

Bayesian Belief Network

Dr. Saed Sayad

University of Toronto

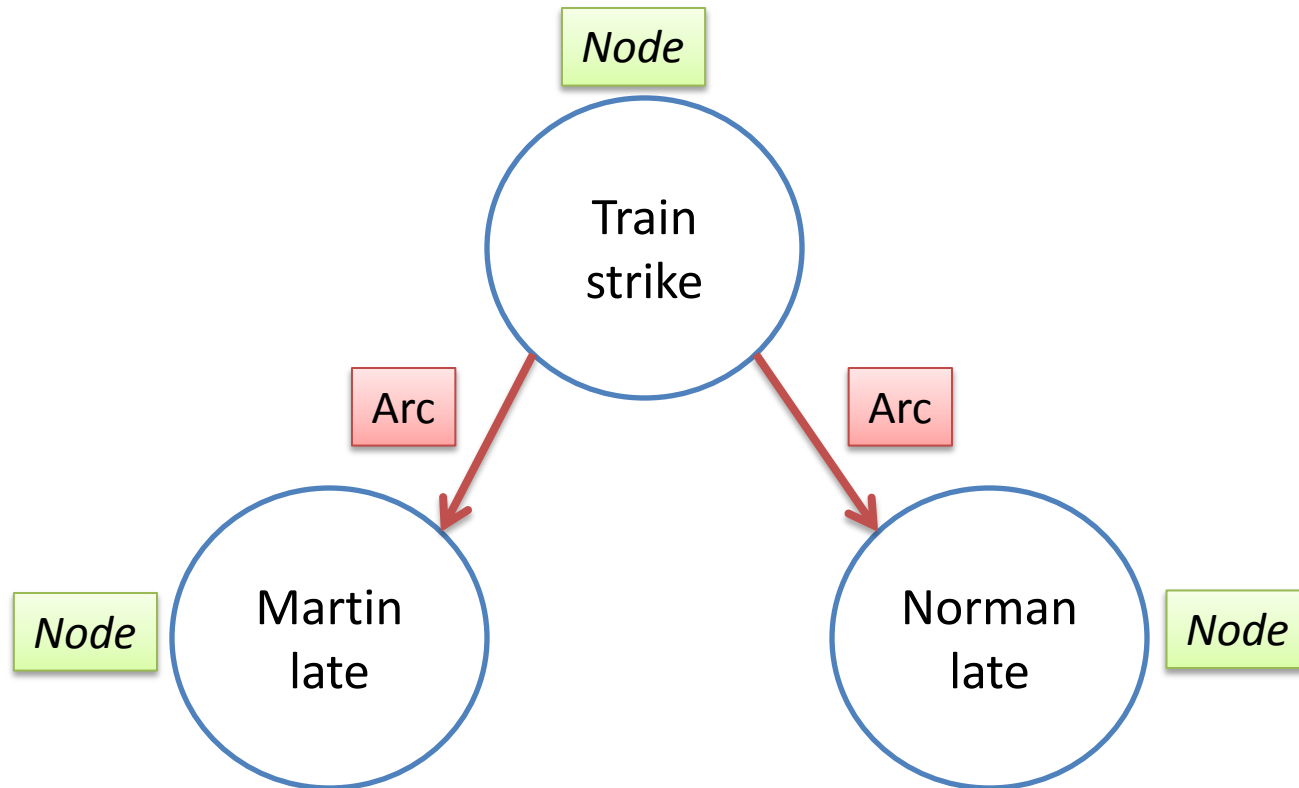
2010

saed.sayad@utoronto.ca

BBN: Definition

- A BBN is a special type of diagram (called a directed graph) together with an associated set of probability tables.
- The graph consists of nodes and arcs.
- The nodes represent variables, which can be discrete or continuous.
- The arcs represent causal relationships between variables.

BBN: Example



Node Probability		Train strike	
		True	False
Norman late	True	0.8	0.1
	False	0.2	0.9

BBN: Key Features

- BBNs enable us to model and reason about *uncertainty*.
- BBNs accommodate both ***subjective*** probabilities and probabilities based on ***objective*** data.
- The most important use of BBNs is in ***revising probabilities*** in the light of actual observations of events.

