Tutorial on How to Measure Link Strengths in Discrete Bayesian Networks

Georgia Tech Research Report: GT-ME-2009-001
September 2, 2009

Imme Ebert-Uphoff *
Adjunct Associate Professor
George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0405
ebert@me.gatech.edu

Abstract

This report discusses measures for link strength in Discrete Bayesian Networks, i.e. measures for the strength of connection along a specific edge. It is a revised version of Report GT-IIC-07-01 (Jan 2007) with improved literature review and explanations.

The target application is the visualization of the strengths of the edge connections in a Bayesian Network learned from data to learn more about the inherent properties of the system. The report reviews existing link strength measures, provides an accessible derivation of the primary measure, proposes some simple variations of the primary measure and compares their resulting properties.

1 The Concept of Link Strength

Boerlage was the first to formally introduce the concept of link strength for Bayesian Networks (Boerlage 1992). Boerlage defines *connection strength* for any pair of nodes (adjacent or not) to measure the strength between those nodes taking any possible path between them into account. In contrast *link strength* (also known as *arc weight*) is defined for a specific edge and measures the strength of connection only along that single edge.

To demonstrate the difference between these concepts in particular for adjacent nodes consider the network in Figure 1. Each of the three nodes only has two states, $True$ and $False$. Let us focus on the connection

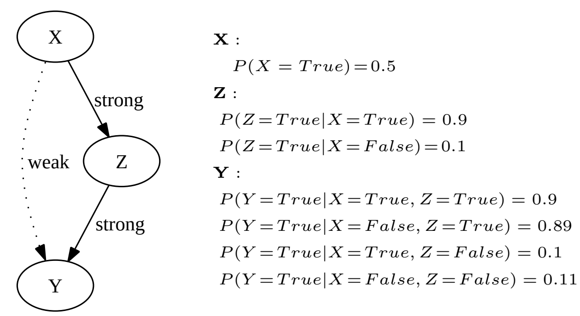

Figure 1: Sample BN with weak link from X to Y, but strong links from X to Z and Z to Y.

between nodes X and Y. For this sample network the *direct* link from X to Y is weak – this is easily seen because the state of X has little effect on the values of $P(Y|X,Z)$. In contrast the *indirect* link from X to Y *through* Z is very strong. According to the above (vague) concept definitions, the connection strength, CS, between X and Y is strong here, but the link strength, LS, of the edge $X \rightarrow Y$ is weak:

$$\begin{aligned} \text{CS}(X,Y) &= \text{strong}, \\ \text{LS}(X \rightarrow Y) &= \text{weak}. \end{aligned}$$

Any pair of measures for link strength and connection strength should yield this result for the above example.

Link strength measures have been proposed for two primary applications:

1. **Visualization:** Boerlage 1992 realized that in graph representations of DAGs it is helpful to show not only the *existence* of arrows, but also the *strength* of the various connections to better understand the network. He proposed two types

*Joint appointment with the Robotics and Intelligent Machines Center, School of Interactive Computing, College of Computing, Atlanta, Georgia 30308.

of enhanced graph representations. One uses connection strength to show the range of influence of a target variable, the other uses link strength to show how much information travels across specific paths. He also realized that these graphs are useful to generate explanations for reasoning in Bayesian Networks.

2. **Approximate Inference:** Since connection strength can be used to determine the range of influence of variables, it can be used in approximate inference to determine which variables can be neglected in the approximation (Jitnah and Nicholson 1998). Connection strength is generally computationally expensive to calculate, so link strength measures are sometimes employed to approximate connection strength quickly. For that purpose the link strength measures must be computationally efficient and combinations of them must yield either bounds on or a decent approximation of connection strength between any two nodes (Jitnah and Nicholson 1998).

The discussion in this paper is targeted toward the first application above. Thus computational complexity is of little concern here.

