP}$-hard in the worst case, similarly to Bayesian Networks. The added efficiency is essential to the model presented here as it was designed to be used in the application domain presented in Chapter 6, which pertains to a complex environment with data of an exponential size captured by a large number of variables.

2. The model presented here uses only κ^{++} as a base for the definition of its various constructs. We do not present a model which uses $\mathfrak{z}$ for the formation of the strengths of its influences. The reason behind this is the similarity between $\mathfrak{z}$ and κ with respect to the range of their ranking values (of being constricted to unsigned integers). A QPN that uses $\mathfrak{z}$ as strength factors will be theoretically identical to a QPN using κ's, which is an-already existing model available in (Renooij et al., 2003). This is not to say that comparing the effectiveness of κ's and $\mathfrak{z}$'s as strength indicators is not worthy of investigation as it is studied in Chapter 6 using the model given in (Renooij et al., 2003) and replacing κ by $\mathfrak{z}$. The idea here is to present a model that may possess additional properties due to the increased range of κ^{++} rankings.

5.1.1 QPNS with κ^{++}-based Strength Indicators

The sign-magnitude features of κ^{++} make it an attractive choice for modeling influences in a Qualitative Probabilistic Network (Section 3.1.1). Using κ^{++}, it is possible to redefine the partial order relations representing influences to possess not only a sign, but instead a sign-magnitude pair.

As a result, the signed integer representing κ^{++} can be used to evaluate the relative strength of influences and to propagate them across the network. Tradeoff resolution comes natural in this case because conflicting signs can be resolved by assessing the magnitude of the influences in conflict. The result is a κ^{++} network capturing the semantics of conditional independence that can be used to propagate beliefs qualitatively and has a built-in conflict-resolution mechanism. In what follows, we define the notion of κ^{++}-based influences.

κ^{++}-based Influences

Four types of κ^{++}-based influences are defined, analogous to those defined over QPNs. They are positive, negative, zero and unknown. This section is restricted to discussing the first three types of influences as a discussion of unknown influences is delayed to section 5.1.1.

Positive Influences: A binary variable W_i is said to positively influence another binary variable W_j if the degree of conditional surprise associated with W_j being true given that W_i is observed is lower than that of W_j being true given that W_i is not observed regardless of the value of any other variable which may directly influence W_j. This implies that for any values w_i of W_i and w_j of W_j, the conditions given in Definition 19 must hold in order to accomplish a positive influence.

Definition 19. $I^{+\delta}_{\kappa^{++}}(W_i, W_j)$ *iff* $\kappa^{++}(w_j|w_i, W) - \kappa^{++}(w_j|\neg w_i, W) < 0$

W represents any variable other than W_i which has a direct influence on W_j, and maybe written as $\pi(W_j)\backslash\{W_i\}$ (where there is more than one such variable, W is thought of as the conjunction of the possible values of such variables (Parsons, 2001)). The influence is denoted by $I^{+\delta}_{\kappa^{++}}(W_i, W_j)$ where δ represents the magnitude of the influence where the $+$ denotes that it is a positive one. The subscript κ^{++} enforces the idea that the influence is defined over κ^{++} values and not probability values as in QPNs. The same nomenclature is followed for negative, zero and unknown influences.

It is important to see that the semantics of κ^{++} guarantee that the constraints

given by the definition holds, which is what we show in Proposition 1.

Proposition 1. *For two binary variables W_i and W_j :*

$$I^{+\delta}_{\kappa^{++}}(W_i, W_j) \rightarrow \kappa^{++}(w_j|w_i, W) - \kappa^{++}(w_j|\neg w_i, W) = \delta \in \mathbb{Z}^-$$

Proof.

There are essentially two cases that result from the inequality $\kappa^{++}(w_j|w_i, W) < \kappa^{++}(w_j|\neg w_i, W)$:

- Case 1: $\kappa^{++}(w_j|w_i, W) \in \mathbb{Z}^-$ and $\kappa^{++}(w_j|\neg w_i, W) \in \mathbb{Z}^+$

  In this case, the fact that $\kappa^{++}(w_j|\neg w_i, W) - \kappa^{++}(w_j|w_i, W) \in \mathbb{Z}^-$ is intuitive.

- Case 2: Both $\kappa^{++}(w_j|w_i, W)$ and $\kappa^{++}(w_j|\neg w_i, W) \in \mathbb{Z}^+$

  In this case, the semantics of κ^{++} enforces that for $\kappa^{++}(w_j|w_i, W)$ to be less surprising than $\kappa^{++}(w_j|\neg w_i, W)$, it must possess a smaller magnitude, which will guarantee the result.

- Case 3: Both $\kappa^{++}(w_j|w_i, W)$ and $\kappa^{++}(w_j|\neg w_i, W) \in \mathbb{Z}^-$

  In this case, the semantics of κ^{++} enforces that for $\kappa^{++}(w_j|w_i, W)$ to be more anticipated than $\kappa^{++}(w_j|\neg w_i, W)$, it must possess a higher magnitude, which will guarantee the result.

□

Negative Influences Similarly to positive influences, a binary variable W_i negatively influences another binary variable W_j if the degree of conditional surprise associated with W_j being true given W_i is observed, $\kappa^{++}(w_j|w_i)$, is higher than that of W_j being true given that W_i is not observed $\kappa^{++}(w_j|\neg w_i)$ regardless of the value of any other variable which may directly influence W_j as given in Definition 20 below.

Definition 20. $I^{-\delta}_{\kappa^{++}(W_i,W_j)}$ *iff* $\kappa^{++}(w_j|w_i, W) - \kappa^{++}(w_j|\neg w_i, W) > 0$

Similarly, the semantics of κ^{++} guarantee that the constraints given by the definition holds as in Proposition 2.

Proposition 2. *For two binary variables W_i and W_j :*

$$I^{-\delta}_{\kappa^{++}}(W_i, W_j) \rightarrow \kappa^{++}(w_j|w_i, W) - \kappa^{++}(w_j|\neg w_i, W) = \delta \in \mathbb{Z}^+$$

Proof.

There are essentially two cases that result from the inequality $\kappa^{++}(w_j|w_i, W) > \kappa^{++}(w_j|\neg w_i, W)$:

- Case 1: $\kappa^{++}(w_j|w_i, W) \in \mathbb{Z}^+$ and $\kappa^{++}(w_j|\neg w_i, W) \in \mathbb{Z}^-$

  In this case, the fact that $\kappa^{++}(w_j|\neg w_i, W) - \kappa^{++}(w_j|w_i, W) \in \mathbb{Z}^+$ is intuitive.

- Case 2: Both $\kappa^{++}(w_j|w_i, W)$ and $\kappa^{++}(w_j|\neg w_i, W) \in \mathbb{Z}^+$

  In this case, the semantics of κ^{++} enforces that for $\kappa^{++}(w_j|w_i, W)$ to be more surprising than $\kappa^{++}(w_j|\neg w_i, W)$, it must possess a greater magnitude, which will guarantee the result.

- Case 3: Both $\kappa^{++}(w_j|w_i, W)$ and $\kappa^{++}(w_j|\neg w_i, W) \in \mathbb{Z}^-$

  In this case, the semantics of κ^{++} enforces that for $\kappa^{++}(w_j|w_i, W)$ to be less anticipated than $\kappa^{++}(w_j|\neg w_i, W)$, it must possess a smaller magnitude, which will guarantee the result.

□

Zero Influences are defined in the same manner and is given in Definition 21.

Definition 21. $I^{0}_{\kappa^{++}(W_i,W_j)}$ *iff* $\kappa^{++}(w_j|w_i, W) - \kappa^{++}(w_j|\neg w_i, W) = 0$

Although the influences given in this work are defined over binary variables, the definitions can be naturally extended to multi-valued variables as we have adopted the order of $w_i > \neg w_i$ to denote that a true value has a higher value than a false one.

Influence Propagation

To combine influences, the $\bigoplus$ and $\bigotimes$ operators are redefined in order to accommodate the signed integer nature of the κ^{++}-based influences.

Chained Influences: As done in (Druzdzel and Henrion, 1993b; Renooij et al., 2003; Wellman, 1990a), we propagate influences along chains using the order of magnitude multiplication operator. Since our influences include sign and magnitude components, these components are handled separately to obtain the net effect on the variables.

The sign portion of the influence is dealt with using sign multiplication as in (Wellman, 1990a) while the magnitude portion is handled in accordance with the rules of order of magnitude multiplication by adding the corresponding values (since the magnitude represent the difference between two κ^{++} values, which are in essence order of magnitude abstractions of the numerical surprise associated with the variable). The result is presented in the table below.

Table 5.1: The $\bigotimes$ Operator Combination Rules

$\bigotimes$	$+ve$	$-ve$	0	?
$+u$	$+(u+v)$	$-(u+v)$	0	?
$-u$	$-(u+v)$	$+(u+v)$	0	?
0	0	0	0	0
?	?	?	?	?

Parallel Influences: For influences in parallel chains, the net effect is decided by that of the strongest influence incident on the node. Accordingly, the effect is achieved via the $\bigoplus$ operator, presented in the table given below.

Example 14. *Figure 5.1 shows a fictitious network whose influences are defined using* κ^{++}. *In the figure, when nodes* A *and* C *are received as evidence, the discovery of the influences in the network propagates as follows. The net influence of node* A *on node* E *through* B *is given by -1* $\bigotimes$ *+2 = -3 because this influence consists of two influences in a chain whose effect is obtained via the* $\bigotimes$ *operator. Similarly, Node* D *receives evidence from both* A *and* C *with the net influence being evaluated as +4* $\bigoplus$ *-5 = +4 because node* D *has two arcs incident on it, which implies that the net effect on* D *is obtained through the discovery of the combined influences in parallel, which*

Table 5.2: The $\bigoplus$ Operator Combination Rules

$\bigoplus$	$+ve$	$-ve$	0	?
$+u$	$+\min\{u, v\}$	a)	$+u$	?
$-u$	b)	$-\min\{u, v\}$	$-u$	?
0	$+v$	$-ve$	0	?
?	?	?	?	?

a) = $+u$, if $u < v$
= $-v$, otherwise
b) = $-u$, if $u < v$
= $+ve$, otherwise

is achieved through the $\bigoplus$ *operator. Similarly, the net influence of* A *and* C *on* E *through* D *is given by +4* $\bigotimes$ *+5 = +9. Finally, node* E *receives as a net influence -3* $\bigoplus$ *+9 = -3. As a result, the net influence of observing* A *and* C *on* E *is a negative one.*

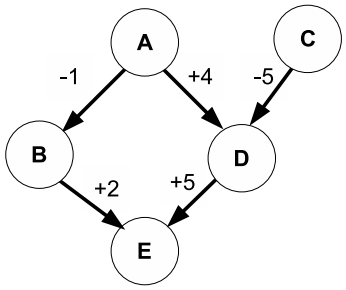


Figure 5.1: Reasoning with a κ^{++}-based Qualitative Probabilistic Network

The Case of Unknown Influences

Because influences only exist when one is able to establish a partial order on the conditional κ^{++} of two variables (Parsons, 2001), it is a weak concept that may not be defined when such order does not exist. However, the more versatile range for strength factors used in κ^{++} networks makes the reaching an unknown influence a less likely outcome compared to κ-based QPNs (which is empirically studied in Chapter 6). Moreover, since our networks are based on κ^{++} values, it is not necessary to resort

to probabilities when an unknown influence is reached and is sufficient to go back to node-based inference on the κ^{++}-level. Although this reduces the efficiency of the inference, it is a necessary last resort when orders are not definable. Moreover, the network retains its qualitative nature as we are still dealing with κ^{++}'s, which are easier to assess than numerical probabilities.

5.2 Temporal Extension of Surprise-based QPNs

The last model presented in this chapter is a temporal extension of QPNs that uses surprise as its uncertainty base. The model is built so that applications that require dealing with time-series data can make use of the propagation rules formulated here.

5.2.1 Terminology

Let U be a set of n events drawn from some distribution Pr and let T be a totally ordered set of m temporal slices such that $T_1...T_m \in T$. We denote the set of events in each temporal slice by U^t $(1 \leqslant t \leqslant m)$ and the set of n events in U^t by A_i^t $(1 \leqslant i \leqslant n)$. A static snapshot of U is termed a temporal snapshot and is given in Definition 22.

Definition 22. ***Temporal Snapshot:***

Let $G = (V(G), A(G))$ be a directed acyclic graph (DAG) such that G is an I-map for Pr, the joint probability distribution defined on U[2]. An instance G_t of G represents a temporal snapshot of G in time slice T_t such that G_t retains the DAG structure of G.

Example 15. *Figure 5.2 represents a fictitious graph G capturing the I-map for Pr, the joint probability distribution on $U = \{A_1, A_2, A_3, A_4\}$. Each instance G_t of G $(1 \leqslant t \leqslant 3$ in the figure) represents a snapshot of G, where the variables in each temporal slice are given by: $U_t = \{A_1^t, A_2^t, A_3^t, A_4^t\}$.*

An instance of the temporal snapshot of G is termed a dynamic instance of the graph G and is defined below.

[2]G is the qualitative probabilistic network representing U

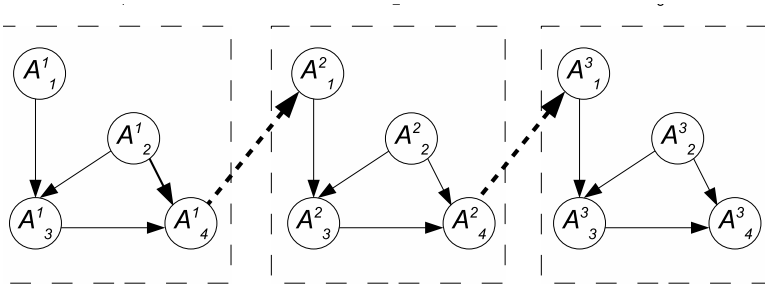


Figure 5.2: An Example of G

Definition 23. ***Dynamic Instance:***

Let G_t be as given in Definition 22. G_t defines a dynamic instance of the QPN whose structure is defined by G and is given by $G_t = (V(G_t), \{A(G_t) \bigcup T(G_t)\})$ [3], where $V(G_t)$ and $A(G_t)$ are instants of $V(G)$ and $A(G)$ respectively at time slot t, and $T(G_t)$ describes the inter-slot conditional dependence between variables in $V(G_t)$ and its immediate neighbor $V(G_{t+1})$.

Example 16. *In the graph given in Figure 5.2, for each G_t, $V(G_t) = U_t$, $A(G_t) = \{(A^t_1, A^t_3), (A^t_2, A^t_3), (A^t_3, A^t_4), (A^t_2, A^t_4)\}$ and $T(G_t) = \{(A^t_4, A^{t+1}_1)\}$.*

Both of $A(G)$ and $T(G)$ encode a set of hyperarcs for G to capture a set of qualitative relations representing how variables influence each other. For this, the set of qualitative influences are re-defined to express inter-slot relations in addition to within-slot ones. Before introducing the new set of influences however, it is necessary to define the proposed Temporal Qualitative Probabilistic Networks (TQPN) below.

Definition 24. ***Temporal QPN:***

Let $(G_1 = (V(G_1), Q(G_1)), ..., G_m = (V(G_m), Q(G_m)))$ be a total ordering of the m instances of G such that $T(G_t) \neq \phi \; \forall \; 1 \leqslant t \leqslant m-1$. Then the compound graph of $G_1, ..., G_m$ defines a Temporal Qualitative Probabilistic Network over G and is given by:

$$\bigcup_{t=1}^{m} G_t = (\bigcup_{t=1}^{m} V(G_t), \bigcup_{t=1}^{m} Q(G_t))$$

[3]For readability purposes, we will refer to $\{A(G_t) \bigcup T(G_t)\}$ as $Q(G_t)$ in this work.

