Quantum d-separation and quantum belief propagation

Robert R. Tucci
tucci@ar-tiste.com

December 14, 2020

Abstract

The goal of this paper is to generalize classical d-separation and classical Belief Propagation (BP) to the quantum realm. Classical d-separation is an essential ingredient of most of Judea Pearl's work. It is crucial to all 3 rungs of what Pearl calls the 3 rungs of Causation. So having a quantum version of d-separation and BP probably implies that most of Pearl's Bayesian networks work, including his theory of causality, can be translated in a straightforward manner to the quantum realm.

1 Introduction

Classical Bayesian networks (bnets), d-separation and belief propagation (BP), the main topics of this paper, were first proposed by Judea Pearl and collaborators. Pearl's results and their history have been amply discussed by him and some coauthors in the following 4 books: Refs.[2, 3, 4, 5]. Tucci has written an open source book entitled Bayesuvius (see Ref.[6]) about classical bnets. Bayesuvius will be cited frequently in this paper as a primary source of background information explained in the same style of notation as this paper.

Quantum Bayesian Networks (qbnets) were first proposed by Tucci in Ref.[12]. Ref.[12] deals with pure quantum states only. Tucci later generalized qbnets to mixed states in Ref.[11].

The goal of this paper is to generalize classical d-separation[1] and classical BP[2] to the quantum realm. Tucci has assumed or implied in previous work that d-separation and BP are valid for qbnets, but he has never explicitly proven this. This paper is intended to be a first step towards filling that gap.

Classical d-separation inspired Tucci to propose the definition of squashed entanglement (SE). d-separation and SE are very closely linked. SE was first proposed by Tucci in a series of 6 papers Refs.[13, 16, 9, 15, 7, 8] spanning the years 1999-2002. Starting around 2004 with Ref.[1] by Christiandl and Winters, other researchers recognized the importance of SE and began writing papers about it. A more complete history of SE is given in Ref.[17]. Subsequently, Tucci has written an open source software program called Entanglish (see Ref.[10]) for calculating SE.

The open-source computer program Quantum Fog (Ref.[14]) written by Tucci assumes that both classical and quantum BP are valid, because it uses the junction tree (JT) algorithm to do both classical and quantum inference with bnets. The JT algorithm[3] is a generalization of BP so as to include bnets with loops.[4]

Classical d-separation is an essential ingredient of most of Judea Pearl's work. It is crucial to all 3 rungs of what Pearl calls the 3 rungs of Causation. So having a quantum version of d-separation and BP probably implies that most of Pearl's bnet work, including his theory of causality, can be translated in a straightforward manner to the quantum realm. This is perhaps not too surprising because most of Pearl's bnet results depend to a large extent on the topology (i.e., graph structure) of the underlying DAG of a bnet, and that DAG should be essentially the same whether we are considering a quantum phenomenon or its classical limit.

[1]Discussed in chapter of Bayesuvius entitled "D-Separation".

[2]Discussed in chapter of Bayesuvius entitled "Message Passing (Belief Propagation)".

[3]Discussed in chapter of Bayesuvius entitled "Juntion Tree Algorithm".

[4] The quantum version of JT used by the current version of Quantum Fog is not quite correct because it does not use the vector amplitudes defined in this paper.

2 Notational Conventions and Preliminaries

This paper will employ the same notational conventions as Bayeuvius (Ref.[6]). Hence, if the reader encounters any notation that is not defined in this paper, he/she should consult Bayesuvius (especially its chapter entitled "Notational Conventions and Preliminaries"), where it is very likely to be defined.

Henceforth, whenever we write $[\mathcal{E}][h.c.]$, where $\mathcal{E}$ is some quantum operator, we will mean $\mathcal{E}\mathcal{E}^\dagger$, where $\mathcal{E}^\dagger$ is the Hermitian conjugate of $\mathcal{E}$. Also, whenever we write $\mathcal{N}(!x)$, we will mean a normalization constant that is independent of x.