Furthermore, structure learning and causal discovery was still in its infancy at the time when Boerlage proposed to use link strength for visualization and thus there exist potential applications today that he could not possibly have foreseen. Most importantly, we believe that Link Strength measures will soon play an important role to identify a system's major causal pathways. For example, using constraint-based structure learning algorithms to learn a system's structure from data often yields large sets of Markov-equivalent DAGs. (Node adjacencies are identical in all of those, but the direction of certain arrows may be reversed.) Using link strength can help researchers to narrow down the alternatives by eliminating those networks with only minor differences (i.e. those where arrows are reversed only for edges with very small link strength), thus yielding a more manageable number of major causal hypotheses to consider.

Link strength could also play an important role to evaluate the quality of structure learning algorithms. Rather than just counting the number of incorrect edges when evaluating models, it may be useful to weigh the incorrect arrows by their link strength values in the correct model. Much research remains concerning the details of this approach.

This document is a revised version of an earlier research report by the same author (Ebert-Uphoff 2007). The main difference is a more thorough literature review and better explanations of the concepts in the context of the literature. The goal of this report is to revive the powerful concept of link strength and to direct more attention to its usefulness for these and other applications. All measures are defined in this document only for *discrete* Bayesian Networks.

2 Background

Link strength measures tend to be built on measures for uncertainty and connection strength. Those are briefly reviewed below.

2.1 Measuring Uncertainty

The most commonly used measure for the uncertainty of a random variable is **Entropy** (Shannon 1949). Entropy of a discrete random variable, X, with states x_i is defined as

$$U(X) = \sum_{x_i} P(x_i) \log_2 \frac{1}{P(x_i)}.$$

While entropy has several limitations (see for example Pearl 1988, pp. 322-323), those are shared by any other measure that is a function of only the *probabilities* of a random variable's states to measure uncertainty. Thus entropy nevertheless remains by far the most popular measure for uncertainty.

2.2 Measuring Connection Strength

Connection strength between X and Y measures how strongly information on the state of X affects the state of Y (and vice versa).

Boerlage defined connection strength for two binary variables X, Y by comparing the distribution for different values of parent X to each other, namely

$$CS^{Boerlage}(X,Y) = d(P(y|+x), P(y|-x)),$$

where $+x$ and $-x$ denote the two possible states of variable X and d is any valid distance measure between two probability distributions (several examples of d are given in Boerlage 1992).

A more standard approach is to compare the distribution of Y *without* any evidence to the distribution of Y if there *is* evidence for X. For example, in their earlier work Nicholson and Jitnah apply the Bhattacharyya distance (Nicholson and Jitnah 1997) to the distributions of Y and $Y|X$. However, that approach yields less suitable results than Mutual Information (Nicholson and Jitnah 1998).

Finally, Mutual Information is the most common implementation of this idea, as proposed by Pearl (Pearl 1988): one simply calculates $U(Y)$ and $U(Y|X)$ and compares them. **Mutual Information (MI)** is defined as

$$MI(X,Y) = U(Y) - U(Y|X), \quad (1)$$

where $U(Y|X)$ is calculated by averaging $U(Y|x_i)$ over all possible states x_i of X, taking $P(x_i)$ into account:

$$U(Y|X) = \sum_{x_i} P(x_i) U(Y|x_i). \quad (2)$$

Simple arithmetic transformations yield the formula:

$$MI(X,Y) = \sum_{x,y} P(x,y) \log_2 \left(\frac{P(x,y)}{P(x)P(y)} \right).$$

2.3 Measuring Shielding Properties

Another useful concept is **Conditional Mutual Information (CMI)** which compares the uncertainty in Y if we know the state of Z to the uncertainty in Y if we know the states of Z *and* X. CMI is defined as

$$MI(X,Y|Z) = U(Y|Z) - U(Y|X,Z),$$

where

$$U(Y|X,Z) = \sum_{x,z} P(x,z) U(Y|x,z) \quad (3)$$
$$= \sum_{x,z} P(x,z) \sum_{y} P(y|x,z) \log_2 \left(\frac{1}{P(y|x,z)} \right).$$

$MI(X,Y|Z)$ can be written as

$$MI(X,Y|Z) = \sum_{x,z} P(x,z) \sum_{y} P(y|x,z) \log_2 \frac{P(y|x,z)}{P(y|z)}. \quad (4)$$

In the context of Bayesian Networks CMI is sometimes used to check how well a set of variables, $\mathbf{Z}$, shields X from Y. For example, Friedman et al. 1999 use conditional mutual information in the context of structure learning to determine how well a node X_i is shielded from node X_j by its assumed set of parents, $PA(X_i)$:

$$M_{shield}(X_i, X_j|\text{current network}) = MI(X_i, X_j|PA(X_i)).$$

Pappas and Gillies 2002 use conditional mutual information to develop an accuracy measure for Bayesian Networks that measures the accuracy of the conditional independencies implied by its structure.