What is Probability?

- Probability theory is the body of knowledge that enables us to reason formally about uncertain events.
- The probability P of an uncertain event A , written $P(A)$, is defined by the frequency of that event based on previous observations.
- This is called *frequency* based probability.

Probability Axioms

1. $P(a)$ should be a number between 0 and 1.
2. If a represents a certain event then $P(a)=1$.
3. If a and b are mutually exclusive events then $P(a \text{ or } b)=P(a)+P(b)$. Mutually exclusive means that they cannot both be true at the same time.

Probability Distributions

- There is a *variable* called A .
- The variable A can have many states:

$$A = \{a_1, a_2, \dots, a_n\}$$

$$\sum_{i=1}^n P(a_i) = 1$$

- We can think of an event as just one state of the variable A .
- The ***probability distribution*** of A , written $P(A)$, is simply the set of values $\{p(a_1), p(a_2), \dots, p(a_n)\}$

Joint Events

- Suppose there are two events A and B that:

$A = \{a_1, a_2, a_3\}$ where $a_1=0, a_2=1, a_3=">1"$

$B = \{b_1, b_2, b_3\}$ where $b_1=0, b_2=1, b_3=">1"$

- The joint event A and B :

$$P(A, B) = P(a_i, b_j) = \left\{ \begin{array}{l} P(a_1, b_1), P(a_1, b_2), P(a_1, b_3), \\ P(a_2, b_1), P(a_2, b_2), P(a_2, b_3), \\ P(a_3, b_1), P(a_3, b_2), P(a_3, b_3) \end{array} \right\}$$

$$i = 1 \dots n; \quad j = 1 \dots m$$

- $P(A, B)$ is called the **joint probability distribution** of A and B . The general form:

$$P(A_1, \dots, A_n) = \prod_{i=1}^n P(A_i \mid \text{parents}(A_i))$$

Marginalisation

- If we know the joint probability distribution $P(A,B)$ then we can calculate $P(A)$ by a formula called ***marginalisation***:

$$P(a) = \sum_i P(a, b_i)$$

- This is because the events (a, b_1) are mutually exclusive.
- When we calculate $P(A)$ in this way from the joint probability distribution we say that the variable B is *marginalised* out of $P(A,B)$.

Belief Measure

- In general, a person belief in a statement \mathbf{a} will depend on some body of knowledge K . We write this as $P(\mathbf{a} | K)$.
- The expression $P(\mathbf{a} | K)$ thus represents a *belief measure*.
- Sometimes, for simplicity, when K remains constant we just write $P(\mathbf{a})$, but you must be aware that this is a simplification.

Conditional Probability

- The notion of degree of belief $P(A | K)$ is an uncertain event A is ***conditional*** on a body of knowledge K .
- In general, we write $P(A | B)$ to represent a belief in A under the assumption that B is known.
- Strictly speaking, $P(A | B)$ is a shorthand for the expression $P(A | B, K)$ where K represents all other relevant information.
- Only when all other information is irrelevant can we really write $P(A | B)$.
- The traditional approach to defining conditional probability is via joint probabilities:

$$P(A | B) = \frac{P(A, B)}{P(B)}$$

Bayes' Rule

- It is possible to define $P(A|B)$ without reference to the joint probability $P(A,B)$ by rearranging the conditional probability formula as follows:

$$P(A|B)P(B) = P(A,B) \text{ \& } P(B|A)P(A) = P(A,B) \Rightarrow$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' Rule

Bayes' Rule (cont.)

- It is common to think of Bayes' rule in terms of updating our belief about a hypothesis A in the light of new evidence B .
- Specifically, our posterior belief $P(A|B)$ is calculated by multiplying our prior belief $P(A)$ by the likelihood $P(B|A)$ that B will occur if A is true.
- The power of Bayes' rule is that in many situations where we want to compute $P(A|B)$ it turns out that it is difficult to do so directly, yet we might have direct information about $P(B|A)$. Bayes' rule enables us to compute $P(A|B)$ in terms of $P(B|A)$.