Consider a DAG with nx nodes $\underline{x}^{nx} = (\underline{x}_i)_{i=0,1,\ldots,nx-1}$. A qbnet consists of such a DAG with a TPM (transition probability matrix) attached to each node, but unlike the TPMs for a classical bnet, these ones are complex valued and normalized differently. If we represent the TPM of node $\underline{x}_j$ by a **probability amplitude** $A(x_j|pa(\underline{x}_j))$, then $A(x_j|pa(\underline{x}_j))$ must satisfy the normalization condition

$$\sum_{x_j} \left|A(x_j|pa(\underline{x}_j))\right|^2 = 1 \; . \tag{1}$$

The amplitude for the full DAG is defined as in the classical case, by multiplying the TPMs of all the nodes:

$$A(x^{nx}) = \prod_j A(x_j|pa(\underline{x}_j)) \; . \tag{2}$$

Note that

$$\sum_{x^{nx}} |A(x^{nx})|^2 = 1 \; . \tag{3}$$

Suppose $\underline{a}.,\underline{b}.$ are disjoint multinodes in $\underline{x}^{nx}$. Denote the complement multinode of $\underline{a}.$ by $\underline{a}.^c = \underline{x}^{nx} - \underline{a}.$ Then we define the **vector amplitude** $|A\rangle(a.)$ and the probability $P(a.)$ by

$$|A\rangle(a.) = |A(a.)\rangle = \sum_{a.^c} A(a., a.^c)|a.^c\rangle \; , \tag{4}$$

with boundary cases $|A\rangle(x^{nx}) = A(x^{nx})$, $|A\rangle = \sum_{x^{nx}} A(x^{nx})|x^{nx}\rangle$. and

$$P(a.) = \sum_{a.^c} |A(a., a.^c)|^2 \; . \tag{5}$$

Note that $\sum_{a.} P(a.) = 1$ so $P(a.)$ is a bona fide probability distribution as the notation implies. Note also that

$$\langle A(a.)|A(a.)\rangle = P(a.) \; , \tag{6}$$

so you can think of $|A\rangle(a.)$ as being a generalized square root of $P(a.)$.

We define the **conditional vector amplitude** $|A\rangle(b.|a.)$ by

$$|A\rangle(b.|a.) = |A(b.|a.)\rangle = \frac{|A\rangle(b.,a.)}{|A\rangle(a.)} \ . \tag{7}$$

This is analogous to the definition of conditional probability, $P(b.|a.) = \frac{P(b.,a.)}{P(a.)}$. Note that $|A\rangle(b.|a.)$ isn't really a ket; it's a tuple of two kets, but it is convenient to represent it as a ket. We can also define the dual $\langle A|(b.|a.)$ of $|A\rangle(b.|a.)$ so that the following equations are satisfied:

$$\langle A|(b.|a.) = \langle A(b.|a.)| = \frac{\langle A|(b.,a.)}{\langle A|(a.)} \ , \tag{8}$$

$$\langle A(b.|a.)|A(b.|a.)\rangle = \frac{\langle A(b.,a.)|A(b.,a.)\rangle}{\langle A(a.)|A(a.)\rangle} = \frac{P(b.,a.)}{P(a.)} = P(b.|a.) \ . \tag{9}$$

Hence, you can think of $|A\rangle(b.|a.)$ as being a generalized square root of $P(b.|a.)$.

Suppose $\underline{a}., \underline{b}., \underline{c}.$ are disjoint multinodes such that $\underline{a}. \cup \underline{b}. \cup \underline{c}. = \underline{x}^{nx}$ Then

$$\begin{aligned} \sum_{b.} |A\rangle(a.,b.) &= \sum_{b.}\sum_{c.} |A\rangle(a.,b.,c.) && (10)\\ &= |A\rangle(a.) && (11)\end{aligned}$$

This is analogous to the probability marginalization rule $\sum_{b.} P(a.,b.) = P(a.)$.

Note that

$$\begin{aligned} \sum_{b.} |A\rangle(a.|b.)|A\rangle(b.) &= \sum_{b.} |A\rangle(a.,b.) && (12)\\ &= |A\rangle(a.) \ . && (13)\end{aligned}$$

This is analogous to the probability splitting rule $P(a.) = \sum_{b.} P(a.|b.)P(b.)$.