3 Existing Link Strength Measures

There is much less literature on link strength than on connection strength and it appears to be harder to measure. A thorough literature review yielded three different link strength measures:

1. Boerlage defined connection strength and link strength for Bayesian Networks with only binary nodes. Building on the definition of $CS^{Boerlage}(X,Y)$ provided above, he defines connection strength in the presence of evidence as

   $$CS(X,Y|\mathbf{e}) = d\left(P(y|+x,\mathbf{e}), P(y|-x,\mathbf{e})\right),$$

   where $\mathbf{e}$ represents the evidence. Link strength is then defined as

   $$LS^{Boerlage}(X \to Y) = \max_{z} CS(X,Y|\mathbf{z}),$$

   where $\mathbf{z}$ denotes all possible state combinations of all parents of Y other than X.

   The central question when using this approach is how to define connection strength for discrete variables, since Boerlage only defined it for binary variables. If one uses the prevalent measure of mutual information for this purpose, then the link strength measure becomes very similar to the third approach below.

2. Lacave and Diez 2004 proposed a measure for the *magnitude of influence* of link $X \to Y$ as

   $$LS^{Lacave}(X \to Y) = \max_{y,x,\mathbf{z}} \left[P(Y \geq y|x,\mathbf{z}) - P(Y \geq y|x_0,\mathbf{z}) \right]$$

   where x, y are the states of nodes X and Y and $\mathbf{z}$ is the state combination of all other parents, $\mathbf{Z}$, of Y. Furthermore, X and Y are assumed to be ordinal variables with their states ordered such that the "lowest" value appears first, i.e. x_0 is the lowest value of X.

   This may be a good approach in the case of a network where node states can be ordered such that nodes tend to have positive influence on each other, i.e. *increasing* the state of a node tends to *increase* the state of its child nodes. However, if no such order can be established, using the reference state x_0 appears to be somewhat arbitrary.

3. The measure presented by Nicholson and Jitnah 1998 and Jitnah 1999 can be written as

   $$LS^{Jitnah-Nicholson}(X \to Y) = \sum_{x,\mathbf{z}} P_{pr}(\mathbf{z}) P_{pr}(x) \sum_{y} P(y|x,\mathbf{z}) \log_2 \frac{P(y|x,\mathbf{z})}{P_{pr}(y|\mathbf{z})} \quad (5)$$

   where the term P_{pr} indicates an approximation of probability that avoids using any inference.

   Nicholson and Jitnah do not provide a derivation of their measure and it already contains some approximate terms. However, it seems more than likely that they derived their measure using conditional mutual information. In fact if we define link strength as mutual information of X, Y conditioned on the set $\mathbf{Z}$ of all other parents of Y, namely

$$LS^{true}(X \rightarrow Y) \quad = \quad MI(X, Y|\mathbf{Z}),$$

we get what we call here the *True Average Link Strength*

$$LS^{true}(X \rightarrow Y) = \sum_{x,\mathbf{z}} P(x, \mathbf{z}) \sum_{y} P(y|x, \mathbf{z}) \log_2 \frac{P(y|x, \mathbf{z})}{P(y|\mathbf{z})}. \quad (6)$$

Equation (6) matches Equation (5) except for the approximations already employed by Nicholson and Jitnah.

Note that all three definitions for link strength share a common feature - they all condition on the set $\mathbf{Z}$ of all other parents of child Y. Why this is a good approach is explained in the next section. Furthermore, Boerlage's and Jitnah's link strength measures are both closely tied to their respective definitions of connection strength measures - Boerlage uses maximization while Nicholson and Jitnah obtain link strength through *averaging* over the states of set $\mathbf{Z}$.