5.2.2 Qualitative Influences in a TQPN

Definition 25. ***Positive TQPN Influence:***

Let G_t and G_{t+1} be two adjacent subgraphs of the TQPN defined over G. Further, let B and C be such that $B, C \in V(G)$. A direct positive influence is exerted by node B over node C, written as $S^+(B, C)$ iff for all values c_i^x of C and b_j^y, b_k^y of B with $b_i^y > b_k^y$, and for all integer values x and y such that $1 \leqslant x, y \leqslant m$ and $x - y \in \{0, 1\}$ we have:

$$\vartheta(C \geqslant c_i^x | b_j^y, w) - \vartheta(C \geqslant c_i^x | b_k^y, w) = \gamma \in \mathbb{Z}^-$$

Where ϑ is an order of magnitude ranking function and γ is as in Section 5.3, the strength of the influence exerted by B on C.

Where w represents any combination of values for the set of nodes W which represent all other direct influences on C other than B. The superscripts x and y denote the temporal slot to which the instances c_i, b_j and b_k belong. The definition necessitates that variables can only directly influence other variables that belong to the same temporal slot ($x = y$) or those that belong to the next immediate slot ($x - y = 1$). Negative, zero and unknown influences are analogously defined.

In order to resolve the likely-to-occur ambiguities, we mimic the mechanisms given in (Renooij and Gaag, 2008) and define *indirect influences* that are augmented with two levels of strength and a multiplication index as given in Definition 26.

Definition 26. ***Strongly Positive TQPN Influence:***

Let B and C be two nodes in the TQPN defined over G. Furthermore, let tr be a trail from B to C. Let W be all the other nodes that can influence C and that do not belong to the trail from B to C. Then the qualitative influence from node B to node C along trail tr is strongly positive with multiplication index μ, $\mu \in \mathbb{N}$, written as $S^{++^\mu}(B, C, tr)$ iff for all values c_i^x of C and b_j^y, b_k^y of B with $b_i^y > b_k^y$:

$$\vartheta(C \geqslant c_i^x | b_j^y, w) - \vartheta(C \geqslant c_i^x | b_k^y, w) = \gamma \leqslant \alpha^\mu$$

Moreover, the qualitative influence of B on C along trail tr is weakly positive with multiplicative index μ, $\mu \in \mathbb{N}$, written as $S^{++}(B, C, tr)$ iff:

$$0 \geqslant \vartheta(C \geqslant c_i^x | b_j^y, w) - \vartheta(C \geqslant c_i^x | b_k^y, w) \geqslant \alpha^\mu$$

Where w represents any combination of values from the set W and $x - y \in \{0, 1\}$. The value μ is given by the length of the trail tr and $\alpha = [0 - 1]$ is the cut-off value used for distinguishing between strong and weak influences and which can be chosen by an expert [4]. In addition to the cut-off value α which distinguish strong from weak influences, influences of the same strength can be compared using their μ value, where higher values indicate a longer trail tr, and as a result, a weaker influence (Renooij and Gaag, 2008).

As a result, surprise-propagation in TQPNs offers multi-level conflict resolution. On the one hand, influence's sign is augmented by the signs multiplication index superscript, and is used as an indicator of its strength. Higher values of multiplication indices indicate a longer path and as a result, a weaker influence. This level of detail enables the use of using the Enhanced QPN $\bigoplus$ and $\bigotimes$ operators described earlier in Tables 3.3 and 3.2 of Chapter 3 to the different types of influences. On top of this, each influence is given a value γ that indicates its ranking strength and is calculated at the time the influence is evaluated. γ is in turn used to resolve conflicts using the propagation rules given in Tables 5.2 and 5.1 when propagation using Tables 3.3 and 3.2 gives an ambiguous sign.

5.3 Qualitative Hidden Markov Models

Hidden Markov Models (HMMs) (Rabiner, 1989) are probabilistic graphical models that capture the dependencies between random variables in time-series data. HMMs can be regarded as a special case of Dynamic Bayesian Networks (Smyth, 1997) that are not as powerful as they are but are far more efficient. The well-known first-order HMM has been particularly successful in several areas of artificial intelligence such

[4] The choice of α is experimentally determined.

as speech recognition (Rabiner, 1989; Rosti and Gales, 2003), robotics (Fox et al., 2006), pattern recognition (Lovell, 2003) and several areas of bioinformatics, such as transmembrane protein classification (Kahsay et al., 2005), to perform predictive and recognitive tasks. The power of HMMs stems from the provision of efficient and intuitive algorithms to perform inference due to the first-order nature of the model as each variable is only dependent on its equivalent in the previous temporal slice and no other variable.

While variations of traditional HMMs proved to be practical in applications where it is feasible to obtain the numerical probabilities required for the specification of the parameters of the model and the probabilities available are descriptive of the underlying uncertainty, the capabilities of HMMs remain unexplored in applications where this convenience is not available (Huang et al. (2001) present an example in Economics). Motivated by such applications, this section presents a HMM that does away with probabilities and instead uses ranking functions for the specification of its various parameters. The ranking-based HMM presented here is a general one in the sense that it is defined for a hypothetical ranking function that can be substituted by κ, κ^{++}, $\mathfrak{z}$ or any other ranking function. Substituting the general qualitative HMM by one that is specific to a ranking function is done in Chapter 6 where the capabilities of the different ranking functions are put to practice.

Section 5.3.1 below provides an introduction to standard Hidden Markov Models before the Qualitative HMM model is presented starting in Section 5.3.2.

5.3.1 Hidden Markov Models

Hidden Markov Models (HMMs) (Rabiner, 1989) are probabilistic graphical models used to represent the behavior of a system which is known to possess a number of states. The states of the model are hidden, in the sense that their operations can only be studied through discrete time series of the observed output produced by the states.

Formally, a HMM=$\{$S,V,π,A,$B\}$ is defined by the following parameters:

1. A finite set of n unobservable (hidden) states S=$\{s_1,...,s_n\}$

2. A finite set of m observable outputs, or the alphabet of the model: V = $\{v_1,...,v_m\}$ that may be produced by the states given in S at any time t.

3. The vector π of the the initial state probability distribution, i.e. the probability of the system being at state s_i at time 0: $P(q_0{=}s_i)$, $\forall\ s_i \in$ S $(1 \leqslant i \leqslant n)$.

4. The matrix $A = [a_{ij}]_{1 \leqslant i \leqslant n}$ which describes the transition probability distribution among associated states. For each entry a_{ij} in A, $a_{ij} = P(q_t = s_i | q_{t-1} = s_j)$, $\forall$ $1 \leqslant i, j \leqslant n$, which describes the probability of the system being in state s_i at time t given that it was in state s_j at time $t-1$. This formulation reflects the Markov property which dictates that the next state is only dependent on the current state, and is independent of previous states. This property also implies that the transition probabilities must satisfy:

$$\sum_{i=0}^{n} \sum_{j=0}^{n} P(q_t = s_i | q_{t-1} = s_j) = 1$$

5. The matrix $B = \{b_j(o_t), 1 \leqslant j \leqslant n\}$ of the emission probabilities of the observable output at a given state $P(o_t = v_i | q_t = s_j)$, which describes the probability of the system producing output v_i at time t given that it is in state s_j $(1 \leqslant i \leqslant m)$. This information reflects the assumption that the output at a given time is only dependent on the state that produced it and is independent of previous output. In other words:

$$\sum_{i=0}^{m} P(o_t = v_i | q_t = s_j) = 1$$

Hence, a HMM can be described by a doubly stochastic structure. The first stochastic process provides a high-level view of the system and is operated by a Markov chain (described by the transition matrix A) governing the transitions among the hidden states. The second stochastic process, on the other hand, is the one governing the production of observable output independently by each state (described by the emission matrix B). This structure provides HMMs with a high degree of flexibility, which makes them attractive for sequential data analysis.

5.3.2 A Qualitative HMM

Having introduced HMMs, it is now possible to define the semantics of its qualitative equivalent, HMM_ϑ. It is a HMM which uses a qualitative ranking function that abstracts numerical probabilities, ϑ, as a measure of the surprise associated with its transitions. This section will assume the existence of $\lambda = (\text{S,V,}\pi\text{,}A\text{,}B)$, a Hidden Markov Model with n possible hidden states and m observable outputs, and whose structure and parameters are specified, and uses it to define the semantics.

Semantics

Introducing ϑ values to the transitions of λ gives the model the following semantics:

1. $\forall \pi_i \in \pi, 1 \leqslant i \leqslant n$, π_i represents the degree of surprise associated with having state i to be true at time 0:

$$\pi_i = \vartheta(q_0 = s_i)$$

2. $\forall a_{ij} \in A$, where $1 \leqslant i, j \leqslant n$, a_{ij} represents the degree of surprise associated with state s_i holding at time t given that state s_j was true at time $t-1$. The resulting matrix A is called the transition ϑ matrix.

$$a_{ij} = \vartheta(s_t = q_i | s_{t-1} = q_j)$$

3. $\forall b_j(o_t = v_i) \in B$, where $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$, b_{ij} represents the degree of surprise associated with state s_j being responsible for producing observable output v_i at time t. The resulting matrix B is called the emission ϑ matrix.

$$b_j(v_i) = \vartheta(o_t = v_i | q_t = s_j)$$

Independence Assumptions

The semantics of order of magnitude abstractions of probability are used to reformulate the independence assumptions to go along with the semantics of λ. The reformulations are presented below.

1. **The limited memory assumption:** states that the degree of surprise associated with observing output v_i at time t being generated by state s_j is only dependent on the degree of surprise associated with state s_j, with any other state or output being irrelevant. This is represented as:

$$\begin{aligned}\vartheta(o_t = v_i | q_t = s_j, q_{t-1} = s_k, ..., q_1 = s_l,\\ v_{t-1} = o_m, ..., v_1 = o_n) = \vartheta(o_t = v_i | q_t = s_j).\end{aligned} \tag{5.1}$$

   Accordingly, the emission ϑ matrix should satisfy:

$$\min_{i=0}^{m} \vartheta(o_t = v_i | q_t = s_j) = \gamma \tag{5.2}$$

   Where:

$$\gamma = \begin{cases} 0 & , \textit{iff } \vartheta : \Omega \rightarrow \mathbb{Z}^+ \cup \{+\infty\} \\ \min\limits_{v \in V}\{\vartheta(v)\} & , \textit{otherwise.} \end{cases}$$

   The above assigns a value of zero to γ if the ranking function used has positive (unsigned) integers as range (as in the case of κ and $\mathfrak{z}$), and the minimum ranking value associated with the domain otherwise (as in the case of κ^{++}).

2. **The Markov assumption:** dictates that the degree of surprise associated with observing state s_i at time t is only dependent on the degree of surprise associated with the previous state, i.e. state s_j at time $t-1$, with all other states and output being irrelevant. This is represented as:

$$\begin{aligned}\vartheta(s_t = s_i | q_{t-1} = s_j, q_{t-2} = s_k, ..., q_1 = s_l,\\ v_t = o_m, ..., v_1 = o_n) = \vartheta(q_t = s_i | q_{t-1} = s_j)\end{aligned} \tag{5.3}$$

   Again, having this assumption in conjunction with the semantics of ϑ yields the following:

$$\min_{i=1}^{n} \vartheta(q_t = s_i | q_{t-1} = s_j) = \gamma \tag{5.4}$$

Where:

$$\gamma = \begin{cases} 0 & , \textit{iff } f\vartheta = \kappa \textit{ or } \mathfrak{z}; \\ \min\limits_{s \in S}\{\vartheta(s)\} & , \textit{otherwise.} \end{cases}$$

Additional Properties

Two interesting concepts arise from the introduction of the semantics of λ. They are:

1. **Output generator:** A state s_i , $1 \leqslant i \leqslant n$ is the generator of output v_j, $1 \leqslant j \leqslant m$ at time t *iff* s_i is the state associated with the minimum degree of surprise of having produced v_j

$$\vartheta(o_t = v_j, q_t = s_i | \lambda) = \min_{i=1}^{n} \vartheta(o_t = v_j, q_t = s_i | \lambda) \tag{5.5}$$

2. **State generator:** A state s_i , $1 \leqslant i \leqslant n$ at time t is the generator of state s_j, $1 \leqslant j \leqslant n$ at time $t+1$ *iff* s_i is the state holding at time t which is associated with the minimum degree of surprise of having preceded state s_j at time $t+1$

$$\vartheta(q_{t+1} = s_j, q_t = s_i | \lambda) = \min_{i=1}^{n} \vartheta(q_{t+1} = s_j, q_t = s_i | \lambda) \tag{5.6}$$

5.3.3 Inference in HMM$_\vartheta$

There are essentially three problems associated with HMM inference. They are:

- **Evaluation:** Given a HMM, λ, and a sequence of observed output O, the evaluation problem is concerned with determining the likelihood of O being a valid sequence produced by λ.

An efficient algorithm, the *forward algorithm* Rabiner (1989) finds a solution to the evaluation problem by performing induction on the length of the output sequence O=$o_1, o_2, ..., o_t$ as follows.

An inductive variable, called the *forward variable* is defined to capture the likelihood of observing the output sequence O of length t and having state s_i $(1 \leqslant i \leqslant n)$ to be the state that produced the last output of the sequence, o_t. Hence the *forward variable* captures the probability $P(O, q_t = s_i|\lambda)$. Induction is then used to derive the probability associated with increasing the length of the output sequence by one (i.e. observing one more letter, o_{t+1} at time $t+1$), and calculating the resulting probability, i.e. $P(O, o_{t+1}, q_t = s_i, q_{t+1} = s_j|\lambda)$.

- **Decoding:** Given the observation sequence $O = o_1, o_2, ..., o_t$ of length t and a model λ=(A,B,π), the decoding problems is concerned with finding the sequence of states $q = q_1, q_2, ..., q_t$ that was most likely to have produced the observation sequence O.

- **Learning:** Given an HMM λ with unknown parameters. The learning problem is concerned with finding out the values of A, B and π from data.

This section presents algorithms for the evaluation and decoding algorithms of HMM$_\vartheta$.

Evaluating Observed Output

The evaluation problem for HMM$_\vartheta$ can be formulated as follows. Given the structure and parameters of a HMM$_\vartheta$, λ, and an output sequence O of length t, the task is to find the likelihood of the sequence O being produced by λ by computing the degree of surprise associated with O given λ.