Suppose $\underline{a}.$ and $\underline{e}.$ are disjoint multinodes, with $\underline{e}.$ representing evidence. Note that

$$\begin{aligned} |A\rangle(a.|e.) &= \frac{|A\rangle(a.,e.)}{|A\rangle(e.)} && (14)\\ &= \frac{|A\rangle(e.|a.)|A\rangle(a.)}{|A\rangle(e.)} \ . && (15)\end{aligned}$$

This is analogous to the classical Bayes rule $P(a.|e.) = \mathcal{N}(!a.)P(e.|a.)P(a.)$.

Let $\mathcal{P}_{\underline{\lambda}}$ be the set of all probability distributions $P_{\underline{\lambda}}(\lambda)$ with $\lambda \in S_{\underline{\lambda}}$.

Let $\mathcal{H}_{\underline{x}}$ represent a Hilbert space spanned by an orthonormal basis $|x\rangle$, where $x \in S_{\underline{x}}$. Let $\mathcal{H}_{\underline{x},\underline{y}} = \mathcal{H}_{\underline{x}} \otimes \mathcal{H}_{\underline{y}}$.

Let $\mathcal{D}_{\underline{x}}$ be the set of density matrices acting on $\mathcal{H}_{\underline{x}}$. Likewise, let $\mathcal{D}_{\underline{x},\underline{y}}$ be the set of all density matrices acting on $\mathcal{H}_{\underline{x},\underline{y}}$. If $\rho_{\underline{x},\underline{y}} \in \mathcal{D}_{\underline{x},\underline{y}}$, then let $\rho_{\underline{x}}$ will denote the partial trace $\text{tr}_{\underline{y}}\rho_{\underline{x},\underline{y}}$.

Let $\mathcal{D}_{\underline{x},\underline{y},\underline{\lambda}^d}$ be the set of all density matrices in $\mathcal{D}_{\underline{x},\underline{y},\underline{\lambda}}$ which are diagonal in λ. In other words, $\rho_{\underline{x},\underline{y},\underline{\lambda}^d} \in \mathcal{D}_{\underline{x},\underline{y},\underline{\lambda}^d}$ if it is of the form

$$\rho_{\underline{x},\underline{y},\underline{\lambda}^d} = \sum_\lambda P_{\underline{\lambda}}(\lambda)|\lambda\rangle\langle\lambda|\rho^\lambda_{\underline{x},\underline{y}} \ , \tag{16}$$

where $\rho^\lambda_{\underline{x},\underline{y}} \in \mathcal{D}_{\underline{x},\underline{y}}$ for all λ and $P_{\underline{\lambda}} \in \mathcal{P}_{\underline{\lambda}}$.

	classical	quantum
Entropy	$H(\underline{x}) = -\sum_x P(x)\ln P(x)$	$S_\rho(\underline{x}) = S(\rho) = -\text{tr}_{\underline{x}}(\rho \ln \rho)$ for $\rho \in \mathcal{D}_{\underline{x}}$
Conditional Entropy	$H(\underline{x}\|\underline{y}) = -\sum_{x,y} P(x,y)\ln P(x\|y)$	$S_\rho(\underline{x}\|\underline{y}) = -S_\rho(\underline{y}) + S_\rho(\underline{x},\underline{y})$ for $\rho \in \mathcal{D}_{\underline{x},\underline{y}}$
Mutual Information	$H(\underline{x}:\underline{y}) = \sum_{x,y} P(x,y)\ln \frac{P(x,y)}{P(x)P(y)}$	$S_\rho(\underline{x}:\underline{y}) = S_\rho(\underline{x}) + S_\rho(\underline{y}) - S_\rho(\underline{x},\underline{y})$ for $\rho \in \mathcal{D}_{\underline{x},\underline{y}}$
Conditional Mutual Information	$H(\underline{x}:\underline{y}\|\underline{\lambda}) = \sum_{x,y,\lambda} P(x,y,\lambda)\ln \frac{P(x,y\|\lambda)}{P(x\|\lambda)P(y\|\lambda)}$	$S_\rho(\underline{x}:\underline{y}\|\lambda) = S_\rho(\underline{x}\|\lambda) + S_\rho(\underline{y}\|\underline{\lambda}) - S_\rho(\underline{x},\underline{y}\|\underline{\lambda})$ for $\rho \in \mathcal{D}_{\underline{x},\underline{y},\underline{\lambda}}$

Table 1: Definitions of various Entropies and Informations.