We feel that Equation (6) provides the most solid basis for a link strength measure, because it is built step by step on a sequence of well established concepts (entropy, MI and CMI) and does not involve any arbitrary choices. Measure (6) is thus considered to be the *primary* link strength measure for the remainder of this paper.

4 Why Conditioning on the Other Parents Works

All existing link strength measures condition on the set $\mathbf{Z}$ of other parents of Y to focus on the connection from parent X to child Y *solely* along edge $X \rightarrow Y$. While this is an intuitive approach it is worthwhile to formally prove that there truly remain no indirect open pathways between X and Y. (for example through common descendants of X and Y).

Theorem 1 *Consider a BN with a link $X \rightarrow Y$. Let $\mathbf{Z} = PA(Y) - \{X\}$ denote the set of all parents of Y other than X. If all other parents, $\mathbf{Z}$, are instantiated and no other nodes are instantiated, then the only open pathway between X and Y is the direct link $X \rightarrow Y$.*

Proof Let us denote the BN as (G,P) where G denotes the DAG and P the joint probability. Let $\widehat{G}$ be a modified DAG generated by deleting edge $X \rightarrow Y$ in G. Since edge $X \rightarrow Y$ does not exist in $\widehat{G}$, set $\mathbf{Z}$ represents *all* parents of Y in $\widehat{G}$. Furthermore, X is not a descendent of Y in $\widehat{G}$ - otherwise the original DAG G would contain a directed cycle. Due to the Markov condition any node in a BN is conditionally independent of its non-descendents given only its parents. Therefore in the BN with DAG $\widehat{G}$, node Y is conditionally independent of X given $\mathbf{Z}$.

Since X and Y are conditionally independent given $\mathbf{Z}$ if the edge from X to Y is removed, it is clear that edge $X \rightarrow Y$ is indeed the *only* path along which information can flow from X to Y in the original network if $\mathbf{Z}$ is instantiated. ∎

This fact ensures that either one of the three link strength measures indeed only measures information flow along the considered edge.

5 Variations of True Average Link Strength

Let us first recall the definition of the primary measure.

Definition *True Average Link Strength* is defined as

$$LS^{true}(X \rightarrow Y) = \sum_{x,\mathbf{z}} P(x, \mathbf{z}) \sum_{y} P(y|x, \mathbf{z}) \log_2 \frac{P(y|x, \mathbf{z})}{P(y|\mathbf{z})}. \quad (7)$$

Interpretation By how much is the uncertainty in Y reduced by knowing the state of X, if the states of all other parent variables are known (averaged over the parent states using their *actual* joint probability)?

5.1 Blind Average Link Strength

A new measure can be derived from True Average Link Strength by disregarding the actual frequency of occurrence of the parent states. Namely we assume that $X, \mathbf{Z}$ are independent and all uniformly distributed, resulting in the following approximations $\hat{P}$:

$$\hat{P}(x, \mathbf{z}) = P(x)P(\mathbf{z}), \ \hat{P}(x) = \frac{1}{\#(X)}, \ \hat{P}(\mathbf{z}) = \frac{1}{\#(\mathbf{Z})}, \quad (8)$$

where $\#(X)$ denotes the number of discrete states of X, etc.

Essentially, this approximation goes one step further in simplifications than the approximations by Jitnah and Nicholson. However, these additional simplifications have a justification of their own. Namely an interesting property of the set of assumptions (8) is that it creates a local measure that depends only on the child node and its conditional probability table, but nothing else in the network. One may argue that for some applications such a local measure is actually more natural, since the connection between parents and child should be independent of any changes in probabilities elsewhere in the network. This discussion is continued in a later section discussing properties of the measures.