Likelihood Ratio

- We have seen that Bayes' rule computes $P(A | B)$ in terms of $P(B | A)$. The expression $P(B | A)$ is called the likelihood of A .
- If A has two values a_1 and a_2 then the following ratio called *likelihood ratio*.

$$\frac{P(B | a_1)}{P(B | a_2)}$$

Chain Rule

- We can rearrange the formula for conditional probability to get the so-called **product rule**:

$$P(A, B) = P(A | B)P(B)$$

- We can extend this for three variables:

$$P(A, B, C) = P(A | B, C)P(B, C) = P(A | B, C)P(B | C)P(C)$$

- And to n variables:

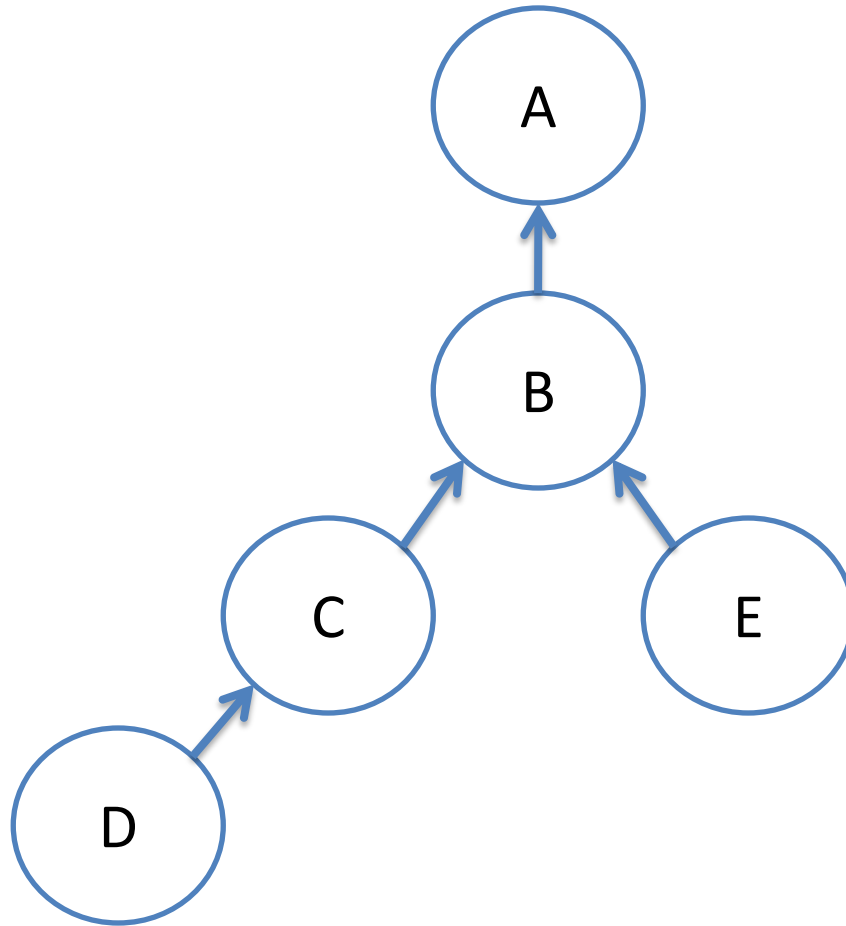
$$P(A_1, A_2, \dots, A_n) = P(A_1 | A_2, \dots, A_n)P(A_2 | A_3, \dots, A_n), \dots, P(A_{n-1} | A_n)P(A_n)$$

- In general we refer to this as the **chain rule**.

Independence and Conditional Independence

- The conditional probability of A given B is represented by $P(A | B)$. The variables A and B are said to be ***independent*** if $P(A) = P(A | B)$ or alternatively if $P(A, B) = P(A)P(B)$.
- A and B are ***conditionally independent*** given C if $P(A | C) = P(A | B, C)$.