Table 1 gives the definitions of various entropies and informations used in classical Shannon Information Theory (SIT), and their counterparts in quantum SIT.

3 Quantum d-separation

The next box comes from the chapter of Bayesuvius entitled "D-Separation".

Claim 1 *(Classical d-separation Theorem)*
Suppose $\underline{A}.,\underline{B}.,\underline{Z}.$ are disjoint multinodes of a DAG G.
$\underline{A}. \perp_G \underline{B}.|\underline{Z}.$ *iff* $H(\underline{A}.:\underline{B}.|\underline{Z}.) = 0$ *for all P compatible with G.*

The proof of this theorem will not be presented here. It isn't presented in the current version of Bayesuvius either. To see it, you will have to look in Ref.[2], and in the original papers cited therein.

The definition of classical d-separation (i.e., of $\underline{A}. \perp_G \underline{B}.|\underline{Z}.$) only depends on the topology of the DAG. We will define quantum d-separation exactly as it it defined classically. Whether the bnet has probabilities or amplitudes for its TPMs does not make a difference at the level of the definition of d-separation. It only becomes important when trying to find a quantum analogue of the whole classical d-separation *theorem* which is stated in the box above. The remainder of this section will be dedicated to finding a quantum analogue to the whole box above.

We start by using the definitions introduced in the previous section to conclude that:

$$S_\rho(\underline{x} : \underline{y} | \underline{\lambda}^d) = \sum_\lambda P(\lambda) \left[S(\rho_{\underline{x}}^\lambda) + S(\rho_{\underline{y}}^\lambda) - S(\rho_{\underline{x},\underline{y}}^\lambda) \right] \tag{17}$$

for $\rho \in \mathcal{D}_{\underline{x},\underline{y},\underline{\lambda}^d}$.

Claim 2 *$S_\rho(\underline{x} : \underline{y} | \underline{\lambda}^d) = 0$ iff $\rho_{\underline{x},\underline{y}}^\lambda = \rho_{\underline{x}}^\lambda \rho_{\underline{y}}^\lambda$ for all λ.*

proof:

$S(\rho_{\underline{x}}^\lambda) + S(\rho_{\underline{y}}^\lambda) - S(\rho_{\underline{x},\underline{y}}^\lambda) = 0$ iff $\rho_{\underline{x},\underline{y}}^\lambda = \rho_{\underline{x}}^\lambda \rho_{\underline{y}}^\lambda$

QED

Next, we express each density matrix $\rho_{\underline{x},\underline{y}}^\lambda$ as

$$\rho_{\underline{x},\underline{y}}^\lambda = UDU^\dagger \ , \tag{18}$$

where U is a unitary matrix and D is a diagonal matrix with non-negative diagonal entries. Now let

$$A(x, y | x_0, y_0, \lambda) = \langle x | \langle y | U | x_0 \rangle | y_0 \rangle \tag{19}$$

and

$$\langle x_0 | \langle y_0 | D | x_0 \rangle | y_0 \rangle = P(x_0, y_0 | \lambda) \ . \tag{20}$$

Hence,

$$\rho_{\underline{x},\underline{y}}^\lambda = \sum_{x_0, y_0} \left[\sum_{x,y} |x\rangle |y\rangle A(x, y | x_0, y_0, \lambda) \right] P(x_0, y_0 | \lambda) \, [h.c.] \tag{21}$$

To make our future expressions more concise, define the two abbreviations $R_0 = (x_0, y_0)$, $R = (x, y)$. Then

$$\rho_{\underline{x},\underline{y},\underline{\lambda}^d} = \sum_{R_0, \lambda} \left[|\lambda\rangle \sum_R |R\rangle \underbrace{A(R | R_0, \lambda) \sqrt{P(R_0 | \lambda)} \sqrt{P(\lambda)}}_{A(R, R_0, \lambda)} \right] [h.c.] \ . \tag{22}$$

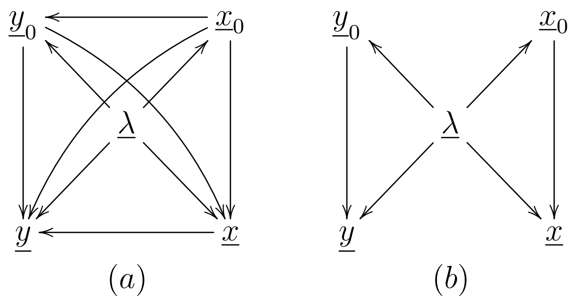

Figure 1: $S(\underline{x} : \underline{y}|\underline{\lambda}^d)$ is nonzero for the qbnet of panel (a) but zero for the qbnet of panel (b).