Definition *Blind Average Link Strength* is defined as

$$LS^{blind}(X \rightarrow Y) \quad = \quad \hat{U}(Y|\mathbf{Z}) - \hat{U}(Y|X,\mathbf{Z}),$$

where

$$\hat{U}(Y|\mathbf{Z}) = \frac{1}{\#(X)\#(\mathbf{Z})} \sum_{x,y,\mathbf{z}} P(y|x,\mathbf{z}) \log_2 \frac{\#(X)}{\sum_x P(y|x,\mathbf{z})},$$

$$\hat{U}(Y|X,\mathbf{Z}) = \frac{1}{\#(X)\#(\mathbf{Z})} \sum_{x,y,\mathbf{z}} P(y|x,\mathbf{z}) \log_2 P(y|x,\mathbf{z}).$$

Note that $\hat{U}(Y|\mathbf{Z})$ and $\hat{U}(Y|X,\mathbf{Z})$ are obtained from $U(Y|\mathbf{Z})$ and $U(Y|X,\mathbf{Z})$ simply by replacing P by approximations $\hat{P}$ of (8). This definition yields the simple formula

$$LS^{blind}(X \rightarrow Y) \quad = \quad \frac{1}{\#(X)\#(\mathbf{Z})} \sum_{x,y,\mathbf{z}} P(y|x,\mathbf{z}) \log_2 \left(\frac{P(y|x,\mathbf{z})}{\frac{1}{\#(X)} \sum_x P(y|x,\mathbf{z})} \right),$$

where $P(y|x,\mathbf{z})$ is given by the conditional probability table of Y and *no* inference is required at all.

Interpretation By how much is the uncertainty in Y reduced by knowing the state of X, if the states of all other parent variables are known (averaged over the parent states assuming all parents are independent of each other and uniformly distributed)?

Comment: This is the simplest and computationally least expensive measure. It is also a local measure, taking only the child and *its* conditional probabilities into account, thus allowing for isolated analysis of child and parents, regardless of the rest of the network.

5.2 Link Strength Percentages

In some cases the *absolute amount* of uncertainty reduction in a variable may provide less insight than the *percentage* of the original uncertainty that was removed. Thus we propose a simple extension of the Link Strength Measures, namely Link Strength Percentage, to be used in conjunction with Link Strength.

Definition *True Average Link Strength Percentage* is defined for $U(Y|\mathbf{Z}) \neq 0$ as

$$LS\%^{true}(X \rightarrow Y) = \frac{LS^{true}(X \rightarrow Y)}{U(Y|\mathbf{Z})} \cdot 100 \qquad (9)$$

Applying independence and uniformity assumptions (8) to the True Average Link Strength Percentage (9) yields the Blind Average Link Strength Percentage.

Definition *Blind Average Link Strength Percentage* is defined for $\hat{U}(Y|\mathbf{Z}) \neq 0$ as

$$LS\%^{blind}(X \rightarrow Y) \quad = \quad \frac{LS^{blind}(X \rightarrow Y)}{\hat{U}(Y|\mathbf{Z})} \cdot 100$$

$LS\%^{true}(X \rightarrow Y)$ is undefined if $U(Y|\mathbf{Z}) = 0$ and $LS\%^{blind}(X \rightarrow Y)$ is undefined if $\hat{U}(Y|\mathbf{Z}) = 0$, which makes perfect sense. If there is zero uncertainty to begin with, then it makes no sense to ask what percentage of it was removed.

6 Properties of Link Strength

This section provides additional intuition on the different link strength measures by presenting some properties and illustrating them by several examples.

6.1 Do The Measures Behave As Desired?

Table 1 shows the results for the network in Figure 1 that demonstrates the difference between connection strength and link strength. Listed are True Average and Blind Average Link Strength for each edge, as well as Mutual Information for each node pair. As expected, both link strength measures yield high values for the edges from X to Z and from Z to Y, while the link strength of edge $X \rightarrow Y$ nearly vanishes. In contrast, X and Y are strongly connected according to mutual information, because information flows through the chain $X \rightarrow Z \rightarrow Y$.