Chain Rule: Example



Joint probability
distribution

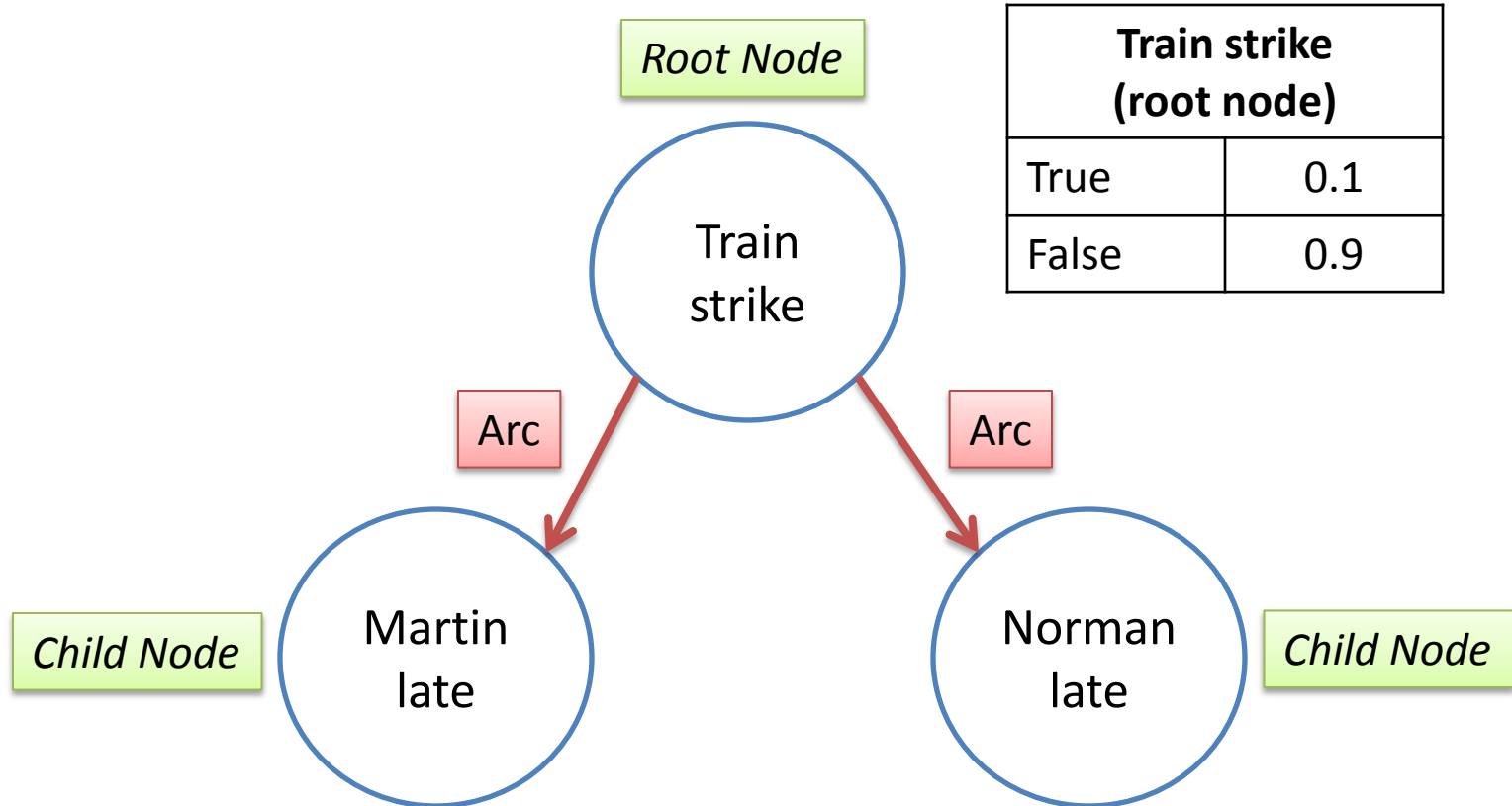
$$P(A,B,C,D,E)=P(A|B,C,D,E)*P(B|C,D,E)*P(C|D,E)*P(D|E)*P(E)$$

$$P(A,B,C,D,E)=P(A|B)*P(B|C,D)*P(C|D)*P(D)*P(E)$$

Bayesian Belief Network

- A BBN is a special type of diagram (called a directed graph) together with an associated set of probability tables.
- The graph consists of nodes and arcs.
- The nodes represent variables, which can be discrete or continuous.
- The arcs represent causal relationships between variables.