We redefine the *forward variable* $f_t(i)$ to be the inductive variable capturing the degree of surprise associated with observing the output sequence O of length t and having state s_i $(1 \leqslant i \leqslant n)$ to be the state that produced the last output of the sequence, o_t at time t, i.e. $\vartheta(O, q_t = s_i|\lambda)$.

$$f_t(i) \quad = \quad \vartheta(O, q_1, q_2, ..., q_{t-1}, q_t = s_i|\lambda) \tag{5.7}$$

If o_t is indeed a valid output generated by state s_i of λ, then state s_i is the output generator of o_t. This enables writing equation (5.7) as a variation of (5.5), which amounts to the following:

$$f_t(i) \quad = \quad \min_{i=1}^{n} \vartheta(o_1, ..., o_t, q_1, ..., q_{t-1}, q_t = s_i|\lambda) \tag{5.8}$$

The Qualitative Forward Algorithm

The algorithm finds the solution by solving for $f_t(i)$ inductively as follows:

1. Initialization:

$$\begin{aligned} f_1(i) &= \vartheta(o_1, q_1 = s_i|\lambda) && (5.9)\\ &= \vartheta(o_1|q_1 = s_i, \lambda) + \vartheta(q_1 = s_i, \lambda) && (5.10)\\ &= b_i(o_1) + \pi_i && (5.11) \end{aligned}$$

   The initialization step applies the inductive variable to the base case for which the length of the output sequence is 1. The conditional propagation associated with ϑ transforms the variable $f_1(i)$ given in (5.9) to the expression given in (5.10). In (5.10), $\vartheta(o_1|q_1 = s_i, \lambda)$ is the emission ϑ value associated with the only output o_1 being produced (by state s_i) at time 0 and $\vartheta(s_i, \lambda)$ is the initial degree of surprise associated with state s_i, which amounts to the expression given in (5.11).

2. Induction:

   The inductive step applies the inductive variable to the case where the sequence O is of length $t + 1$ and where state s_j is responsible for producing the output o_{t+1}. We hence devise a new variable $f_{(t+1)}(j)$ which represents the degree of surprise associated with observing an output sequence of length $t+1$ with state

s_j being the one that produced $o_{(t+1)}$, given that $f_t(i)$ holds at time t. The inductive variable $f_{(t+1)}(j)$ is given in equation (5.12) and is derived below.

Starting with the forward variable obtained in equation (5.8), and using order of magnitude conditional propagation, $f_{t+1}(j)$ can be rewritten by assigning ψ the values of $o_1, ..., o_t, q_1, ..., q_t$ and ϕ those of $o_{(t+1)}, q_{(t+1)} = s_j, \lambda$ as given below.

$$f_{t+1}(j) = \min_{j=1}^{n}[\vartheta(o_1, ..., o_t, o_{(t+1)}, q_1, ..., q_t, q_{(t+1)} = s_j|\lambda))$$

$$= \min_{j=1}^{n}[\vartheta(o_1, ..., o_t, q_1, ..., q_t|\ o_{(t+1)}, q_{(t+1)} = s_j, \lambda)+ \vartheta(o_{(t+1)}, q_{(t+1)} = s_j, \lambda)]$$

The above equation is further rewritten using two properties, one is property (5.1) of HMM$_\vartheta$, making the term $\vartheta(o_1, ..., o_t, q_1, ..., q_t|o_{(t+1)}, q_{(t+1)} = s_j, \lambda)$ simply $\vartheta(o_1, ..., o_t, q_1, ..., q_t|\lambda)$ because the elements of $o_1, ..., o_t, q_1, ..., q_t$ are independent of $o_{(t+1)}$ and $q_{(t+1)}$ according to the memoryless independence assumption. The second property is order of magnitude conditional propagation rule with ψ being $o_{(t+1)}$ and ϕ being $q_{(t+1)} = s_j$.

$$f_{t+1}(j) = \min_{j=1}^{n}[\vartheta(o_1, ..., o_t, q_1, ..., q_t|\lambda) + \vartheta(o_{(t+1)}\ |q_{(t+1)}= s_j, \lambda)+ \vartheta(q_{(t+1)} = s_j, \lambda)]$$

The inductive hypothesis dictates that in order for $f_{t+1}(j)$ to be true, $f_t(i)$ must have been true. This makes the state that held at time t to be s_i, and the resulting equation is shown below.

$$f_{t+1}(j) = \min_{j=1}^{n}[\vartheta(o_1, ..., o_t, q_1, ..., q_t = s_i|\lambda)+\vartheta(o_{(t+1)}|q_{(t+1)} = s_j, \lambda)+ \vartheta(q_{(t+1)} = s_j|q_{(t)} = s_i\lambda)]$$

In the above equation, it can be clearly seen that the first term is the inductive variable $f_t(i)$, the second term is a_{ij}, while the third is $b_j(o_{(t+1)})$, which is only one value, and hence is taken out of the sum to give equation (5.12) given below, which is the final form of the inductive step.

$$f_{t+1}(j) = \min_{i=1}^{n}[f_t(i) + a_{ij}] + b_j(o_{(t+1)}) \tag{5.12}$$

As made clear in the derivation, the inductive step of equation (5.12), which computes $f_{t+1}(j)$, is only executed if at time t, the degree of surprise of having state s_i producing output o_t, given output sequence O and the complete specification of the structure and parameters of λ has been computed by $f_t(i)$.

3. Termination:

   Given that the inductive step computes the forward variable at every time step until t, starting with the base case at time 1, the inductive algorithm correctly terminates at step t, by calculating $f_t(i)$, $\forall\ 1 \leqslant i \leqslant n$, and consequently finding the degree of surprise associated with observing the last output o_t of the sequence O.

$$\vartheta(O|\lambda) = \min_{i=1}^{n} f_t(i) \tag{5.13}$$

The Decoding Problem

Given the observation sequence $O = o_1, o_2, ..., o_t$ of length t and a model λ=(A,B,π), the decoding problems is concerned with finding the sequence of states $q = q_1, q_2, ..., q_t$ that is associated with the minimum surprise ϑ of having produced the observation sequence O.

In order to evaluate candidate sequences of states, we require a quantity representing the degree of surprise associated with the most-likely sequence being one which ends with state $q_t = i$. We denote this quantity by $\delta_t(i)$.

$$\delta_t(i) = min_{q_1 q_2 ..., q_t} \kappa(q_1 ... q_{t-1}, o_1 ... o_t, q_t = i) \tag{5.14}$$

In order to use the score function to find the best sequence q, we should be be able to answer the question: what is the degree of surprise associated with the most-likely sequence being one which ends with state q_{t+1} being state j, given that the degree of surprise associated with the most-likely sequence being one which ends with state $q_t = i$ is $\delta_t(i)$? The answer is found by induction on the length of the sequence q as shown below.

$$\begin{aligned}\sigma_{(t+1)}(j) &= min_{\pi_1,...,\pi_t}\kappa(o_1...o_t,q_1,...,q_t,\ o_{t+1},q_{t+1}{=}j)\\ &= min_{q_1,...,q_t}\ [\kappa(o_{t+1},q_{t+1} = j \mid o_1...o_t,\ q_1,...,q_t) + \kappa(o_1...o_t,q_1,...,q_t)]\end{aligned}$$

Taking into account Markov and Independence assumptions
and redistributing the rest:
$= min_{q_1,...,q_t}\ [\kappa(o_{t+1},q_{t+1} = j \mid q_t) + \kappa(o_1...o_{t-1},q_1,...,q_{t-1},\ o_t,\ q_t)]$

However, the sequence that minimized the degree of surprise was the one that ended with state i and which was given by the equation 5.14. This makes the above:
$= min_t[\ \kappa(o_{t+1},q_{t+1} = j|q_t = i) + min_{q_1q_2...,q_t}\kappa(q_1...q_{t-1},o_1...o_t,q_t = i)$
$= min_t[\ \kappa(o_{t+1},q_{t+1} = j|q_t = i) + \sigma_t(i)]$
$= b_j(o_{t+1}) + min_t\ [a_{ij} + \sigma_t(i)]$

$$\sigma_{t+1}(j) = b_j(o_{t+1}) + min_t[a_{ij} + \sigma_t(i)] \tag{5.15}$$

A Qualitative Viterbi Algorithm

The algorithm keeps track of the argument which has minimized 5.15 at every time t and state j. For this, a vector $\varrho_t(j)$ is used. Hence, the qualitative viterbi algorithm can be described via the following steps:

1. Initialization

$$\sigma_t(i) = \pi_i + b_i(o_1), 1 \leqslant i \leqslant N \tag{5.16}$$

$$\varrho_1(i) = 0 \tag{5.17}$$

2. Recursion

$$\sigma_t(j) = b_j(o_t) + min_{1 \leqslant i \leqslant N}[a_{ij} + \sigma_{t-1}(i)] \quad 2 \leqslant t \leqslant T, 1 \leqslant j \leqslant N \quad (5.18)$$

$$\varrho_t(j) = argmin_{1 \leqslant i \leqslant N}[a_{ij} + \sigma_{t-1}(i)] \quad 2 \leqslant t \leqslant T, 1 \leqslant j \leqslant N \quad (5.19)$$

3. Termination

$$P^* = min_{1 \leqslant i \leqslant N}[\sigma_T(i)] \quad (5.20)$$

$$q_T^* = argmin_{1 \leqslant i \leqslant N}[\sigma_T(i)] \quad (5.21)$$

4. Path (state sequence) Backtracking

$$q_t^* = \varrho_{t+1}(q_{t+1}^*) \quad t = T-1, T-2, ..., 1 \quad (5.22)$$

5.4 Summary

This chapter introduced three graphical models, surprise-based QPNs, their temporal extensions (TQPNs) and a surprise-based Hidden Markov Model (HMM_ϑ). These models incorporate the ranking functions defined in Chapter 4 as either bases for the specification of the various model parameters or as a conflict resolution mechanism.

The aim was to create a set of models that enable the use of the ranking functions in a real-life applications. Hence, having developed the three models, the next chapter devises a set of experiments that test the capabilities of the ranking functions as well as examine their use in a complex application domain.

Chapter 6

Qualitative Aspects of Genetic Regulation [1]

As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.

- Albert Einstein, *answering the question: "how can it be that mathematics, being after all a product of human thought which is independent of experience, is so admirably appropriate to the objects of reality?"*

In this dissertation, the use of qualitative uncertainty formalisms has been motivated by having substitutes to quantitative methods when the precision they offer is either unattainable or unnecessary to perform plausible reasoning. This is especially evident in the discussion at the beginning of Chapter 3. This chapter is concerned with presenting the application domain chosen to implement the ideas presented so far: bioinformatics. This domain is intriguing as it does not have the lack of numerical information at the heart of its current problems; on the contrary, it suffers from

[1] This chapter incorporates the outcome of a joint research undertaken under the supervision of Professors Ahmed Tawfik and Alioune Ngom. The key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of co-authors was primarily through the provision of advice when needed.

the opposite issue: the abundance of numerical data.

Besides personal interest in bioinformatics, the reason behind choosing it as an application domain is to demonstrate that qualitative methods for dealing with uncertainty are not only an alternative for when data is not available, but are also useful where quantitative approaches have been proposed and successfully used. This is because qualitative equivalents of the quantitative methods available can serve as a guide for a better analysis of the available methods to obtain more biological insight, given that the large amount of data is what has made formulating mechanisms to provide better biological insight work in progress (Friedman, 2004). In other words, they can be used to perform an initial analysis to filter the data available, which aids in reducing the complexity of the full analysis performed by the quantitative methods.

This chapter presents two studies that aim at performing types of analysis on microarray gene expression data (introduced in Section 6.2). These studies represent the results of implementations that test the models developed in Chapter 5 and aim at 1) examining the ranking power of the functions formulated in Chapter 4 2) examining how well are the advantages supported in a real-life complex environment 3) demonstrating that a qualitative analysis can provide a good aid to existing quantitative approaches with respect to the application domain and improve the results available in terms of accuracy and efficiency.

The chapter is structured as follows. Sections 6.1 - 6.3 introduce concepts that are essential to the application domain along with the vocabulary associated with it. In Section 6.4, a set of findings that motivate the use of qualitative approaches for handling uncertainty in the domain of choice are introduced. Sections 6.5 - 6.6 contain the studies conducted to test the three aims given above. The chapter ends with a summary given in Section 6.7.

6.1 Gene Expression in the Cell

With only a few exceptions, every cell in the body of an organism contains a full set of chromosomes containing identical genes. Modeling the interactions of genetic

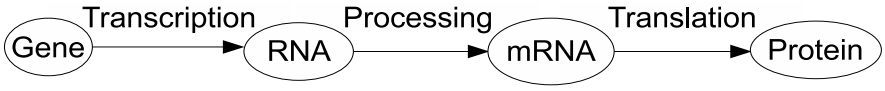


Figure 6.1: The Gene Expression Process

information in the cell have received increasing attention in recent years. The general aim is to study the variables making up the internal state of the cell in order to determine its general behavior by studying the interactions taking place on the gene level. More specifically, the process by which inheritable information located on the genes, i.e. DNA, is made into a functional end product (protein) is studied. This enables studying the behavioral properties of living organisms as the protein end products perform most of the critical cellular functions.

The process by which information from a gene is used in the synthesis of a functional gene product, or protein, is called *gene expression* (Hunter, 2004). The different proteins synthesized from different genes perform virtually every function within the cell and are essential for its survival. Hence, the process of gene expression is continuously taking place at the cell, producing the protein products necessary to perform the diverse functions that the cell requires to adapt to its changing needs (Someren et al., 2002).

A simplified model of the expression process is shown in Figure 6.1 and consists of three stages, each ending with the actualization of an end-product. First, genes *transcribe* the information contained within the DNA into RNA, which is then *processed* into messenger RNA (mRNA). mRNA is in turn receptive to other chemical products that *translate* it into functional protein, which performs most of the critical functions of cells. The process of expressing the information located on the DNA of a single gene to a functional protein is so important to modern biology that it is called the *Central Dogma* of molecular biology (Crick, 1970) and serves as the base for studying the flow of information during the cellular life cycle in order to understand biological systems at the cellular level (Hunter, 2004).

The expression process is triggered by many factors including chemical cellular factors (e.g. change in PH level, temperature, etc...) and the expression of other

genes. More specifically, the protein end products resulting form the expression of many genes trigger the expression of other genes. The result is a complex web of interactions among the genetic components and end-products (Pisabarro et al., 2008).

6.2 Measuring Expression Levels: Microarray Technology

Under any set of conditions, only a fraction of the genes are active, or *expressed*, and it is this subset that delivers the unique properties to each cell type and perform the functions required at the time. Gene expression is a highly complex and tightly regulated process that allows a cell to respond dynamically both to environmental stimuli and to its own changing needs. This mechanism acts as both an *activation/supression* switch to control which genes are expressed in a cell as well as a *volume control* that increases or decreases the level of expression of particular genes as necessary. As a result, in order to obtain insight to how the cell responds to the changing needs of itself and the environment, one must measure the expression level of the different genes during different time slots or under different cellular conditions (D'hæseleer, 2000).

As an organism's genome may consist of tens of thousands of genes, identifying the genes that are expressed under any conditions entails the need for simultaneously measuring the expression levels of a large number of genes and a method for distinguishing those genes that are expressed form those that are not. This is usually performed via experiments that are conducted using *Microarray technology*. A microarray is a tool for analyzing gene expression that consists of a small membrane or glass slide containing thousands of spots, each containing a single-stranded DNA sequence of one gene (also called cDNA, for complementary DNA) (Horak and Snyder, 2002). Hence, the thousands of spots of a microarray contain single DNA strands for thousands of genes. Each spot in the microarray enables the measurement of the expression level of the corresponding gene as will be explained in the example below.