Table 1: Results for Sample Network in Figure 1

	LS^{true}	LS^{blind}	MI
$X \rightarrow Y$	0.000	0.000	0.311
$X \rightarrow Z$	0.531	0.531	0.531
$Z \rightarrow Y$	0.204	0.516	0.515

6.2 Scale

A few comments on scale are in order. It is guaranteed that the values of the measures increase monotonously when uncertainty is reduced, but the scale of the actual values is *not* linear and *not* intuitive - just as is

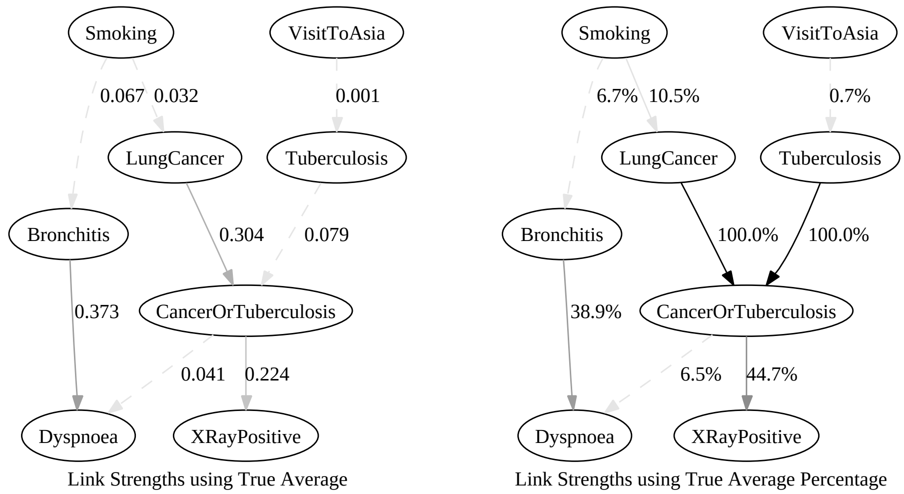

Figure 2: True Average Link Strength (left) and Percentage (right) for Asia Model.

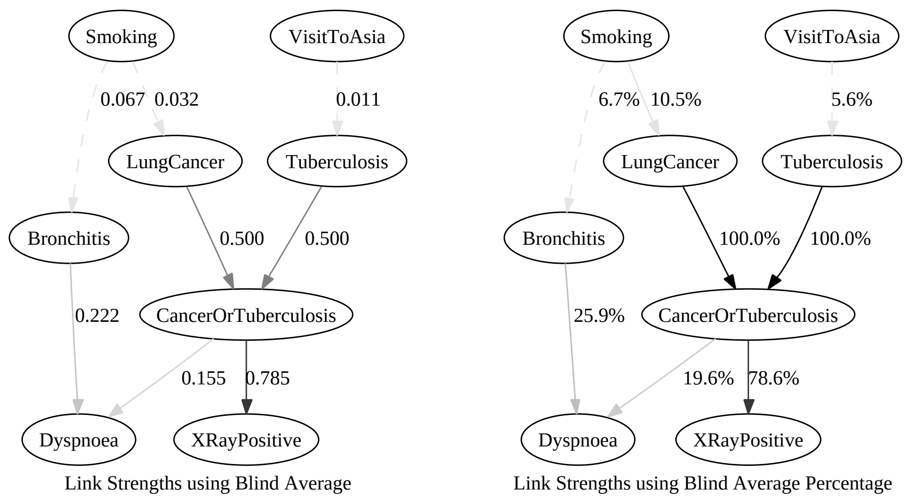

Figure 3: Blind Average Link Strength (left) and Percentage (right) for Asia Model.

the case for values of entropy. This needs to be considered when choosing a threshold for when a connection is considered "strong". While link strength *percentages* are also affected by the nonlinear scale (e.g. a reduction by 50% means that the value of a nonlinear measure is reduced by 50%), it nevertheless appears to provide a better indication since the same nonlinear effect appears in the numerator and denominator and may be 'somewhat' canceled. However, more research on this topic is required to develop practical guidelines.

6.3 True Versus Blind Average Link Strength

Let us consider the *Visit to Asia* network introduced in (Lauritzen and Spiegelhalter 1998). Figures 2 and 3 show True Average Link Strength and Blind Average Link Strength. In the link strength graphs, the value of the link strength is indicated both by the number next to the arrow and by the gray scale of the arrow (if the arrow would otherwise be invisible, a dashed light gray line is used instead).