From the definitions of conditional probabilities and conditional amplitudes, we get

$$A(R|R_0, \lambda) = A(y|x, R_0, \lambda)A(x|R_0, \lambda) \tag{23}$$

and

$$P(R_0|\lambda) = P(y_0|x_0, \lambda)P(x_0|\lambda) \ . \tag{24}$$

Therefore,

$$A(R, R_0, \lambda) \ = \ A(y|x, R_0, \lambda)A(x|R_0, \lambda)\sqrt{P(y_0|x_0, \lambda)P(x_0|\lambda)P(\lambda)} \ . \tag{25}$$

Eq.(25) can be represented graphically by the qbnet Fig.1(a). The TPMs, printed in blue, for the 2 qbnets (a) and (b) of Fig.1, are as follows.

$$A(\lambda) = \sqrt{P(\lambda)} \tag{26}$$

$$A(x_0|\lambda) = \sqrt{P(x_0|\lambda)} \tag{27}$$

$$A(y_0|x_0, \lambda) = \left\{ \begin{array}{ll} \sqrt{P(y_0|x_0, \lambda)} & \text{for Fig.1}(a) \\ \sqrt{P(y_0|\lambda)} & \text{for Fig.1}(b) \end{array} \right. \tag{28}$$

$$A(x|R_0, \lambda) = \left\{ \begin{array}{ll} A(x|R_0, \lambda) & \text{for Fig.1}(a) \\ A(x|x_0, \lambda) & \text{for Fig.1}(b) \end{array} \right. \tag{29}$$

$$A(y|x, R_0, \lambda) = \left\{ \begin{array}{ll} A(y|x, R_0, \lambda) & \text{for Fig.1}(a) \\ A(y|y_0, \lambda) & \text{for Fig.1}(b) \end{array} \right. \tag{30}$$

It's easy to check that

- $\underline{x} \perp_G \underline{y}|\underline{\lambda}$ is false and $S(\underline{x} : \underline{y}|\underline{\lambda}^d) \neq 0$ in Fig.1(a), whereas
- $\underline{x} \perp_G \underline{y}|\underline{\lambda}$ is true and $S(\underline{x} : \underline{y}|\underline{\lambda}^d) = 0$ in Fig.1(b).

So far, we have shown how, given any density matrix $\rho \in \mathcal{D}_{\underline{x},\underline{y},\underline{\lambda}}$, one can construct a qbnet. This method of constructing a qbnet from a density matrix ρ (or vice versa, constructing a ρ from a qbnet) can be generalized to finding a qbnet for any $\rho \in \mathcal{D}_{\underline{x}^{nx}}$. for arbitrary nx. Now we argue that the proof of the quantum d-separation theorem should be formally identical to the proof of the classical d-separation theorem. The only difference between the proofs is that whenever a probability occurs in the classical proof, it must be replaced by a vector amplitude in the quantum proof. Of course, probabilities and vector amplitudes are normalized differently, but that should not change the form of the proofs. Note that we have been careful to show that vector amplitudes can be conditioned and satisfy a splitting rule, just like probabilities do. Also, we have been careful to define d-separation $\underline{A}. \perp_G \underline{B}.|\underline{Z}.$ to be identical for the classical and quantum cases. Hence, we argue that, without looking at the details of the proof of the classical d-separation theorem, one can conclude that the following theorem must be true:

Claim 3 *(Quantum d-separation Theorem)*
Suppose $\underline{A}., \underline{B}., \underline{Z}.$ are disjoint multinodes of a DAG G.
$\underline{A}. \perp_G \underline{B}.|\underline{Z}.$ iff $S_\rho(\underline{A}. : \underline{B}.|\underline{Z}.^d) = 0$ for all ρ compatible with G.