Example 17. *Figure 6.2 shows an example of an experiment aiming at discovering*

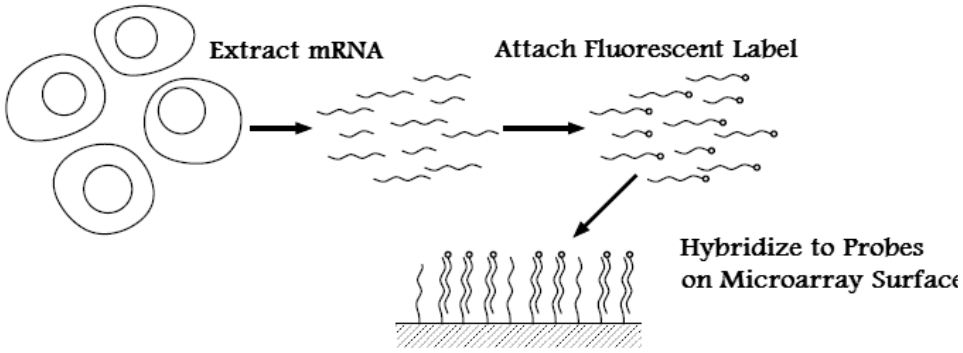


Figure 6.2: A Experiment for Measuring Gene Expression Levels

the expression profile (i.e. which genes are expressed and which are repressed) of a certain cell culture under a certain set of conditions (e.g. the absence of Oxygen). The experiment starts by obtaining a cellular sample from the culture after allowing it to be in the condition being examined (and as a result, given the organism the chance to adapt to the conditions by launching the expression of the genes required for the set of conditions). mRNA molecules corresponding to the expressed genes are then extracted from the cellular culture [2] *and is colored with a fluorescent for ease of detection. Then, the extracted mRNA is poured over the microarray containing the DNA of the respective cellular culture. If the DNA at a certain spot has a corresponding mRNA in the extracted cellular sample, it is bound to it* [3] *and the corresponding spot on the microarray is illuminated with the fluorescent that the mRMA was colored with. Hence, the simultaneous measurement of mRNA hybridization to cDNA strands enables their use as indications of the expression level of the corresponding gene Friedman (2004).*

[2]Details pertaining to the technical biological aspects will not be covered here. The interested reader may refer to (D'hæseleer, 2000) for an elaborate discussion.

[3]The process is called hybridization (D'hæseleer, 2000).

6.3 Networks Representing Genetic Interactions

The data obtained from microarray experiments is presented as collections of gene expression patterns that can be used to study the effects of certain treatments, diseases, and developmental stages on gene expression. As the presence of certain proteins and expression of certain genes can have an activating or repressing effect on the expression levels of other genes, making sense out of the data involves understanding the mechanisms controlling the expression and the causal relations that exist among genes (Ram et al., 2006) in which the expression of one or more genes causes the expression (or inhibition) of other genes. The general aim is to reach a functional understanding of the mechanisms governing the interactions among the cellular genetic components. Such understanding is linked to medical breakthroughs such as a general classification of clinical entities, e.g. different tumor types (Khan et al., 1998; Alizadeh et al., 2000), and the inception of novel techniques for the prevention and treatment of human diseases Hood et al. (2004) by changing the diagnosis process to focus on causes rather than symptoms (as many diseases are characterized by abnormal gene expression (Lubovac and Olsson, 2003)).

The large size of the data and the large number of interactions have motivated the use of graphical models to ease the process of capturing the complex relations among the genes and end products (Friedman, 2004). These models are generally termed *genetic networks* and are directed graphs in which every node represents a gene or a functional protein and every edge represents a regulation relationship Noveen et al. (1998). The network models how genes influence (through activation or inhibition) other genes in a complex web of interactions and are useful for analyzing genomes as they make explicit different types of interactions among genetic elements, which enables their usage as a road map for functional studies of the genomes of different organisms (Pisabarro et al., 2008).

Uncovering the topology of the network from the kind of data available is a challenging problem for which many techniques exist (Wessels et al., 2001). Specifically, the complexity of the task stems from the fact that not only the kind of data available is of high dimensionality and suffers from great noise (Friedman, 2004), but also

because the data, termed *microarray expression data*, provides the expression levels of a large number of genes (usually tens of thousands) at different but relatively few (usually a few dozens) experiments [4]. Hence, it is usually sparse, which makes uncovering causal relations more difficult.

The above issues of microarray expression data, along with the stochastic nature of the gene regulatory system have motivated Bayesian approaches for learning the structure of genetic networks (Murphy and Mian, 1999; Friedman, 2004; Zou and Conzen, 2005a; Zhang et al., 2007), which have been successfully used to learn large scale networks. For instance, Roland (2004) use learned Bayesian networks to uncover the mechanisms underlying the progressive genetic changes in the development of urothelial bladder cancer. However, approaches using BNs remain far from being efficient, specially given the data's large size (Chickering et al., 2004), which has motivated different other approaches and their conjunction with the Bayesian model. For example, Zainudin and Deris (2008) use k-means clustering to establish clusters of co-expressed genes and then learns the corresponding BNs of the discovered clusters.

6.4 Motivating the Use of Qualitative Probability

The relative success of Bayesian approaches to reconstruct genetic networks motivated this search for alternatives. On the one hand, (Dynamic) Bayesian Networks have been successfully used to detect conditional (in)dependence and time-delay relations that help uncover the structure of the gene expression networks (Liu and wing Kin Sung, 2006; Murphy and Mian, 1999). However, it is the qualitative nature of the information extracted from the data that brought about many of the benefits of the model, e.g. the type of the effect one gene having on another being activating or repressing. Hence, if one is to formulate a model that is specifically tailored to represent this information (in addition to other qualitative information the quantitative Bayesian approaches may not be able to capture), then more insight maybe obtained regarding the functional interactions governing the data.

[4]The reasons behind this difference between the number of genes and the number of experiments will be made clear in section 6.4

In light of the above, the next sections present a taxonomy of reasons to raise concerns with respect to using Bayesian techniques to discover the various interactions among the genes of the network. The taxonomy is the result of a comparative study of the literature that addresses fundamental concerns regarding the use of quantitative methods in general and Bayesian techniques in specific.

6.4.1 Robustness

1. Apart from being intrinsically noisy and difficult to analyze, microarray data do not correspond to the expression levels of protein end-products as microarray measures the amount of cRNA (or cDNA) hybridization to target mRNA transcripts (Friedman, 2004). Therefore, the numbers used as probabilities in current studies assume a direct correlation between mRNA and protein expression levels. This assumption however, is until now not well supported as current studies of this type of correlations have given varying results depending on 1) the technology used and 2) biological factors that remain poorly understood (Nie et al., 2006; Pascal et al., 2008; Guo et al., 2008), and as a result mRNA to protein expression correlations vary from being poor (Pascal et al., 2008) to moderately good (Guo et al., 2008). Moreover, each mRNA transcript may correspond to a number of functionally different proteins by undergoing a post-translation chemical modification, where the type and number of proteins resulting from a single mRNA depend on different cellular and chemical factors (D'hæseleer, 2000). This renders the interpretation of mRNA expression levels as probabilities a less than an ideal alternative.

   It is worth nothing that large-scale protein expression measurements are currently present but tend to be extremely noisy and lack a great deal of sensitivity and specificity (D'hæseleer, 2000). This however, makes protein expression data a good candidate for a qualitative formalism.

2. Analyzing microarray data for accuracy is difficult as the nature of and cost associated with current microarray technologies represent obstacles for repeating the experiments, which make statistical studies of a single measurement unattainable (Filkov et al., 2002). Therefore, it is difficult to rely on the numbers as they repre-

sent outcomes of a single, non-repeated experiment. This is especially important in determining how dependable the data is given the sparse nature of the resulting measurements as a single experiment yields very small number of samples (intervals under which the experiment is performed) especially when compared to the number of variables involved (the number of genes studied can reach tens of thousands for a single experiment) (D'hæseleer, 2000). Also, the dynamic nature of the expression process and the fact that it depends on factors that may not be known (D'hæseleer, 2000) makes the numbers more untrustworthy because it is currently not known whether the variables affecting the expression at different intervals are constant through the experiment (Pisabarro et al., 2008). Therefore, building a network that involves thousands of genes from dozens of examples of their expression levels does not assure the distinction between true gene-to-gene correlation and spurious ones (Friedman, 2004).

6.4.2 Possibility for Extension

Biological pathways are intricate in nature, and their discovery remains an ongoing challenge. Moreover, it is now accepted that in order to obtain a biological insight, it is viable to examine data from different view points in the aim of forming an integral examination of cellular interactions, e.g. gene expression and protein-protein interactions (Friedman, 2004). Given this, discovering a biological pathway may require information for which there does not exist quantitative information (even noisy information). Having a model that can do away with this type of information makes it more portable and more capable of dealing with the surprises that may encounter the discovery process (Iyenga and McGuire, 2007).

In order to perform such integral studies, we must first understand the biological principles that couple the measurements. In fact, it has been shown through stability analyses of gene expression models that describing models of gene networks requires information on both mRNA and protein levels (Hatzimanikatis and Lee, 1999). Hence, while awaiting the development of acceptable large protein chip technology, a qualitative model can present a viable alternative.

6.4.3 The Abundance of Qualitative Information

The uncertainty surrounding microarray data does not prevent the extraction of useful qualitative information that can be used to uncover the underlying genetic interaction and effectively reason about it to obtain biological insight. In fact, microarray data contains information pertaining the conditional independence among the genes in question, variable time delays and the combined effects of complexes of end products over genes. Although this information can be modeled correctly using Bayesian networks (as done in (Liu and wing Kin Sung, 2006)), there are other information of a strictly qualitative nature that can be extracted from the data. For instance, the sought-after relations among the genetic components are monotonic in nature, i.e. a gene or a set of genes, represented by variables, exhibit one type of influence on another set of genes of either stimulation or inhibition given a set of conditions. This has been further studied in bioinformatics and a set of gene expression *network motifs* (Milo et al., 2002) have been uncovered. These are the basic building blocks that define patterns of interconnections that recur in many different parts of a gene expression network at high frequencies.

Therefore, instead of constructing the joint distribution governing the conditional probabilities of genes given other genes, qualitative formalisms such as Qualitative Probabilistic Networks (Wellman, 1990a) are capable of explicitly modeling the influences underlying the conditional probabilities in a explicit way and can be used to either uncover the network model or perform inference on an existing one in a more efficient (see section 6.4.4) and robust (see section 6.4.1) manner.

6.4.4 Computational Efficiency

The dynamic nature of microarray time-series data requires the use of Dynamic Bayesian Networks as regular Bayesian Networks (Murphy and Mian, 1999). Despite the recent efforts to develop algorithms that are tailored to provide more efficient computations for uncovering genetic interactions (Murphy and Mian, 1999; Zou and Conzen, 2005b; Zhang et al., 2007), inference in Dynamic Bayesian Networks remains NP-hard (Chickering et al., 2004) as opposed to the polynomial-time arc-traversal

algorithm for inference in QPNs (Druzdzel and Henrion, 1993c).

6.5 The First Study: A Qualitative HMM for Gene Expression Data

In this section, qualitative HMM models are implemented to conduct an analysis of gene expression data. Traditional HMMs have been used to cluster time-series of gene expression data in the aim of finding the correlations among different genes (e.g. (Schliep et al., 2003), (Zeng and Garcia-Frias, 2006)). The qualitative HMM constructed in Chapter 5 can be applied to the same problem, and serve to create pre-clusters that the existing quantitative HMMs can use as a guide for a better analysis.

6.5.1 Aim

The aim of the experiments conducted in this section is to demonstrate the advantages of κ^{++} and $\mathfrak{z}$ over κ as measures of surprise by comparing the HMMs that use κ, κ^{++} and $\mathfrak{z}$ (denoted in this section by HMM_{κ}, $\text{HMM}_{\kappa^{++}}$ and $\text{HMM}_{\mathfrak{z}}$) and the classification capabilities of their associated qualitative forward algorithms.

Moreover, the experiment conducted here also aims producing results that can at least be used as a pre-clustering model that may be used to obtain useful insight about the data without having deal with the numerical aspects of the expression process. The qualitative HMMs can be useful in the following ways.

1. $\text{HMM}_{\kappa^{++}}$ and $\text{HMM}_{\mathfrak{z}}$ can be used as a pre-clustering medium that guides the initial stages of clustering using quantitative HMMs for very large time-series gene expression data.

2. $\text{HMM}_{\kappa^{++}}$ and $\text{HMM}_{\mathfrak{z}}$ will not suffer from the assumption that the actual numbers labeled in expression levels (which are the the levels of mRNA hybridization) are dealt with as the probability of expression of a certain gene. With

$\text{HMM}_{\kappa^{++}}$ and $\text{HMM}_{\mathfrak{z}}$, the expression level merely measures the degree of surprise of a gene being expressed; a less strict statement than that made by the probability-based assumption.

6.5.2 The Structure of the HMMs

Given a matrix M$< n, m >$ corresponding to a time-series microarray data set, construct a HMM to model the stochastic behavior of the matrix M as follows:

- Construct the set of states $S = \{s_1, ..., s_n\}$, where $\forall s_i \in S : s_i$ represents the hidden behavior of gene i $(1 \leqslant i \leqslant n)$, i.e. the behavior governing the time-series for gene i.
- Construct the set of observation variables $O = \{o_1, ..., o_m\}$, where $\forall o_t \in O : o_t$ represents the expression level of some gene at time t $(1 \leqslant t \leqslant m)$. Hence, the matrix $B = \{b_j(o_t), 1 \leqslant j \leqslant n\}$ represents the observed expression level of gene j at time t.

6.5.3 Data Set

For the purpose of the initial examination of the performance of $\text{HMM}_{\kappa^{++}}$ and $\text{HMM}_{\mathfrak{z}}$ embodied in this section, two data sets are used. The first is a small set of simulated time-series data describing the expression levels of 550 genes for a 5-step time series, whose usage aims at testing the models a small-sized data set. The second is the *Escherichia coli* time-series data set, for which the algorithms designed in Chapter 5 are evaluated by comparing the results obtained to those given in the literature.

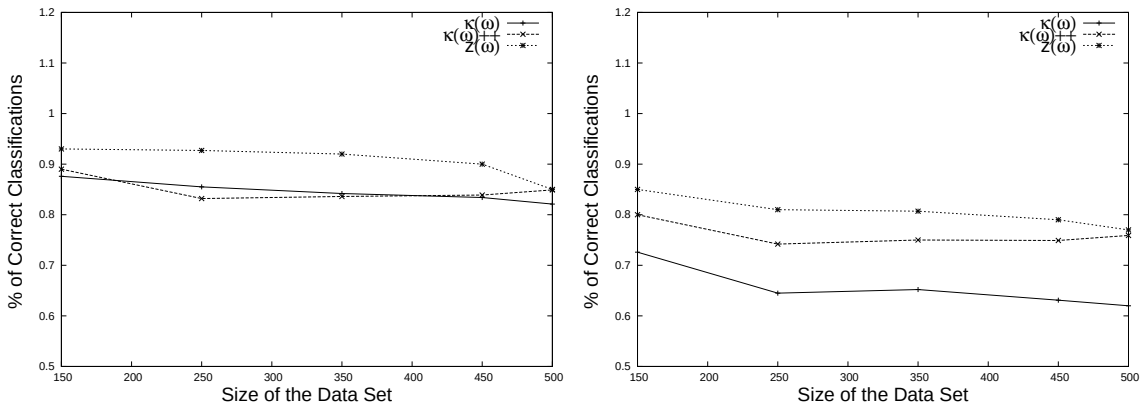
6.5.4 Obtaining $\text{HMM}_{\kappa^{++}}$ and $\text{HMM}_{\mathfrak{z}}$

Ideally, we would like the HMM to be trained with ranking values instead of numerical probabilities. This, however, requires a qualitative version of the learning algorithms, which is currently under development. Therefore, the HMM was trained with the

well-known Baum-Welch algorithm Rabiner (1989), which iteratively searches for the HMM parameters by maximizing the likelihood of the observations given the model, $P(O|\lambda)$. We use the Baum-Welch to obtain a HMM = (S,V,π,A,B) that uses regular probabilities. The κ^{++} and $\mathfrak{z}$ values of the corresponding qualitative HMMs are then obtained from the probability values of the π vector, the A and B matrices by mapping the probability values to κ^{++} and $\mathfrak{z}$ as given in Equations 4.3 and 4.5 of Chapter 4.