As indicated by the True Average Percentages on the right of Figure 2 most links are quite strong. All connections except for the one from *Visit to Asia* to *Tuberculosis* can be classified as significant.

The Blind Average Value Percentage for the link from for *Visit to Asia* to *Tuberculosis* on the right of Figure 3 is much higher though, indicating that the reason for the low True Average Percentage is the low probability of state $True$ for *Visit to Asia*. In a nutshell, one could say that in this example *True Average Link Strength (and Percentage) only considers the benefit of the information of variable Visit to Asia for the* **average** *patient. In contrast Blind Average Link Strength (and*

Percentage) considers all patient categories equally – in this case the small group of patients actually having traveled to Asia is given equal weight to the large group not having traveled there – and thus gives **more attention to special cases (small groups) and the value of information of variable** *Visit to Asia* **for that special group**.

This difference is typical for the different viewpoints of True Average and Blind Average. Either viewpoint is valid, but one should be aware of them when choosing a measure for a particular application.

6.4 Detecting Deterministic Relationships

This section illustrates interesting properties of the Link Strength *Percentages* for deterministic functions. By deterministic function we mean that the state of a child is completely known if the states of all of its parents are known, i.e. there is *no* uncertainty involved.

Definition A node Y is a *deterministic child* of its parents, $P_1, \ldots, P_n$, if

$$\forall \text{ states } y, \forall \text{ parent states } p_1, \ldots, p_n : \quad P(y|p_1, \ldots, p_n) \in \{0, 1\}.$$

Proposition 6.1 *If Y is a deterministic child of its parents, then both its True Average and Blind Average Link Strength Percentage from any parent P is 100%:*

$$\forall P \in parents(Y): \quad LS^{true}\%(P \to Y) = 100\%$$
$$\forall P \in parents(Y): \quad LS^{blind}\%(P \to Y) = 100\%.$$

Proof See Appendix A.

The question arises whether the reverse is also true, i.e. if the link strength percentages of all parents to a child are 100% does that imply that the child is deterministic? This is indeed the case for Blind Average Percentage, but not for True Average Percentage, as evident from the following two Propositions.

Proposition 6.2 *If $LS^{blind}\%(P \to Y) = 100\%$ for at least one parent P of a node Y, then Y is a deterministic child of its parents.*

Proof See Appendix A.

Remark: it follows that if $LS^{blind}\%(P \to Y) = 100\%$ for *one* of Y's parents, that the same must hold for *all* of Y's parents.

Proposition 6.3 *Even if $LS^{true}\%(P \to Y) = 100\%$ for all parents P of node Y, then Y is not necessarily a deterministic child of its parents.*

Proof See Appendix A.

To see the usefulness in particular of Proposition 2 we revisit the Visit to Asia Example. Looking at the plot for the Blind Average Link Strength Percentage (right plot in Figure 3) immediately shows that *CancerOrTuberculosis* is a deterministic child of its parents – which, admittedly, in this case could have been guessed from its name, too. Other cases are less obvious, in particular if a large network is learned from data and this property can be helpful to identify deterministic and *nearly* deterministic child nodes.

7 Conclusions

We believe that Link Strength measures will play an important role in the context of constraint-based structure learning algorithms **to derive hypotheses of a system's primary causal pathways** from data. One problem when using constraint-based structure learning algorithms is the generally large number of Markov-equivalent DAGs returned by each algorithm. One way in which link strength may be helpful is that it could help one reduce the number of models to look at. Although the concept of link strength was already proposed and demonstrated for Bayesian Networks back in 1992, it has yet to find widespread use in the visualization and evaluation of Bayesian networks learned from data. We hope that this introductory paper, through its review of existing link strength definitions, definitions of Blind Average Link Strength and Link Strength percentages, and its examples and discussion of link strength properties, will increase the use of link strength for those and other purposes.

A Appendix: Proofs for Propositions 6.1 to 6.3

Proof of Proposition 6.1: If node Y is a deterministic child of its parents then it follows $U(Y|X, \mathbf{Z}) = 0$ and $\hat{U}(Y|X, \mathbf{Z}) = 0$ in the definitions of True/Blind Average Link Strengths, which then yields the desired result.