BBN: Example



Train strike (root node)	
True	0.1
False	0.9

Node Conditional Probability		Train strike	
		True	False
Martin late	True	0.6	0.5
	False	0.4	0.5

Node Conditional Probability		Train strike	
		True	False
Norman late	True	0.8	0.1
	False	0.2	0.9

Entering Evidence and Propagation (1)

- Having entered the probabilities we can now use Bayesian probability to do various types of analysis.
- For example, we want to calculate the (unconditional) probability that 'Norman is late'.

$$\begin{aligned} P(\text{Norman late}) &= P(\text{Norman late} \mid \text{Train strike}) * P(\text{Train strike}) + \\ & P(\text{Norman late} \mid \text{No train strike}) * P(\text{No train strike}) = \\ & (0.8 * 0.1) * (0.1 * 0.9) = 0.17 \end{aligned}$$

This is called the *marginal probability*.
$$P(a) = \sum_i P(a, b_i)$$

Entering Evidence and Propagation (2)

- The most important use of BBNs is in *revising probabilities* in the light of actual observations of events.
- Suppose, for example, that we *know* there is a train strike. In this case we can **enter the evidence** that ‘train strike=true’.
- The conditional probability tables already tell us the revised probabilities for Norman being late (0.8) and Martin being late (0.6).

Entering Evidence and Propagation (3)

- Suppose, however, that we do not know if there is a train strike but do know that Norman is late.
- Then we can enter the evidence that 'Norman late=true' and we can use this observation to determine:
 - a) The (revised) probability that there is a train strike, and
 - b) The (revised) probability that Martin will be late.

Entering Evidence and Propagation (4)

- a) $P(\text{Train strike} | \text{Norman late}) = P(\text{Norman late} | \text{Train strike}) * P(\text{Train strike}) / P(\text{Norman late}) = (0.8 * 0.1) * 0.17 = 0.47$
- b) $P(\text{Martin late}) = P(\text{Martin late} | \text{Train strike}) * P(\text{Train strike}) + P(\text{Martin late} | \text{No train strike}) * P(\text{No train strike}) = (0.6 * 0.47) + (0.5 * 0.53) = 0.55$
- ❖ When we enter evidence and use it to update the probabilities in this way we call it **propagation**.

Hard and Soft evidence

- **Hard evidence (instantiation)** for a node X is evidence that the state of X is definitely a particular value.
- **Soft evidence** for a node X is any evidence that enables us to update the prior probability values for the states of X .

BBN Connections

1. Serial Connection

- Any evidence entered at the beginning of the connection can be transmitted along the directed path providing that no intermediate node on the path is instantiated (which thereby blocks further transmission).