6.5.5 Experiment and Analysis

This experiment aims at testing the classification capability of the qualitative forward algorithm associated with HMM$_\vartheta$ using three different ranking functions κ, κ^{++} and $\mathfrak{z}$. Two separate experiments are performed using two data sets. The first is that of simulated while the second is the much larger *Escherichia coli* data set. The experiments are performed by running the qualitative algorithm on each data set using different values of ϵ for each ranking function (and hence varying the level of abstraction of the probability values).

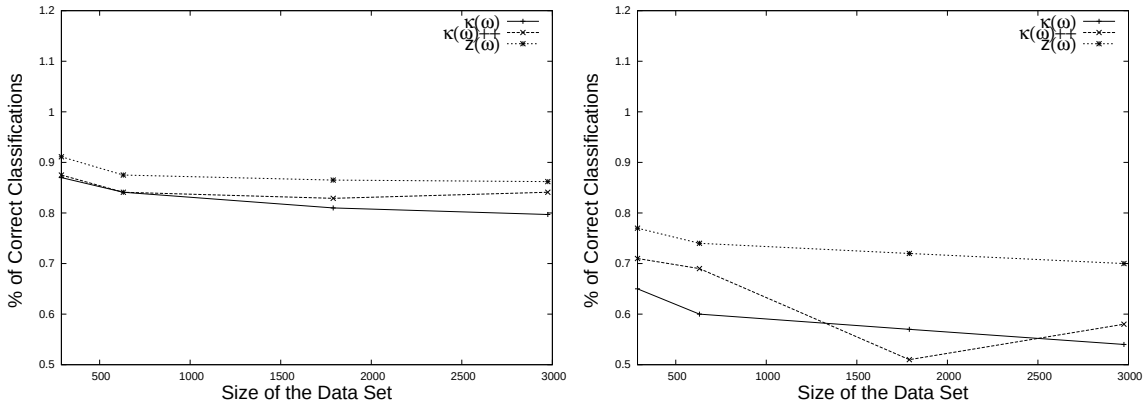


(a) Using an Infinitesimal ϵ (b) Using a Non-infinitesimal ϵ

Figure 6.3: HMM Classification Results Using an Artificial Data Set

The results are summarized in Figures 6.3 and 6.4. Both figures compare the classification of the different rank-based HMMs for infinitesimal and non-infinitesimal ϵ values using differently-sized portions of the data sets. In both figures, $\mathfrak{z}$ performs the best classification while κ averages as the worst. Also, unlike the other two

functions, the performance of κ^{++} fluctuates. This is an expected behavior given the non-monotonic behavior of κ^{++} for some of its negative range, which makes the performance of κ^{++} sensitive to the values given as rankings in the data set and causes the fluctuation depending on whether or not the rankings included in the data set are within the non-monotonic range.



(a) Using an Infinitesimal ϵ (b) Using a Non-infinitesimal ϵ

Figure 6.4: HMM Classification Results Using the Escherichia Coli Data Set

Moreover, the results can be analyzed in terms of the following parameters:

Infinitesimal vs. Non-infinitesimal ϵ:

When an infinitesimal ϵ is used (as in Figures 6.3a and 6.4a), κ's performance is at times comparable to that of κ^{++} (apart from the fluctuation). However, when a non-infinitesimal ϵ is used (as in Figures 6.3b and 6.4b), κ's performance drastically deteriorates as the number of correctly-classified cases decreases considerably. As for κ^{++}, using a non-infinitesimal ϵ seems to increase the fluctuation in its classification quality.

Synthetic vs. Real Data

A close examination of Figures 6.3a and 6.4a shows that the difference in performance between experiments where infinitesimal ϵ values are used and those where non-

infinitesimal ϵ values are used is much greater in real data than when using artificial data. This is evident even when conducting experiments on data sets of comparable sizes. For instance, the difference between the percentages of correctly-classified data of Figures 6.3a and 6.3b for HMM_3 when a data set of size 500 is used (approximately 88% in Figure 6.3a and %79 in Figure 6.3b) is much smaller than the difference between the percentages of correctly-classified data using HMM_3 for a data set of the same size as Figures 6.4a and 6.4b show (approximately %89 in Figure 6.4a and %75 in Figure 6.4b).

6.6 The Second Study: Discovering Genetic Network Motifs Using TQPNs

As Section 6.3 demonstrated, gene expression networks tend to be very complex with a large number of nodes and arcs connecting them. This has motivated studies that define simple patterns of interconnections between small groups of nodes. These patterns appear at high frequencies in naturally-occurring networks (including biological networks) and tend to increase in number monotonically as the size of the network increases. This is in contrast to synthetic, randomly-generated networks in which such patterns tend to sharply decrease in number as the size of the network grows (Shen-Orr et al., 2002). Hence, these patterns define subgraphs that occur at high frequencies in the network and which can serve as building blocks of the network. Such patterns have been termed *regulatory network motifs* (Shen-Orr et al., 2002; Milo et al., 2002) and have been shown to carry significant information about the network's overall organization and functionality (Hinman et al., 2003). The motifs present a way of uncovering the structural design principles of gene expression networks is by breaking down their complex wiring into basic components.

Shen-Orr et al. (2002) identify three motifs that occur frequently in gene expression networks that have been shown to appear at frequencies greater than 10 standard deviations greater than their mean number of appearances in randomized networks (Shen-Orr et al., 2002). These motifs are the *feed-forward* , in which a node X reg-

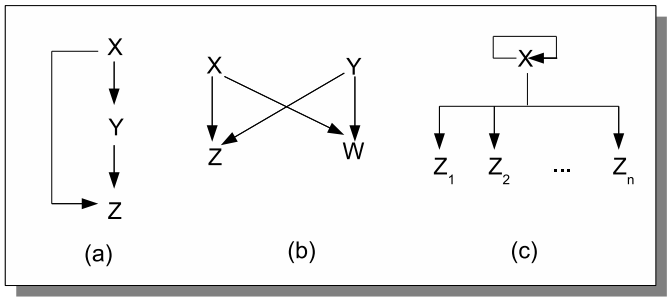


Figure 6.5: (a) The Feed-forward loop motifs (b) The Bi-fan motif (c) The Single-input Module Motif.

ulates another node Y such that they both regulate a third node Z, *bi-fan motifs*, in which two nodes concurrently regulate two other nodes, and *single-input module motifs* which define a set of nodes under the control of the same type of regulation (positive or negative) of one node, and are shown in Figure 6.5.

The study presented in this section can be summarized as using TQPN influences defined in Chapter 5 for the modeling and discovery of genetic network motifs and testing the method for the identification of network motifs using time series gene expression data of *Saccharomyces Cerevisiae* (yeast). It consists of three experiments whose specific aims, descriptions and results are discussed in the respective sections.

6.6.1 Using Influences to Define Genetic Network Motifs

If one to represent the gene-to-gene interactions in an expression experiment using a TQPN, where each subgraph G_t, $1 \leqslant t \leqslant m$ represents a snapshot of the genetic interactions of the cell during time slot T_t modeled by a QPN, then $A_1^t, ..., A_n^t \in U^t$ represents the expression levels of all the genes involved at slot T_t. In this context, a qualitative influence naturally corresponds to a regulatory relation between two nodes (genes). As a result, defining the motifs given in Figure 6.5 is directly obtained from the construct of the TQPN [5] as given in Definitions 27 - 29 below.

Definition 27. ***Feed-forward loops*** *A feed-forward loop exists in a genetic network*

[5]The reader can refer to Section 5.2 of Chapter 5 for a rehash of the concepts of TQPN.

modeled by a TQPN defined over G iff for two subgraphs G_t and G_{t+1}[6]*:*

$$S^{\delta_1}(A_i^t, A_i^{t+1}, tr_1) \wedge S^{\delta_2}(A_i^t, A_i^{t+1}, tr_2), \text{ where } tr_1 \neq tr_2$$

Where $\delta_1, \delta_2 \in \{++, --, +, -, ?, 0\}$. The above definition states that a feed-forward loop exists on a variable (gene) A_i if it influences its own expression through two different trails (by stimulating different genes that will subsequently stimulate its expression). Bi-fans are similarly defined below.

Definition 28. ***Bi-fans***

A bi-fan among four genes $A_a^t, A_b^t, A_c^{t+1} and A_d^{t+1}$ exists in a genetic network modeled by a TQPN defined over G iff for two subgraphs G_t and G_{t+1}

$$S^{\delta_1}(A_a^t, A_c^{t+1}, 1) \wedge S^{\delta_2}(A_b^t, A_c^{t+1}, 1) \wedge S^{\delta_3}(A_a^t, A_d^{t+1}, 1) \wedge S^{\delta_4}(A_b^t, A_d^{t+1}, 1).$$

Where $\delta_1, \delta_2, \delta_3$ and $\delta_4 \in \{++, --, +, -, ?, 0\}$.

Definition 29. ***Single Input Module (SIM)***

A SIM motif of a gene X_t on n other genes $A_1^{t+1}, ..., A_n^{t+1}$ exists in a genetic network modeled by a TQPN defined over G iff for two subgraphs G_t and G_{t+1}

$$S^{\delta}(X_t, A_1^{t+1}, 1) \wedge .... \wedge S^{\delta}(X_t, A_n^{t+1}, 1)$$

Where $\delta \in \{++, --, +, -, ?, 0\}$.

6.6.2 The First Experiment: Uncovering the Network Motifs Using TQPNs

We conducted a set of experiments to verify the mapping between qualitative influences and the motifs formalized in definitions 27, 28 and 29. The data set used for the

[6]Note that only two time slots are sufficient for the definition of the loop as TQPNs naturally preserve the Markov property.

Table 6.1: Nodes = Number of Nodes (Genes) in the Run, Edge_{avg} = Average Number of Edges for 10 Runs of Networks of Size N Feed-forward = Average Number of Feed-forward Motifs for the 10 Runs, Bi-fan = Average Number of Bi-fan Motifs for the 10 Runs

Nodes	Edge_{avg}	Feed-forward	Bi-fan
85	154	16	209
185	372	18	430
285	518	21	825
385	698	29	1092
485	912	46	1437
585	997	52	1745

purpose is based on the YPD (Yeast protein database) (S2) and was obtained from the data set used in (Milo et al., 2002) and contains 1079 interactions of 688 genes describing the regulation relationships of the transcriptional regulatory network of *Saccharomyces Cerevisiae*. The data comprises of three columns representing regulating genes, regulated genes and the mode of regulation. Not only that the number of motifs detected by our influences matches those of (Milo et al., 2002), but also upon retesting the hypothesis with differently-sized subsets of the data set, the number of motifs discovered by our influences was found to monotonically increase with the size of the data (as expected in real biological networks) as Table 6.1 shows.

The latter finding was achieved by constructing six additional experiments each testing the hypothesis for a subset of the full data set having a specific size. Each experiment consisted of 10 runs, all of the same size (number of nodes) but differ in connectivity (number of arcs). The algorithm describing the mapping of Section 5.2.2 was tested on each of the 60 resulting runs and used to output the number of feed-forward loops and bi-fan motifs in each run. The results given in Table 6.1 visibly show the monotonic increase of the number of motifs with the number of nodes in the interaction data set.

6.6.3 The Second Experiment: Discovering the Overall Motif Sign Using Surprise-based QPNs

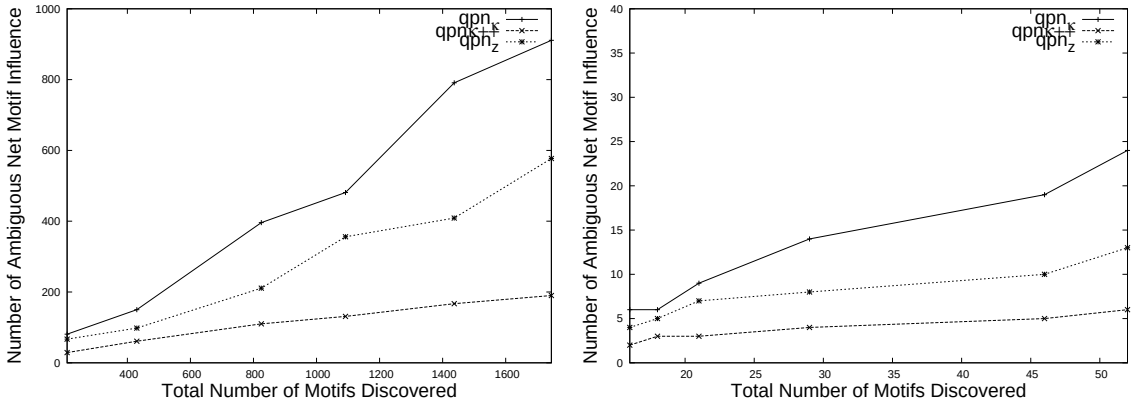
Having developed a method for discovering network motifs in the previous section, this section's experiment is tailored to evaluate the strengths of the ranking provided by κ, κ^{++} and $\mathfrak{z}$ in evaluating the net sign and strength of the motifs discovered. Hence, the experiment described here does not aim at devising a model of a practical use but is merely of an importance with respect to assessing the range of values assigned as rankings for each of the functions. Devising the experiment involved the following steps:

1. A QPN is implemented such that it uses a ranking function for the definition of its influences and uses the resulting strength assigned by the ranking for conflict resolution. The resulting three implementations are QPN_{κ}, $\text{QPN}_{\kappa^{++}}$ and $\text{QPN}_{\mathfrak{z}}$.

2. For each motif discovered, the signs and strengths of its arcs are mapped into a binary influence of the QPN under consideration.

3. For each motif, the net influence's magnitude and sign are evaluated.

The results of the experiment are shown in Figure 6.6, where Part 6.6a shows the number of ambiguous bi-fan motifs while Part 6.6b shows the number of feed-forward loop motifs for the motifs discovered and displayed in Table 6.1. In both cases, the QPN using κ^{++} as an indicator of influence strength leads in terms of minimizing the number of motifs evaluated as ambiguous. κ performs very poorly while $\mathfrak{z}$'s performance lies somewhere in between. The results reflect the advantages of the signed-integer values used in κ^{++} in increasing the range of values that the rank can hold, which leads to decreasing the possibility of a sign conflict whose causing influences have equal strengths. Also, the fact that $\text{QPN}_{\mathfrak{z}}$ had a better performance than QPN_{κ} supports the discussion given in Section 4.6.4 (and shown in Figure 4.5), which demonstrated that $\mathfrak{z}$ assumed a wider unsigned-integer range than κ, because it is this wider range that reduces the chance of obtaining equal magnitudes to two influences of conflicting signs.



(a) Bi-fan (b) Feed-forward

Figure 6.6: The Number of Ambiguous Motifs Using QPN$_\kappa$, QPN$_{\kappa^{++}}$ and QPN$_\jmath$

6.6.4 The Third Experiment: Reconstructing the Genetic Network from Expression Data

The second set of experiments were conducted to build qualitative influences between genes by examining their expression levels, map the relevant influences to network motifs and use them to guide the construction of a DBN. The aim of the experiment is to assess the accuracy of the approach in recovering the structure of the DBN from the expression data with the aid of the discovered motifs by comparing it to the unguided DBN approach of Zou and Conzen (2005a).