Proof of Proposition 6.2: From $LS^{blind}\%(P \to Y) = 100\%$ follows $\hat{U}(Y|X, \mathbf{Z}) = 0$, thus

$$\sum_{x,y,\mathbf{z}} P(y|x, \mathbf{z}) \log_2 P(y|x, \mathbf{z}) = 0.$$

Each term $P(y|x, \mathbf{z}) \log_2 P(y|x, \mathbf{z})$ is positive and vanishes if and only if $P(y|x, \mathbf{z}) = 0$ or $P(y|x, \mathbf{z}) = 1$. Thus in order for the whole sum to vanish, we must have $\forall x, y, \mathbf{z} : \quad P(y|x, \mathbf{z}) \in \{0, 1\}$. Thus Y is a deterministic child of its parents.

Proof of Proposition 6.3: The following degenerate case serves as a counter example. Y has two parents, X, Z, which each can only take states 0 and 1. Let us say that $x = 0$ and $z = 0$ always, thus $P(x = 0, z = 0) = 1$ and $P(x, z) = 0$ otherwise. Define $Y = (x + z) * (\text{random number})$, then $U(Y|x = 0, z = 0) = 0$ and $U(Y|x, z) \neq 0$ otherwise. Thus all products $P(x, z)U(Y|x, z)$ vanish and $U(Y|X, Z) = 0$, although Y is clearly *not* a deterministic child of its parents.

Comment: One may argue that The inability of the True Average Link Strength Percentage to guarantee that a node is a deterministic child comes from the fact that the definition of whether a child is deterministic is *independent of the joint probability of the node's parents*, while True Average Link Strength Percentage *disregards parent state combinations with zero joint probability*. Thus one may argue that this difference is philosophical in nature and that True Average Link Strength Percentage is also a good indicator for deterministic relationships. Nevertheless, it is more prudent to use Blind Average Link Strength Percentage for that purpose.

References

Boerlage, B., 1992, "Link Strengths in Bayesian Networks ," Master's thesis, Dept. of Computer Science, The University of British Columbia.

Ebert-Uphoff, I., 2007, "Measuring Connection Strengths and Link Strengths in Discrete Bayesian Networks ," Technical Report GT-IIC-07-01, Georgia Institute of Technology.

Friedman, N., Nachman, I., and Peér, D., 1999, "Learning Bayesian Network Structure from Massive Datasets: The "Sparse Candidate" Algorithm ," In *Uncertainty in Artificial Intelligence (UAI'99)*.

Jitnah, N., 1999, *Using Mutual Information for Approximate Evaluation of Bayesian Networks* PhD thesis, School of Computer Science and Software Engineering, Monash University, Clayton, Victoria, Australia.

Lacave, C. and Diez, F., 2004, "The Elvira GUI: a tool for generating explanations for Bayesian networks ," *submitted journal paper*.

Lauritzen, S. L. and Spiegelhalter, S., 1998, "Local computations with probabilities on graphical structures and their application to expert systems ," *J. Royal Statistics Society B*, 50(2):157–194.

Nicholson, A. and Jitnah, N., 1997, "Treenets: A framework for anytime evaluation of belief networks ," In *First International Joint Conference on Qualitative and Quantitative Pracical Reasoning (ECSQARU-FAPR'97)*. Springer.

Nicholson, A. and Jitnah, N., 1998, "Using Mutual Information to determine Relevance in Bayesian Networks ," In *5th Pacific Rim International Conference on Artificial Intelligence (PRICAI'98)*, pages 399–410, Singapore. Springer.

Pappas, A. and Gillies, D., 2002, "A New Measure for the Accuracy of a Bayesian Network ," In *Second Mexican International Conference in Artificial Intelligence (MICAI 2002)*, Merida, Yucatan, Mexico.

Pearl, J., 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Interference* Morgan Kaufman Publishers, San Mateo, CA, revised second printing edition.

Shannon, C. and Warren, W., 1949, *The Mathematical Theory of Communication* University of Illinois Press, Urbana and Chicago.