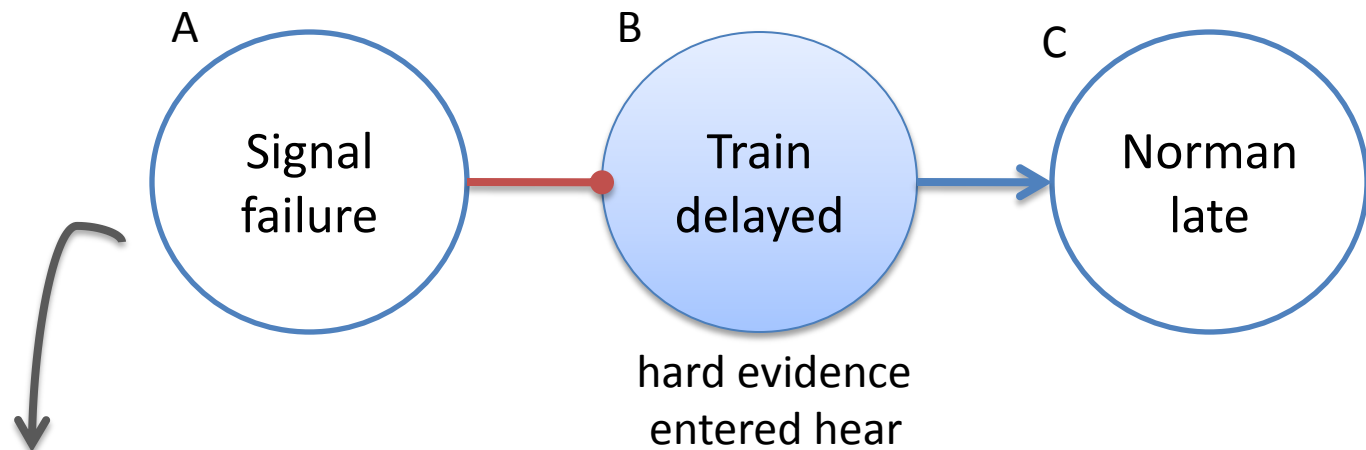
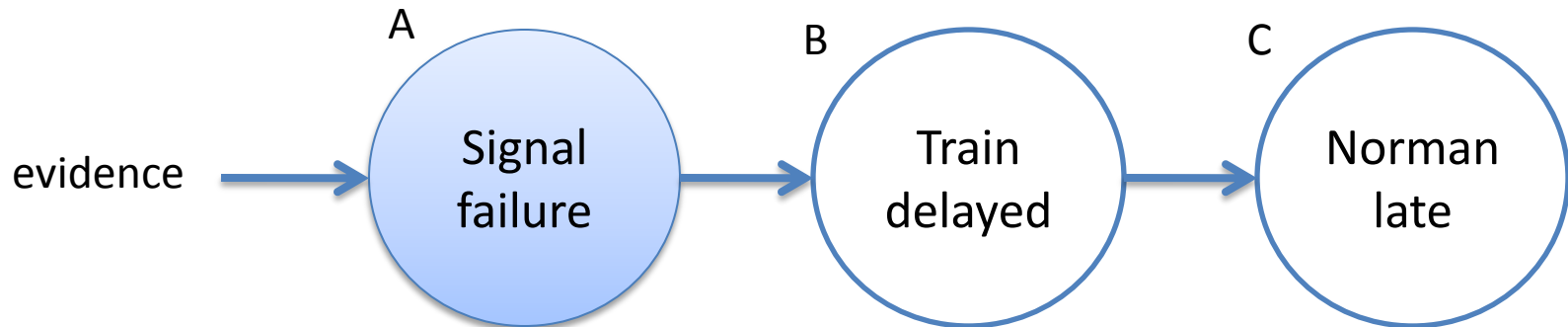
2. Diverging Connection

- Evidence can be transmitted between two child nodes of the same parent providing that the parent is not instantiated.

3. Converging Connection

- Evidence can only be transmitted between two parents when the child (converging) node has received some evidence (which can be soft or hard).