For this experiment, we used the *Saccharomyces Cerevisiae* time series data from Choo et al (Cho et al., 1998), which contains data for ten time points. The first step was to examine the microarray data to investigate the strength of the various regulatory interactions by assigning each pair of genes a correlation coefficient γ capturing the degree to which two genes are co-expressed. We used cut-off values of $\gamma_+ \geqslant 1.2$ for a positive regulation and $\gamma_- \leqslant 0.7$ for a negative regulation to separate possible direct regulation from spurious interactions and used an approach similar to that of (Zou and Conzen, 2005a) to identify potential regulators and regulees. The cut-off values were chosen to match those of (Zou and Conzen, 2005a) for a controlled experiment.

We then designed an algorithm that reads through the collected pairs and their normalized expression levels and builds a database of qualitative influences that are detected by examining the genes pair-wise. We constructed an $M \times M$ matrix of influences exhibited among the genes. Each cell in the matrix contains the information required regarding the possible influence of a sign, a rank calculated from the expression levels (we collected both κ^{++} and $\mathfrak{z}$ ranks). In our experiment, an unknown or a zero sign given in cell $m[i][j]$ designates a no correlation between the respective genes (at locations i and j). The mapping presented in Section 5.2.2 is used to construct the set of feed-forward loop motifs discovered in the data.

The set of motifs constructed is then used as prior knowledge to guide the construction of the yeast gene regulatory network using (Zou and Conzen, 2005a)'s method, referred to in this work as DBN$_{ZC}$. We evaluated the method in terms of accuracy of the reconstructed network. More specifically, the guidance provided by the motifs discovered increased the specificity [7] as Table 6.2 shows.

Table 6.2: The Result of comparing the analysis provided by DBN$_{ZC}$ with the same method guided with our qualitative network motifs for the yeast transcription dataset comprising of 116 genes. The only prior knowledge included is the knowledge of our qualitative motifs and nothing about the yeast cell cycle is given to test the hypothesis of an improved detection of regulator-target relations and a better construction of the target network. I = Identified Relationships, M = Misidentified Relationship, S = Specificity

Method	I	M	S
DBN$_{ZC}$	17	3	9.8%
DBN$_{ZC}$ + Qualitative Motifs	26	2	10.7%

[7] Specificity is the percentage of correctly predicted known gene relationships out of the total number of predicted gene relationships.

6.7 Chapter Summary

This chapter presented two studies and four experiments that examine the performance of the ranking functions in a bioinformatics setting.

The first study uses qualitative HMMs to classify gene expression data by mapping their expression levels to tentative rankings that are used to cluster the genes into functional groups. The experiment is performed for the three qualitative HMMs devised in Chapter 5, mainly HMM_{κ}, $\text{HMM}_{\kappa^{++}}$ and $\text{HMM}_{\mathfrak{z}}$ in an aim to evaluate the ranking capabilities of the three functions in this environment. The experiment shows that the ranking capability of $\mathfrak{z}$ precedes the other two functions in terms of performance while κ generally performs the worst.

The second study consists of three experiments all revolving around the idea of the identification of simple network motifs making up the building blocks of the gene-to-gene interactions in large gene regulatory networks. While the first experiment presented in Section 6.6.2 describes a procedure that successfully discovers the elementary network motifs by mapping them to TQPN influences (introduced in Chapter 5), the second experiment presented in Section 6.6.3 evaluates the semantical richness of the ranking functions by using them to find the overall motif sign and strength for the motifs discovered in Section 6.6.2. Since the performance of the function in this experiment relies on the width of the range assigned as ranks, κ^{++} performed best as it produced the minimum number of ambiguities compared to κ and $\mathfrak{z}$, while κ once again performed the worst due to its limited range (even compared to $\mathfrak{z}$). The last experiment uses the models developed to reconstruct genetic networks from expression data and uses both $\mathfrak{z}$ a κ^{++} as strength indicators. The method described in 6.6.4 has a performance that exceeds that of quantitative methods present in the literature.

Chapter 7

Conclusions, Loose Ends and Vision

A mathematician is a machine for turning coffee into theorems.

- Paul Erdõs

In this dissertation, we have presented surprise as an alternative notion for defining order-of-magnitude abstractions of probabilistic systems and used it to construct several graphical models that use qualitative surprise instead of qualitative probability to perform plausible reasoning. The main contributions this work has resulted are the following:

1. Two surprise-based ranking functions κ^{++} and $\mathfrak{z}$ whose main features are improving the semantics and ranking capabilities of existing rank functions. The two functions succeed in achieving the aims to different degrees and but overall, they represent an improvement to existing ranking functions. The functions use order of magnitude abstractions of conditional propagation rules for revision and update and each have different semantics and range based on the numerical surprise measure that each function abstracts.

2. A qualitative probabilistic network (QPN) that uses κ^{++} and $\mathfrak{z}$ to define its influences instead of probabilities. The QPN model defines qualitative influences as having both a sign and a strength factor that are based on the semantics of the ranking function used to define the QPN. Experimental results show that the richer semantics of κ^{++} and $\mathfrak{z}$ contribute to a better conflict-resolution mechanism for the QPN model proposed compared to using other ranking functions such as κ.

3. A temporal extension of the QPN model given above termed TQPNs. In addition to associating strength factors to its influences using the sign-magnitude pair, TQPNs use another level of influence strength measure based on the length of the trail through which the influence is exerted. These two levels of conflict resolution are implemented because the number of possible ambiguous signs increases in large networks, specially ones that model causality through time.

4. A qualitative Hidden Markov Model (HMM) that uses order of magnitude ranking functions instead of probabilities for the specification for its parameters. Along with the HMM come two algorithms that use the qualitative framework to perform recognition and prediction tasks.

5. A method for discovering elementary network motifs of gene regulatory networks by mapping the motifs structure to TQPN influences and constructing the latter from gene expression data.

6. An algorithm for reconstructing genetic regulatory network from microarray gene expression data that constructs tentative rankings from the expression data and maps them to TQPN influences and learning the regulatory network accordingly. The algorithm's performance has been empirically shown to be accurate and more efficient than existing quantitative bayesian methods.

7.1 Limitations and Loose Ends

1. As various discussions throughout the dissertation have shown, the semantics envisioned for ranking functions of having not only different levels of surprise associated

with events but also different levels of expectedness turned out not to be an easily achievable task. This is because a qualitative function that is based on a surprise index whose definition seems to promise this added richness suffers from some drawbacks including biased ranges for levels of surprise and expectedness as well as exhibiting non-monotonic behavior for some of its values. As a result, it becomes inevitable that the ranking offered by the function will not agree with that provided by probabilities with respect to the values for which the ranking function exhibits a non-monotonic behavior. Therefore, instead of increasing the type of values included in the ranking (of being surprising or expected), we have opted to devising a ranking function that increases the range of surprise measures. This made having a better and more versatile ranking that does not assign equal surprise classes to states that should belong to different classes a key point to the advantages. In addition to this, the new ranking functions offer domain independence as a key feature so that it becomes possible to compare the surprise associated with events that belong to different distributions.

2. Although this dissertation has demonstrated different ways of incorporating the proposed formalisms with other qualitative probability frameworks, combining the functionalities achieved here with quantitative probabilistic methods, which was initially part of the original vision, has not yet been studied. This aspect of the original plan is left as part of the future research and is discussed in the next section.

7.2 Vision for Future Directions

This section outlines some of the questions while developing the dissertation. We discuss them briefly here as possible directions for future research.

7.2.1 Collaboration with Quantitative Formalisms

Although qualitative formalisms can present a good alternative when efficient and less data-intensive methods are required for reasoning, the real power lies not in the use of qualitative and quantitative methods independently, but it is when they are

used in collaboration to achieve a certain goal by using probabilistic methods where appropriate and substituting them with qualitative approaches when the task at hand requires so. Identifying the conditions under which each formalism is used and the steps required to go from one level of abstraction to the other (or subsequently, to no abstraction) is currently an open problem.

7.2.2 Qualitative Entropy

The quantitative surprise measure used for bases of formulating the new ranking functions are closely related to the notion of information gain. This gives rise to the interesting question: is it possible to define the notion of qualitative entropy?

7.2.3 Better Regulation

We are currently working on the realization of a model for completely reconstructing gene regulatory networks using TQNs. We are at the stage of incorporating time lags into the model and testing the hypothesis of 'the full specification of conditional probabilities is not necessary to reconstruct the regulatory relations in a gene regulatory network and only a subset of the quantitative data available is required. Because TQPNs deploy arc-based reasoning, they are expected to be much more efficient than their quantitative equivalents.

Bibliography

Ernest Adams. The Logic of Conditionals. *Synthese Library*, 86, 1975.

Ash Alizadeh, Michael Eisen, Eric Davis, Chi Ma, Izidore Lossos, Andreas Rosenwald, Jennifer Boldrick, hajeer Sabet, Truc Tran, Xin Yu, John Powell, Liming Yang, Gerald Marti, Troy Moore, James Hudson, Lisheng Lu, David Lewis, Robert Tibshirani, Gaven Sherlock, Wing Chan, Timothy Greiner, Dennis Weisenburger, James Armitage, Roger Warnke, Ronald Levy, Wyndham Wilson, Michael Grever, John Byrd, David Botstein, Patrick Brown, and Louis Staudt. Distinct Types of Diffuse Large B-cell Lymphoma Identified by Gene Expression Profiling. *Nature*, 403(6769):503–511, 2000.

Chris Bailey-Kellog and Feng Zhao. Qualitative Spatial Reasoning: Extracting and Reasoning with Spatial Aggregates. *AI Magazine*, 24(4):47–60, 2003.

Yoseph Barash and Nir Friedman. Context-Specific Bayesian Clustering for Gene Expression Data. *Journal of Computational Biology*, 9(2):169–191, 2002.

Vijay Bawa. Optimal Rules for Ordering Uncertain Prospects. *Journal of Financial Economics*, 2:95–121, 1975.

M.J. Bayarri and James Berger. Quantifying Surprise in the Data and Model Verification. *Bayesian Statistics*, 6:53–82, 1998.

M.J. Bayarri and J. Morales. Bayesian Measures of Surprise for Outlier Detection. *Journal of Statistical Planning and Inference*, 111:3–22, 2003.

Salem Benferhat, Didier Dubois, and Henri Prade. Argumentative Inference in Uncertain and Inconsistent Knowledge Bases. In *International Conference on Uncertainty in Artificial Intelligence*, pages 411–419, 1993.

Mark Beumer. Qualitative Probabilistic Networks in Medical Diagnosis. Technical Report 0477672, Universiteit van Amsterdam, 2006.

Janneke Bolt, Silja Renooij, and Linda Van Der Gaag. Upgrading Ambiguous Signs in QPNs. In *International Conference on Uncertainty in Artificial Intelligence*, pages 73–80, 2003a.

Janneke Bolt, Linda Van Der Gaag, and Silja Renooij. Introducing Situational Influences in QPNs. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 113–124, 2003b.

Craig Boutilier. A Unified Model of Qualitative Belief Change: a Dynamical Systems Perspective. *Artificial Intelligence*, 98:281–316, 1997.

Jack Breeze and Kenneth Fertig. Decision Making with Interval Influence Diagrams. In *International Conference on Uncertainty in Artificial Intelligence*, pages 467–478, 1991.

Wray Buntine. Operations for Learning with Graphical Models. *Journal of Artificial Intelligence Research*, 2:159–225, 1994.

Enrique Castillo, José Manuel Gutierrez, and Ali Hadi. A new Method for Efficient Symbolic Propagation in Bayesian Networks. *Networks*, 28:31–43, 1996.

Enrique Castillo, José Manuel Gutierrez, and Ali Hadi. Sensitivity Analysis in Discrete Bayesian Networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 27:412–423, 1997.

Hei Chan and Adnan Darwiche. When do Numbers Really Matter? *Journal of Artificial Intelligence Research*, 17:265–287, 2002.

Hei Chan and Adnan Darwiche. Sensitivity Analysis in Bayesian Networks: from Single to Multiple Parameters. In *International Conference on Uncertainty in Artificial Intelligence*, pages 67–75, 2004.

Tim Chard. Qualitative Probability Versus Quantitative Probability in Clinical Diagnosis: a Study Using a Computer Simulation. *Medical Decision Making*, 11(1): 38–41, 1991.

Peter Cheeseman. In Defense of Probability. In *International Joint Conference on Artificial Intelligence*, pages 1002–1009, 1985.

Peter Cheeseman. Discussion of the paper by Lauritzen and Spiegelhalter. *Journal of Royal Statistical Society*, B(50):203, 1988.

David Chickering, David Heckerman, and Christopher Meek. Large-Sample Learning of Bayesian Networks is NP-Hard. *The Journal of Machine Learning Research*, 5: 1287–1330, 2004.

Raymond Cho, Michael Cambell, Elizabeth Winzeler, Lars Steinmetz, Andrew Conway, Lisa Wodicka, Tyra Wolfsberg, Andrei Gabrielian, David Landsman, David Lockhart, and Ronald Davis. A Genome-wide Transcirptional Analysis of the Mitotic Cell Cycle. *Molecular Cells*, 2(1):65–73, 1998.

Gustave Choquet. Theory of Capacities. *Annals of of the Institute Forrier*, 5:131–295, 1953.

Veerle Coopé, Niels Peek, Jaap Ottenkamp, and Dik Habbema. Using Sensitivity Analysis for Efficient Quantification of a Belief Network. *Artificial Intelligence in Medicine*, 17:223–247, 1999.

Veerle Coopé, Habbema Dik, and Linda Van Der Gaag. Sensitivity Analysis: An Aid for Probability Elicitation. *Knowledge Engineering Review*, 15:1–18, 2000.

Gregory Cooper. The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks. *Artificial Intelligence*, 42:393–405, 1990.

Gregory Cooper and Edward Herskovits. a Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning*, 9:309–347, 1992.

Veerle Coupé and Linda Van Der Gaag. Properties of Sensitivity Analysis of Bayesian Belief Networks. *Annals of Mathematics and Artificial Intelligence*, 36(4):323–356, 2002.

Francis Crick. Central Dogma of Molecular Biology. *Nature*, 227:561–563, 1970.

Paul Dagum and Michael Luby. An Optimal Approximation Algorithm For Bayesian Inference. *Artificial Intelligence*, 93:1–27, 1993.

Bruce D'ambrosio. Parallelizing Probabilistic Inference: Some Early Explorations. In *International Conference on Uncertainty in Artificial Intelligence*, page 5966, 1992.

Bruce D'ambrosio. Symbolic Probabilistic Inference in Large BN2O Networks. In *International Conference on Uncertainty in Artificial Intelligence*, pages 128–135, 1994.

Bruce D'Ambrosio. Inference in Bayesian Networks. *AI Magazine*, 20(2):21–36, 1999.

Adnan Darwiche. *A Symbolic Generalization of Probability Theory*. PhD thesis, Computer Science Department, Stanford University, Palo Alto, 1992.

Adnan Darwiche. Argument Calculus and Networks. In *International Conference on Uncertainty in Artificial Intelligence*, pages 420–427, 1993.

Adnan Darwiche and Moisés Goldzmidt. On the Relation Between Kappa Calculus and Probabilistic Reasoning. In *International Conference on Uncertainty in Artificial Intelligence*, pages 145–153, 1994.