Define the squashed entanglement of a density matrix $\rho_{\underline{x},\underline{y}}$ by

$$E_{sq}(\rho_{\underline{x},\underline{y}}) = \frac{1}{2}\min_{\rho\in\mathcal{D}} S_\rho(\underline{x} : \underline{y}|\underline{\lambda}^d) \tag{31}$$

where $\mathcal{D} = \{\rho \in \mathcal{D}_{\underline{x},\underline{y},\underline{\lambda}^d} \ | \ \mathrm{tr}_{\underline{\lambda}^d}\rho_{\underline{x},\underline{y},\underline{\lambda}^d} = \rho_{\underline{x},\underline{y}}\}$. Then the quantum d-separation theorem immediately(?) implies the following.

Claim 4 *Suppose $\underline{A}., \underline{B}.$ are disjoint multinodes of a DAG G,*
($\underline{A}. \perp_G \underline{B}.|\underline{Z}.$ for some $\underline{Z}.$ such that $\underline{A}., \underline{B}., \underline{Z}.$ are disjoint multinodes of G) iff $E_{sq}(\rho_{\underline{A}.,\underline{B}.}) = 0$.

4 Quantum Belief Propagation

The rules (and their proof) of classical BP can be found in the Bayesuvius chapter entitled "Message Passing (Belief Propagation)". Just like the proof of the classical d-separation theorem, the proof of the rules for classical BP relies on 3 ingredients:

- the topology of a DAG
- the definition of conditional probabilities as ratios of joint probabilities
- the splitting rule for probabilities

Since these 3 ingredients are also available in the quantum side if we replace probabilities by vector amplitudes, we can conclude that the rules for quantum BP are formally the same as those for classical BP, modulus the replacement of probabilities by vector amplitudes. The difference in normalization of probabilities and vector amplitudes does not make the rules for classical and quantum BP different because these rules are defined up to a normalization constant.

In Bayesuvius, the BP chapter entitled "Message Passing (Belief Propagation)" considers a general case of classical BP (viz., BP for polytrees (BP-Gen)) and a special case of classical BP (viz., BP for bipartite bnets (BP-BB)). We end this section with 2 subsections dedicated to the quantum analogues of the rules for BP-Gen and BP-BB. Those 2 subsections are simply exact quotes from the BP chapter in Bayesuvius, except that all P's have been crossed out and replaced by $|A\rangle$'s.

4.1 Quantum BP for polytrees

Let $\underline{a}^{na} = (\underline{a}_i)_{i=0,1,\ldots,na-1}$ denote the parents of $\underline{x}$ and $\underline{b}^{nb} = (\underline{b}_i)_{i=0,1,\ldots,nb-1}$ its children.

Define

$$
\begin{aligned}
\pi_{\underline{x}}(x) &= \sum_{a^{na}} \overset{|A\rangle}{\cancel{P}}(x|a^{na}) \prod_i \pi_{\underline{x} \Leftarrow \underline{a}_i}(a_i) && (32) \\
&= E_{\underline{a}^{na}}[\overset{|A\rangle}{\cancel{P}}(x|a^{na})] && (33)
\end{aligned}
$$

(boundary case: if $\underline{x}$ is a root node, use $\pi_{\underline{x}}(x) = \overset{|A\rangle}{\cancel{P}}(x)$.) and

$$
\lambda_{\underline{x}}(x) = \prod_i \lambda_{\underline{b}_i \Rightarrow \underline{x}}(x) \ . \tag{34}
$$

(boundary case: if $\underline{x}$ is a leaf node, use $\lambda_{\underline{x}}(x) = 1$.)

- **RULE 1: (red parent)**

$$\underbrace{\lambda_{\underline{x}\Rightarrow\underline{a}_i}(a_i)}_{OUT} = \mathcal{N}(!a_i)\sum_x \left[\underbrace{\lambda_{\underline{x}}(x)}_{IN}\sum_{(a_k)_{k\neq i}}\left(\stackrel{|A\rangle}{P}(x|a^{na})\prod_{k\neq i}\underbrace{\pi_{\underline{x}\Leftarrow\underline{a}_k}(a_k)}_{IN}\right)\right] \tag{35}$$

$$= \mathcal{N}(!a_i)\sum_x \left[\lambda_{\underline{x}}(x)E_{(\underline{a}_k)_{k\neq i}}[\stackrel{|A\rangle}{P}(x|a^{na})]\right] \tag{36}$$

$$= \mathcal{N}(!a_i)E_{(\underline{a}_k)_{k\neq i}}E_{\underline{x}|a^{na}}\lambda_{\underline{x}}(x) \tag{37}$$

(boundary case: if $\underline{x}$ is a root node, use $\lambda_{\underline{x}\Rightarrow\underline{a}_i}(a_i) = \mathcal{N}(!a_i)$.)