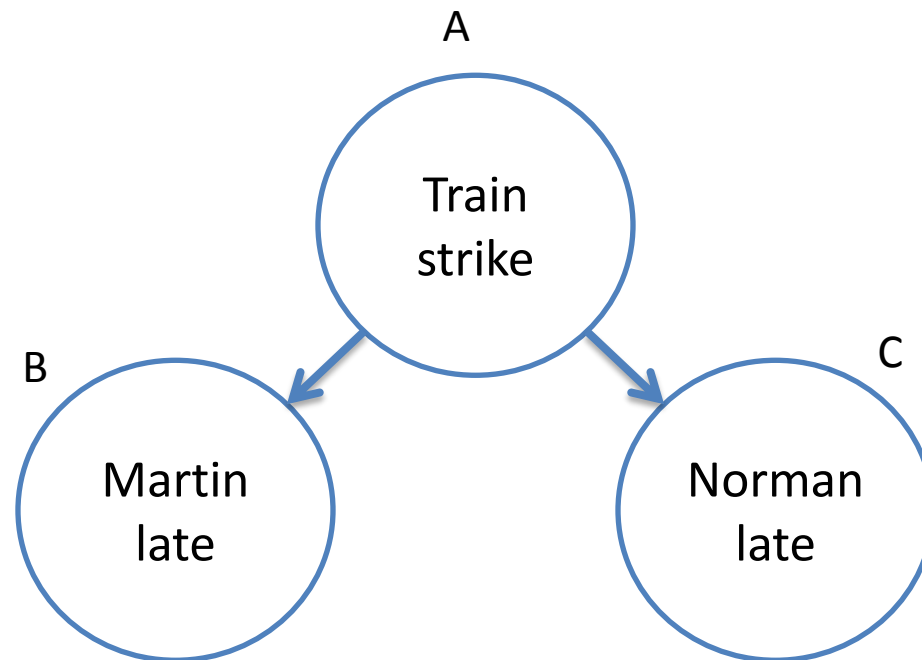
BBN: Serial Connection



The evidence from A cannot be transmitted to C because B **blocks** the channel.

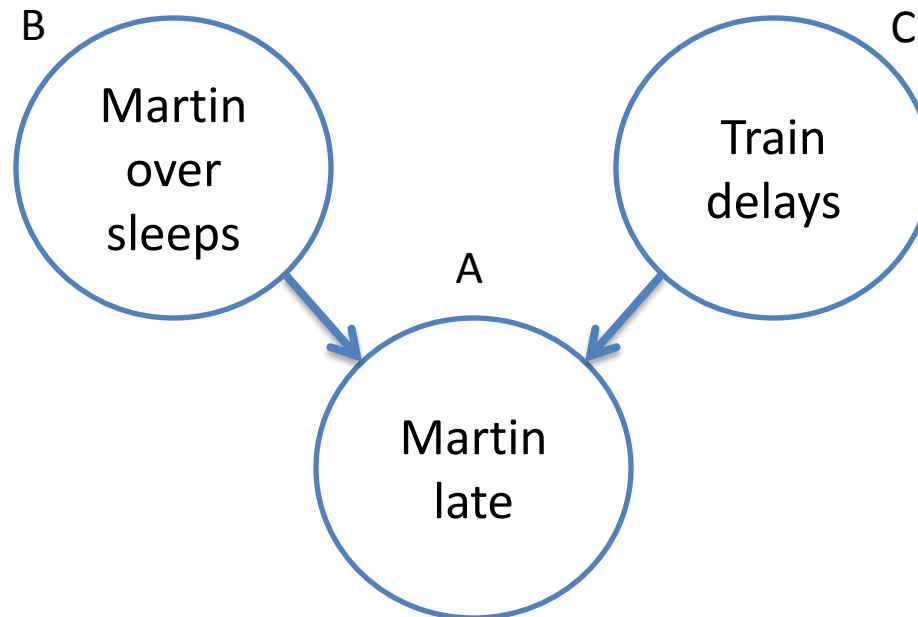
In summary, in a serial connection evidence can be transmitted from A to C unless B is instantiated. Formally we say **that A and C are d-separated given B.**

BBN: Diverging Connection



Evidence can be transmitted from B to through a diverging connection A, unless A is instantiated. We say that **B and C are d-separated given A.**

BBN: Converging Connection



Evidence can be transmitted from B to through a converging connection A, if anything is known about A. We say that **B and C are conditionally dependent on A**. In the other word, **B and C are only d-separated when there is no evidence about A**.

d-separated and d-connected

- Two nodes X and Y in a BBN are d-separated if, for all paths between X and Y , there is an intermediate node A for which either:
 - The connection is serial or diverging and the state of A is known for certain; or
 - The connection is diverging and neither A nor any of its descendants have received any evidence at all.
- In general, two nodes which are not d-separated are said to be **d-connected**.

References

- <http://www.dcs.qmul.ac.uk/~norman/BBNs/BBNs.htm>
- http://en.wikipedia.org/wiki/Bayesian_network