Thomas Dean and Keiji Kanazawa. A model for reasoning about persistence and causation. *Artificial Intelligence*, 93(102):1–27, 1989.

Rina Dechter. Bucket Elimination: A Unifying Framework for Probabilistic Inference. In *International Conference on Uncertainty in Artificial Intelligence*, pages 211–219, 1996.

Peter D'hæseleer. *Reconstructing Gene Networks from Large Scale Gene Expression Data*. PhD thesis, University of New Mexico, Albuquerque, 2000.

Francisco Diez. Distributed inference in bayesian networks. *Cybernetics and Systems*, 25(1):3961, 1994.

Francisco Diez. Local Conditioning in Bayesian Networks. *Artificial Intelligence*, 87: 1–20, 1996.

Marek Druzdzel. *Probabilistic Reasoning in Decision Support Systems: From Computation to Commonsense*. PhD thesis, Department of Engineering and Public Policy, Carnegie Mellon University, 1993.

Marek Druzdzel and Linda Van der Gaag. Elicitation of Probabilities for Belief Networks: Combining Qualitative and Quantitative Information. In *International Conference on Uncertainty in Artificial Intelligence*, pages 141–148, 1995.

Marek Druzdzel and Max Henrion. Interausal Reasoning iwth Uninstantiated Ancestor Nodes. In *International Conference on Uncertainty in Artificial Intelligence*, pages 317–325, 1993a.

Marek Druzdzel and Max Henrion. Belief Propagation in Qualitative Probabilistic Networks. In *International Workshop on Qualitative Reasoning and Decision Technologies*, pages 451–460, 1993b.

Marek Druzdzel and Max Henrion. Efficient Reasoning in Qualitative Probabilistic Networks. In *National Conference on Artificial Intelligence*, pages 548–553. AAAI, 1993c.

Marek Druzdzel and Linda Van Der Gaag. Building Probabilistic Networks: "Where Do the Numbers Come From?" Guest Editors Introduction. *IEEE Transactions on Knowledge and Data Engineering*, 12(4):481–486, 2000.

Didier Dubois. Possibility Theory and Statistical Reasoning. *Computational Statistics and Data Analysis*, 51(1):47–69, 2006.

Didier Dubois and Henri Prade. Fuzzy Logics and the Generalized Modus Ponens Revisited. *Cybernetics and Systems*, 15:293331, 1984.

Didier Dubois and Henri Prade. *Possibility Theory An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.

Didier Dubois and Henri Prade. The logical View of Conditioning and its Application to Possibility and Evidence Theories. *International Journal of Approximate Reasoning*, 4(23-46), 1990.

Didier Dubois and Henri Prade. Epistemic Entrenchment and Possibilistic Logic. *Artificial Intelligence*, 50:223–239, 1991.

Didier Dubois and Henri Prade. Belief Revision and Updates in Numerical Formalisms: An Overview, with new Results for the Possibilistic Framework. In *International Joint Conference on Artificial Intelligence*, pages 620–625, 1993.

Didier Dubois and Henri Prade. Possibility Theory, Probability Theory and Multiple-valued Logics: A Clarification. *Annals of Mathematics and Artificial Intelligence*, 32:35–66, 2001.

Didier Dubois and Henri Prade. A Survey of Belief Revision and Updating Rules in Various Uncertainty Models. *International Journal of Intelligent Systems*, 9: 61–100, 2007.

Didier Dubois, Henri Prade, Luìs Godó, and Ramon Lòpez de Mántaras. A Symbolic Approach to Reasoning with Linguistic Quantifiers. In *International Conference on Uncertainty in Artificial Intelligence*, pages 74 – 82, 1992.

Didier Dubois, Hung Nguyen, and Henri Prade. Fuzzy Sets and Probability: Misunderstandings, Bridges and Gaps. In *Fundamentals of Fuzzy Sets*, pages 343–438, 2000.

Vladimir Filkov, Steven Skiena, and Jii Zhi. Analysis Techniques for Microarray Time-Series Data. *Journal of Computational Biology*, 9:317–330, 2002.

John Fox, Paul Krause, and Simon Ambler. Arguments, Contradictions and Practical Reasoning. In *European Conference on Artificial Intelligence*, pages 623–627, 1992.

Maria Fox, Malik Ghallab, Guillaume Infantes, and Derek Long. Robot Introspection Through Learned Hidden Markov Models. *Artificial Intelligence*, 170(2):59–113, 2006.

Nir Friedman. Learning Belief Networks in the Presence of Missing Values and Hidden Variables. In *International Conference on Machine Learning*, pages 125–133, 1997.

Nir Friedman. Inferring Cellular Networks Using Probabilistic Graphical Models. *Science*, 303:799–805, 2004.

Nir Friedman, Kevin Murphy, and Sean Russel. Learning the Structure of Dynamic Probabilistic Networks. In *International Conference on Uncertainty in Artificial Intelligence*, 1998.

Lise Getoor, Nir Friedman, Daphne Koller, and Benjamin Taskar. Learning Probabilistic Models of Relational Structure. In *International Joint Conference on Artificial Intelligence*, pages 1300–1309, 1999.

Lise Getoor, Nir Friedman, Daphne Koller, and Benjamin Taskar. Learning Probabilistic Models of Link Structure. *Journal of Machine Learning Research*, 3:679–707, 2002.

Zoubin Ghahramani. Learning Dynamic Bayesian Networks. *Lecture Notes in Computer Science*, 1387:168–172, 1998.

Phan Giang and Prakash Shenoy. On Transformations Between Probability and Spohnian Disbelief Functions. In *International Conference on Uncertainty in Artificial Intelligence*, pages 236–244, 1999.

Gerla Giangiacomo. *Fuzzy Logic: Mathematical Tools for Approximate Reasoning*. Kluwer Academic Publishers, 2001.

Moisés Goldszmidt. Qualitative Probabilities: A Normative Framework for Commonsense Reasoning. Technical report, University of California, 1992.

Moisés Goldszmidt. Fast Belief Update Using Order-of-Magnitude Probabilities. In *International Conference on Uncertainty in Artificial Intelligence*, pages 99–107, 1995.

Moisés Goldszmidt and Judea Pearl. Reasoning with Qualitative Probabilities Can be Tractable. In *International Conference on Uncertainty in Artificial Intelligence*, pages 112–120, 1992.

Moisés Goldszmidt and Judea Pearl. Qualitative Probabilities for Default Reasoning, Belief Revision, and Causal Modeling. *Artificial Intelligence*, 84:145–151, 1996.

Yanfang Guo, Peng Xiao, Shufeng Lei, Feiyan Deng, Gary Xiao, Yaozhong Liu, xiangding Chen, Liming Li, Shan Wu, Yan Chen, Hui Jiang, Lijun Tan, Jingyun Xie, Xuezhen Zhu, Songping Liang, and Hongwen Deng. How is mRNA Expression Predictive for Protein Expression: A Correlation Study on Human Circulating Monocytes. *Acta Biochimica et Biophysica Sinica*, 40(5):426–436, 2008.

Haipeng Guo and William Hsu. A Survey of Algorithms for Real-Time Bayesian Network Inference. In *Joint AAAI-02/KDD-02/UAI-02 Workshop on Real-Time Decision Support and Diagnosis Systems*, 2002.

David Hand. Protection or Privacy? Data Mining and Personal Data. In *Proceedings of the 10th Pacific-Asia Conference, Advances in Knowledge Discovery and Data Mining (PAKDD)*, pages 1–10, 2006.

Vassily Hatzimanikatis and Kelvin Lee. Dynamic Analysis of Gene Networks Requires Both mRNA and Protein Expression Information. *Metabolic Engineering*, 1:1–7, 1999.

Patrick Hayes. The Naïve Physics Manifesto. In *Expert Systems in the Microelectronic Age*, pages 242–270, 1978.

Patrick Hayes. Naïve Physics I: Ontology for Liquids. *Formal Theories of the Commonsense World*, a:71–107, 1985a.

Patrick Hayes. The Second Naïve Physics Manifesto. *Formal Theories of the Commonsense World*, b:1–36, 1985b.

David Heckerman. A Tutorial on Learning Bayesian Networks. Technical Report MSR-TR-95-06, Microsoft Reearch, Advanced Technology Division, 1995.

Max Henrion and Marek Druzdzel. Qualitative Propagation and Scenario-based Approaches to Explanation of Probabilistic Reasoning. In *International Conference on Uncertainty in Artificial Intelligence*, pages 17–32, 1991.

Max Henrion, Gary Provan, Bart Del Favero, and Graig Sanders. An Experimental Comparison of Numerical and Qualtiative Probabilistic Reasoning. In *International Conference on Uncertainty in Artificial Intelligence*, pages 319–326, 1994.

Max Henrion, Mark Pradhan, Bart Del Favero, Kevin Huang, and Ganesh Prvan. Why is Diagnosis Using Belief Networks Insensitive to Imprecision in Probabilities. In *International Conference on Uncertainty in Artificial Intelligence*, pages 307–314, 1996.

Veronica Hinman, Albert Nguyen, Andrew Cameron, and Eric Davidson. Developmental Gene Regulatory Network Architecture Across 500 Million Years of Echinoderm Evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23):13356–61, 2003.

Leroy Hood, James Heath, Michael Phelps, and Biaoyang Lin. Systems Biology and New Technologies Enable Predictive and Preventive Meidcine. *Science*, 306:640–643, 2004.

Christine Horak and Michael Snyder. Global Analysis of Gene Expression in Yeast. *Functional Integrative Genomics*, 2:171–180, 2002.

Xuedong Huang, Alex Acero, and Hsiao-Wuen Hon. *Spoken Language Processing: A Guide to Theory, Algorithm and System Development.* Prentice Hall PTR, 2001. ISBN 0130226165.

Franz Huber. Ranking Functions and Rankings on Languages. *Artificial Intelligence*, 170:462–471, 2006.

Daniel Hunter. Parallel Belief Revision. In *International Conference on Uncertainty in Artificial Intelligence*, pages 241–251, 1990.

Daniel Hunter. Non-monotonic Reasoning and the Revrsibility of Belief Change. In *International Conference on Uncertainty in Artificial Intelligence*, pages 159–164, 1991.

Lawrence Hunter. Life and Its Molecules, A Brief Introduction. *AI Magazine*, 25(1): 9–22, 2004.

Zina Ibrahim, Ahmed Tawfik, and Alioune Ngom. A Qualitative Hidden Markov Model for Spatio-temporal Reasoning. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 707–718, 2007.

Zina Ibrahim, Ahmed Tawfik, and Alioune Ngom. Qualitative Hidden Markov Models for Classifying Gene Expression Data. In *Twenty-eighth SGAI International Conference on Artificial Intelligence*, 2008.

Zina Ibrahim, Alioune Ngom, and Ahmed Tawfik. Qualitative Motif Detection in Gene Regulatory Networks. In *IEEE International Conference on Bioinformatics and Biomedicine*, pages 124–129, 2009a.

Zina Ibrahim, Ahmed Tawfik, and Alioune Ngom. Surprise-Based Qualitative Probabilistic Networks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 228–239, 2009b.

Zina Ibrahim, Ahmed Tawfik, and Alioune Ngom. A Surprise-based Qualitative Probability Calculus. In *The Twenty Second Conference of the Florida Artificial Intelligence Research Society*, pages 573–574, 2009c.

Zina Ibrahim, Ahmed Tawfik, and Alioune Ngom. Towards Qualitative Outlier Detection. In *The Twenty Third Conference of the Florida Artificial Intelligence Research Society*, page To Appear, 2010.

Sriram Iyenga and Mary McGuire. Imprecise and Qualitative Probability in Systems Biology. In *International Conference on Systems Biology*, 2007.

Finn Jensen. Cautious Propagation in Bayesian Networks. In *International Conference on Uncertainty in Artificial Intelligence*, pages 323–328, 1995.

Finn Jensen. *Bayesian Networks and Decision Graphs*. Springer-Verlag, 2001.

Finn Jensen, Steffen Lauritzen, and Kristian Olesen. Bayesian Updating in Recursive Graphical Models by Local Computations. *Computational Statistics Quarterly*, 4: 269–282, 1990.

Finn Jensen, Søren Aldenryd, and Klaus Jensen. Sensitivity Analysis in Bayesian Networks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 243–250, 1995.

Daniel Kahneman, Paul Slovic, and Amos Tversky. *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982.

Robel Kahsay, Guang Gao, and Li Liao. An improved Hidden Markov Model for Transmembrane Protein Detection and Topology Prediction and its Applications to Complete Genomes. *Bioinformatics*, 21(9):853–1858, 2005.

Richard Karp, Roland Stoughton, and Ka Yee Yeung. Algorithms for Choosing Differential Gene Expression Experiments. In *International Conference on Computational Molecular Biology*, pages 208–217, 1999.

Jeroen Keppens. Towards Qualtative Approaches to Bayesian Evidential Reasoning. In *IEEE International Conference on Automation and Logistics*, pages 17 – 25, 2007.

Javed Khan, Richard Simon, Michael Bittner, Yidong Chen, Stephen Leighton, Thomas Pohida, Paul Smith, Yuan Jiang, Gerald Gooden, Jeffrey Trent, and Paul Meltzer. Gene Expression Profiling of Aveolar Rhabdomyosarcoma with cDNA Microarrays. *Cancer Research*, 58:5009–5013, 1998.

Jin Kim and Judea Pearl. A Computational Model for Causal and Diagnostic Reasoning in Inference Systems. In *International Joint Conference on Artificial Intelligence*, pages 190–193, 1983.

Uffe Kjærulff. dHugin: A Computational System for Dynamic Time-Sliced Bayesian Networks. *International Journal of Forecasting, Special Issue on Probability Forecasting*, 11(1):89–111, 1995.

Ladislav Kohout and Brian Gaines. The Logic of Protection. In *GI*, pages 736–751, 1975.

Kevin Korb and Ann Nicholson. *Bayesian Artificial Intelligence*. CRC Press, 2003. ISBN 1584883871.

Alexander Kozlov and Jaswinder Pal Singh. A parallel Lauritzen-Spiegelhalter Algorithm for Probabilistic Inference. In *the ACM/IEEE conference on Supercomputing*, pages 320–329, 1994.

Alexander Kozlov and Jaswinder Pal Singh. Parallel Implementations of Probabilistic Inference. *Computer*, 29(12):3340, 1996.

Johan Kwisthout and Linda Van Der Gaag. The Computational Complexity of Sensitivity Analysis and Parameter Tuning. In *International Conference on Uncertainty in Artificial Intelligence*, pages 349–356, 2008.

Henry Kyberg. Higher Order Probabilities. In *International Conference on Uncertainty in Artificial Intelligence*, pages 15–22, 1989.

Henry Kyburg. Probability and the Logic of Rational Belief. Technical report, Wesleyan Universit Press, 1961.

Kathryn Blackmond Laskey. Sensitivity Analysis for Probability Assessments in Bayesin Networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 25: 901–909, 1995.

Steffen Lauritzen and David Spiegelhalter. Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems. *Journal of Royal Statistical Society*, 50B(2):157–224, 1988.

Xiaolin Li, Xiangdong He, and Senmiao Yuan. Learning Bayesian Networks Structures from Incomplete Data: An Efficient Approach Based on Extended Evolutionary Programming. In *Advances in Knowledge Discovery and Data Mining*, pages 474–479, 2005.

Zhaoyu Li and Bruce D'ambriosio. Efficient Inference in Bayes Networks As a Combinatorial Optimization Problem. *International Journal of Approximate Reasoning*, 11(1):55–81, 1994.