- **RULE 2: (red child)**

$$\underbrace{\pi_{\underline{b}_i\Leftarrow\underline{x}}(x)}_{OUT} = \mathcal{N}(!x)\underbrace{\pi_{\underline{x}}(x)}_{IN}\prod_{k\neq i}\underbrace{\lambda_{\underline{b}_k\Rightarrow\underline{x}}(x)}_{IN} \tag{38}$$

(boundary case: if $\underline{x}$ is a leaf node, use $\pi_{\underline{b}_i\Leftarrow\underline{x}}(x) = \mathcal{N}(!x)\pi_{\underline{x}}(x)$.)

In the above equations, if the range set of a product is empty, then define the product as 1; i.e., $\prod_{k\in\emptyset}F(k) = 1$.

Claim: Define

$$BEL^{(t)}(x) = \mathcal{N}(!x)\lambda^{(t)}_{\underline{x}}(x)\pi^{(t)}_{\underline{x}}(x) \; . \tag{39}$$

Then

$$\lim_{t\to\infty} BEL^{(t)}(x) = \stackrel{|A\rangle}{P}(x|\epsilon) \; . \tag{40}$$

This says that the belief in $\underline{x} = x$ converges to $\stackrel{|A\rangle}{P}(x|\epsilon)$ and it equals the product of messages received from all parents and children of $\underline{x} = x$.

4.2 Quantum BP for Bipartite Bnets

1. **Traversing an x (i.e., root) node.**

   For $i = 0, 1, \ldots, nx - 1$, if $\alpha \in nb(i)$, then,

$$m^{(t)}_{\alpha\Leftarrow i}(x_i) = \prod_{\beta\in nb(i)-\alpha} m^{(t-1)}_{\beta\Rightarrow i}(x_i) \; , \tag{41}$$

   whereas if $\alpha \notin nb(i)$

$$m^{(t)}_{\alpha\Leftarrow i}(x_i) = m^{(t-1)}_{\alpha\Leftarrow i}(x_i) \; . \tag{42}$$

2. **Traversing an f (i.e., leaf) node.**

   For $\alpha = 0, 1, \ldots, nf-1$, if $i \in nb(\alpha)$, then

$$
\begin{aligned}
m^{(t)}_{\alpha \Rightarrow i}(x_i) &= \sum_{(x_k)_{k \in nb(\alpha)-i}} f_\alpha(x_{nb(\alpha)}) \prod_{k \in nb(\alpha)-i} m^{(t-1)}_{\alpha \Leftarrow k}(x_k) && (43)\\
&= E^{(t-1)}_{(x_k)_{k \in nb(\alpha)-i}} [f_\alpha(x_{nb(\alpha)})] \, , && (44)
\end{aligned}
$$

   whereas if $i \notin nb(\alpha)$

$$
m^{(t)}_{\alpha \Rightarrow i}(x_i) = m^{(t-1)}_{\alpha \Rightarrow i}(x_i) \, . \tag{45}
$$

In the above equations, if the range set of a product is empty, then define the product as 1; i.e., $\prod_{k \in \emptyset} F(k) = 1$.

Claim:

$$
\overset{|A\rangle}{\cancel{P}}(x_i|\epsilon) = \lim_{t \to \infty} \mathcal{N}(!x_i) \prod_{\alpha \in nb(i)} m^{(t)}_{\alpha \Rightarrow i}(x_i) \tag{46}
$$

and

$$
\overset{|A\rangle}{\cancel{P}}(x_{nb(\alpha)}|\epsilon) = \lim_{t \to \infty} \mathcal{N}(!x_{nb(\alpha)}) f_\alpha(x_{nb(\alpha)}) \prod_{k \in nb(\alpha)} m^{(t)}_{\alpha \Leftarrow k}(x_k) \, . \tag{47}
$$

In the classical case, $f_\alpha(x_{nb(\alpha)})$ stands for a real valued function (e.g., $P(f_\alpha = 1|x_{nb(\alpha)})$), whereas in the quantum case, it stands for a vector amplitude (e.g., $|A\rangle(f_\alpha = 1|x_{nb(\alpha)})$).