Chaolin Liu and Michael Wellman. Incremental Tradeoff Resolution in Qualitative Probabilistic Networks. In *International Conference on Uncertainty in Artificial Intelligence*, pages 338–345, 1998.

Tie-Fei Liu and wing Kin Sung. Learning Gene Network Using Conditional Dependence. In *The IEEE International Conference on Tools with Artificial Intelligence*, pages 800–804, 2006.

Brian Lovell. Hidden Markov Models for Spatio-Temporal Pattern Recognition and Image Segmentation. In *International Conference on Advances in Pattern Recognition*, volume 1, pages 60–65, 2003.

Zelmina Lubovac and Björn Olsson. Towards Reverse Engineering of Genetic Regulatory Networks. Technical report, University of Skövde, 2003.

Peter Lucas. Bayesian Network Modelling through Qualitative Patterns. *Artificial Intelligence*, 163(2):233–263, 2005. ISSN 0004-3702.

Anders Madsen and Finn Jensen. Parallelization of Inference in Bayesian Networks. Technical report, Department of Computer Science, Aalborg University, 1999.

Subramani Mani, Marco Valtorta, and Suzanne McDermott. Building Bayesian Network Models in Medicine: The MENTOR Experience. *Applied Intelligence*, 22(2): 93–108, 2005.

John McCarthy. Programs with Common Sense. In *The Teddington Conference on the Mechanization of Thought Processes*, pages 75–91, 1959.

Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827, 2002.

Kevin Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California Berkeley, 2002.

Kevin Murphy and Sean Mian. Modelling Gene Expression Data using Dynamic Bayesian Networks. Technical Report, University of California, Berkely, 1999.

Iftach Nachman, Gal Elidan, and Nir Friedman. "Ideal Parent" Structure Learning for Continuous Variable Networks. In *International Conference on Uncertainty in Artificial Intelligence*, pages 400–409, 2004.

Vasanth Namasivayam, Animesh Pathak, and Viktor Prasanna. Scalable Parallel Implementation of Bayesian Network to Junction Tree Conversion for Exact Inference. In *International Symposium on Computer Architecture and High Performance Computing*, 2006.

Eric Neufeld. Defaults and Probabilities; Extensions and Coherence. In *International Conference on Principles of Knowledge Representation and Reasoning*, pages 312–323, 1989.

Eric Neufeld. A Probabilistic Commonsense Reasoner. *International Journal of Intelligent Systems*, 5:565–594, 1990.

John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, third edition, 1953.

G. Ng and K. Ong. Using a Qualitative Probabilistic Network to Explain Diagnostic Reasoning in an Expert System for Chest Pain Diagnosis. In *Computers in Cardiology*, pages 569–572, 2000.

Lei Nie, Gang Wu, and Weiwen Zhang. Correlation of mRNA Expression and Protein Abundance Affected by Multiple Sequence Features Related to Translational Efficiency in Desulfovibrio vulgaris: A Quantitative Analysis. *Genetics*, 174:2229–2243, 2006.

David Nikovski. Constructing Bayesian Networks for Medical Diagnosis from Incomplete and Partially Correct Statistics. *IEEE Transactions on Knowledge and Data Engineering*, 12(4):509–516, 2000.

Alexande Noveen, Volker Hartenstein, and Cheng-Ming Chuong. *Gene Networks and Supernetworks: Evolutionary Conserved Gene Interactions*, pages 371–391. Landes Bioscience, 1998.

Simon Parsons. *Qualitative Methods for Reasoning Under Uncertainty*. MIT Press, 2001.

Simon Parsons. Qualitative Probability and Order of Magnitude Reasoning. *International Jounal on Uncertainty and Fuzziness in Knowledge-Base Systems*, 11(3): 373–390, 2003.

Simon Parsons and Ebrahim Mamdani. On Reasoning in Networks with Qualitative Uncertainty. In *International Conference on Uncertainty in Artificial Intelligence*, pages 435–442, 1993.

Laura Pascal, Lawrence True, David Campbell, Eric Deutsch, Michael Risk, Ilsa Coleman, Lillian Eichner, Peter Nelson, and Alvin Liu. Correlation of mRNA and Protein Levels: Cell Type-specific Gene Expression of Cluster Designation Antigens in the Prostate. *BMC Genomics*, 23(9):246–258, 2008.

Judea Pearl. Fusion, Propagation, and Structuring in Belief Networks. *Artificial Intelligence*, 29(3):241–288, 1986.

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

Judea Pearl. Bayesian Decision Methods. *Readings in Uncertain Reasoning*, pages 345–352, 1990.

Judea Pearl. From Conditional Oughts to Qualitative Decision Theory. In *International Conference on Uncertainty in Artificial Intelligence*, pages 12–22, 1993.

Judea Pearl and Thomas Verma. A Theory of Inferred Causation. In *International Conference on Principles of Knowledge Representation and Reasoning*, pages 441–452, 1991.

Judea Pearl, Dan Geiger, and Thomas Verma. Conditional Independence and its Representation. *Kybernetika*, 25(2), 1989.

David Pennock. Logarithmic Time Parallel Bayesian Inference. In *International Conference on Uncertainty in Artificial Intelligence*, pages 431–438, 1998.

Antonio Pisabarro, Gumer Pérez, José Luis Lavín, and Lucía Ramírez. Genetic Networks for the Functional Study of Genomes. *Briefings in Functional Genomics and Proteomic*, 7(4):249–263, 2008.

Lawrence Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–289, 1989.

Olivier Raiman. Order of Magnitude Reasoning. In *National Conference on Artificial Intelligence*, pages 100–104, 1986.

Ramesh Ram, Madhu Chetty, and Trevor Dix. Causal Modeling of Gene Regulatory Network. In *IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology*, pages 1–8, 2006.

Vahid Ramezani and Steven Marcus. Estimation of Hidden Markov Models: Risk-Sensitive Filter Banks and Qualitative Analysis of their Sample Paths. *IEEE Transactios on Automatic Control*, 47(12):1000–2009, 2002.

Silja Renooij. *Qualitative Approaches to Quantifying Probabilistic Networks*. PhD thesis, Proefschrift Universiteit Utrecht, 2001.

Silja Renooij and Linda Van Der Gaag. Exploiting Non-monotonic Influences in Qualitative Belief Networks. In *International Conference on Information Processing and Management of Uncertainty*, pages 1285–1290, 2000.

Silja Renooij and Linda Van Der Gaag. Propgation of Multiple Observations in QPNs Revisited. In *European Conference on Artificial Intelligence*, pages 665–669, 2002.

Silja Renooij and Linda Van Der Gaag. Evidence-Invariant Sensitivity Bounds. In *International Conference on Uncertainty in Artificial Intelligence*, pages 479–486, 2004.

Silja Renooij and Linda Van Der Gaag. Enhanced Qualitative Probabilistic Networks for Resolving Trade-offs. *Artificial Intelligence*, 172(12-13):1470–1494, 2008.

Silja Renooij, Linda Van Der Gaag, Simon Parsons, and Shaw Green. Pivotal Pruning of Tradeoffs in QPNs. In *International Conference on Uncertainty in Artificial Intelligence*, pages 515 – 522, 2000.

Silja Renooij, Linda Van Der Gaagand, and Simon Parsons. Context-specific Sign Propagation in Qualitative Probabilistic Networks. *Artificial Intelligence*, 140:207–230, 2002.

Silja Renooij, Simon Parsons, and Pauline Pardieck. Using Kappas as Indicators of Strength in Qualitative Probabilistic Networks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 87–99, 2003.

Conrad Christianand Erfle Holgerand Warnat Patrickand Daigle Nathalieand Lrch Thomasand Ellenberg Janand Pepperkok Rainerand Eils Roland. Automatic Identification of Subcellular Phenotypes on Human Cell Arrays. *Genome Research*, 14 (6):1130–6, 2004.

A. Rosti and M. Gales. Factor Analysed Hidden Markov Models for Speech Recognition. Technical Report 453, Cambridge University Engineering Department, 2003.

Stuart Russell and Peter Norvig. *Artificial Intelligence a Modern Approach*. Prentice Hall, Second Edition edition, 2003.

Leonard Savage. *The Foundations of Statistics*. Dover, Second edition, 1972.

Alexander Schliep, Alexander Schönhuth, and Christine Steinhoff. Using Hidden Markov Models to Analyse Gene Expression Time Course Data. *Bioinformatics*, 19(1):i255–i263, 2003.

Ross Shachter. Evaluating Influence Diagrams. *Operations Research*, 34(6):871–882, 1986.

Ross Shachter. Bayes-ball: The Rationall Pastime for Determining Irrelevance and Requisit Information in Belief Networks and Influence Diagrams. In *International Conference on Uncertainty in Artificial Intelligence*, pages 48–487, 1988.

Ross Shachter. An Ordered Examination of Influence Diagrams. *Networks*, 20(5): 535–563, 1990.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1988.

Glenn Shafer and Judea Pearl. *Readings in Uncertain Reasoning*. Morgan Kaufmann Publishers Inc., 1990. ISBN 1-55860-125-2.

Glenn Shafer and Prakash Shenoy. Probability Propagation. *Annals of mathematics and Artificial Intelligence*, 2:283–306, 1990.

Shai Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network Motifs in the Transcriptional Regulation Network of Escherichia Coli. *Nature Genetics*, 31:64–68, 2002.

Prakash Shenoy. On Spohn's Theory of Epistemic Beliefs. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems: Uncertainty in Knowledge Bases*, pages 2–13, 1990.

Prakash Shenoy. Uncertain Reasoning. Technical Report BUS934, School of Business, University of Kansas, 1998.

Solomon Shimony and Carmel Domshlak. Complexity of Probabilistic Reasoning in Directed-path Singly-connected Bayes Networks. *Artificial Intelligence*, 151(1-2): 213–225, 2003.

Padhraic Smyth. Belief Networks, Hidden Markov Models, and Markov Random Fields: A unifying view. *Pattern Recognition Letters*, 18(11-13):1261–1268, 1997.

Eugene Van Someren, Lodewyk Wessels, Eric Backer, and Marcel Reinders. Genetic Network Modeling. *Pharmacogenomics*, 3(4):507–25, 2002.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Predication and Search*. Springer Verlag, 1993.

Wolfgang Spohn. A General Non-Probabilistic Theory of Inductive Reasoning. In *International Conference on Uncertainty in Artificial Intelligence*, pages 149–158, 1988a.

Wolfgang Spohn. Ordinal Conditional Functions. A Dynamic Theory of Epistemic States. *Causation in Decision, Belief Change, and Statistics*, 2:105–134, 1988b.

Valentina Tamma and Simon Parsons. Argumentation and Qualitative Probabilistic Reasoning Using the Kappa Calculus. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 680–691, 2001.

Ahmed Tawfik. *Changing Times: An Investigation in Probabilistic Temporal Reasoning*. PhD thesis, Deaprtment of Computer Science, University of Saskatchewan, 1997.

Ahmed Tawfik and Eric Neufeld. Surprises in Probabilistic Reasoning. In *Florida Artificial Intelligence Symposium*, pages 45–50, 1996.

Thomas Verma and Judea Pearl. An Algorithm for D eciding if a Set of Observed Independencies Has a Causal Explanation. In *International Conference on Uncertainty in Artificial Intelligence*, pages 323–330, 1992.

William Weaver. Probability, Rarity, Interest and Surprise. *Scientific Monthly*, 67: 390–392, 1948.

Kurt Weichselberger and Sigrid Pöhlmann. *A Methodology for Uncertainty in Knowledge-Based Systems*. Springer Verlag, 1990.

Michael Wellman. Fundamental Concepts of Qualitative Probabilistic Networks. *Artificial Intelligence*, 44:257–303, 1990a.

Michael Wellman. Graphical Inference in Qualitative Probabilistic Networks. *Networks*, 30:687–701, 1990b.

Michael Wellman. Some Varieties of Qualitative Probability. In *International Conference on Information Processing and the Management of Uncertainty*, pages 437–442, 1994.

Michael Wellman. Incremental Tradeoff Resolution in Qualitative Probabilistic Networks. In *International Conference on Uncertainty in Artificial Intelligence*, pages 338–345, 1998.

Michael Wellman and Max Henrion. Qualitative Intercausal Relations, or Explaining "explaining away ". In *International Conference on Principles of Knowledge Representation and Reasoning*, pages 535–546, 1991.

Michael Wellman and Max Henrion. Explaining 'Explaining Away'. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):287–292, 1993.

Lodewyk Wessels, Eugene Van Someren, and Marcel Reinders. A Comparison of Genetic Network Models. In *Pacific Symposium on Biocomputing*, volume 6, pages 508–519, 2001.

Nick Wilson. An Order of Magnitude Calculus. In *International Conference on Uncertainty in Artificial Intelligence*, pages 548–555, 1995.

Man Wong and Yuan Guo. Discover bayesian Networks from Incomplete Data Using a Hybrid Evolutionary Algorithm. In *International Conference on Data Mining*, pages 1146–1150, 2006.

Dan Wu and Cory Butz. On the Complexity of Probabilistic Inference in Singly Connected Bayesian Networks. In *International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 581–590, 2005.

Yinglong Xia and Viktor Prasanna. Node level primitives for parallel exact inference. In *International Symposium on Computer Architecture and High Performance Computing*, pages 221–228, 2007.

Ronald Yager. A Foundation for a Theory of Possibility. *Journal of Cybernetics*, 10: 177–204, 1980.

Lotfi Zadeh. Fuzzy Sets as the Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978. (Reprinted in F uzzy Sets and Systems 100 (Supplement): 9-34, 1999.).

Suhaila Zainudin and Safaai Deris. Combining Clustering and Bayesian Network for Gene Network Inference. In *International Conference on Intelligent Systems Design and Applications*, pages 557–563, 2008.

Yujing Zeng and Javier Garcia-Frias. A Novel HMM-based Clustering Algorithm for the Analysis of Gene Expression Time-Course Data. *Computational Statistics & Data Analysis*, 50:2472–2494, 2006.

Nevin Lianwen Zhang and David Pool. Exploiting Causal Independence in Bayesian Network Inference. *Journal of Artificial Intelligence Research*, 5:301–328, 1996.

Yu Zhang, Zhidong Deng, Hongshan Jiang, and Peifa Jia. Inferring Gene Regulatory Networks from Multiple Data Sources Via a Dyanamic Bayesian Network with Structural EM. *Data Integration in the Life Sciences*, pages 204–214, 2007.

Min Zou and Suzanne Conzen. A New Dynamic Bayesian netowrk (DBN) Approach for Identifying Gene Regulatory networks from Time Course Microarray Data. *Bioinformatics*, 2(1):71–79, 2005a.

Min Zou and Suzanne Conzen. A New Dynamic Bayesian netowrk (DBN) Approach for Identifying Gene Regulatory networks from Time Course Microarray Data. *Bioinformatics*, 2(1):71–70, 2005b.

Geoffrey Zweig. *Speech Recognition with Dynamic Bayesian Networks*. PhD thesis, University of California, Berkeley, 1997.

Vitae Auctoris

Zina Ibrahim was born in 1980 in Baghdad, Iraq. She graduated from Sayfol International School in Kuala Lumpur, Malaysia in 1998. From there she went on to the University of Windsor where she obtained her B.Sc. in Computer Science with a Software Engineering focus in 2002 and M.Sc. in Computer Science in 2004. She is currently a candidate for the Doctoral degree in Computer Science at the University of Windsor and hopes to graduate in Winter 2010.