A Appendix: Reduced qbnet

$$
\begin{tikzcd}
 & \underline{\lambda} \arrow[dl] \arrow[dr] & \\
\underline{Y} & & \underline{X} \arrow[ll]
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \underline{\lambda} \arrow[dl] \arrow[dr] & \\
\underline{Y} & & \underline{X}
\end{tikzcd}
$$

(a) (b)

Figure 2: The 2 qbnets in Fig.1 can sometimes be reduced to these 2 qbnets.

Let $\underline{X} = (\underline{x}, \underline{x}_0)$ and $\underline{Y} = (\underline{y}, \underline{y}_0)$. Note that if, as indicated in Eq.(49), $A(x|R_0, \lambda)$ is independent of y_0, then the 2 qbnets in Fig.1 can be reduced to the

2 qbnets in Fig.2, The node TPMs, printed in blue, of the qbnets in Fig.2, are as follows:

$$A(\lambda) = \sqrt{P(\lambda)} \tag{48}$$

$$A(x, x_0|\cancel{y_0}, \lambda) = \left\{ \begin{array}{ll} A(x|\overset{x_0}{\cancel{R_0}}, \lambda)\sqrt{P(x_0|\lambda)} & \text{for Fig.2}(a) \\ A(x|x_0, \lambda)\sqrt{P(x_0|\lambda)} & \text{for Fig.2}(b) \end{array} \right. \tag{49}$$

$$A(y, y_0|x, x_0, \lambda) = \left\{ \begin{array}{ll} A(y|x, R_0, \lambda)\sqrt{P(y_0|x_0, \lambda)} & \text{for Fig.2}(a) \\ A(y|y_0, \lambda)\sqrt{P(y_0|\lambda)} & \text{for Fig.2}(b) \end{array} \right. \tag{50}$$

References

[1] Matthias Christandl and Andreas Winter. squashed entanglement: An additive entanglement measure. *Journal of Mathematical Physics*, 45(3):829–840, Mar 2004.

[2] Judea Pearl. *Probabilistic Inference in Intelligent Systems*. Morgan Kaufmann, 1988.

[3] Judea Pearl. *Causality: Models, Reasoning, and Inference, Second Edition*. Cambridge University Press, 2013.

[4] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

[5] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic Books, 2018.

[6] Robert R. Tucci. Bayesuvius (book). `https://github.com/rrtucci/Bayesuvius/raw/master/main.pdf`.

[7] Robert R. Tucci. Entanglement of bell mixtures of two qubits. `https://arxiv.org/abs/quant-ph/0103040`.

[8] Robert R. Tucci. Entanglement of distillation and conditional mutual information. `https://arxiv.org/abs/quant-ph/0202144`.

[9] Robert R. Tucci. Entanglement of formation and conditional information transmission. `https://arxiv.org/abs/quant-ph/0010041`.

[10] Robert R. Tucci. Entanglish (software). https://github.com/rrtucci/Entanglish.

[11] Robert R. Tucci. An introduction to quantum Bayesian networks for mixed states. https://arxiv.org/abs/1204.1550.

[12] Robert R. Tucci. Quantum Bayesian nets. https://arxiv.org/abs/quant-ph/9706039.

[13] Robert R. Tucci. Quantum entanglement and conditional information transmission. https://arxiv.org/abs/quant-ph/9909041.

[14] Robert R. Tucci. Quantum Fog (software). https://github.com/artiste-qb-net/quantum-fog.

[15] Robert R. Tucci. Relaxation method for calculating quantum entanglement. https://arxiv.org/abs/quant-ph/0101123.

[16] Robert R. Tucci. Separability of density matrices and conditional information transmission. https://arxiv.org/abs/quant-ph/0005119.

[17] Wikipedia. Squashed entanglement. https://en.wikipedia.org/wiki/Squashed_entanglement.