Korb & Nicholson, 2004

2.4.4.1 Causal chains

Consider a causal chain of three nodes, where A causes B which in turn causes C, as shown in Figure 2.4(a). In our medical diagnosis example, one such causal chain is "smoking causes cancer which causes dyspnoea." Causal chains give rise to conditional independence, such as for Figure 2.4(a):

$$P(C|A \wedge B) = P(C|B)$$

This means that the probability of C, given B, is exactly the same as the probability of C, given both B *and* A. Knowing that A has occurred doesn't make any difference to our beliefs about C *if we already know that B has occurred.* We also write this conditional independence as: $A \perp\!\!\!\perp C|B$.

In Figure 2.1(a), the probability that someone has dyspnoea depends directly only on whether they have cancer. If we don't know whether some woman has cancer, but we do find out she is a smoker, that would increase our belief both that she has cancer and that she suffers from shortness of breath. However, if we already *knew* she had cancer, then her smoking wouldn't make any difference to the probability of dyspnoea. That is, dyspnoea is conditionally independent of being a smoker *given* the patient has cancer.

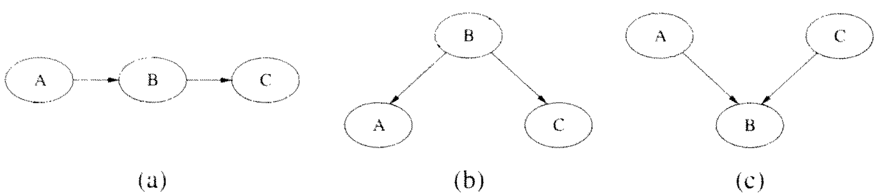

FIGURE 2.4
(a) Causal chain; (b) common cause; (c) common effect.

2.4.4.2 Common causes

Two variables A and C having a common cause B is represented in Figure 2.4(b). In our example, cancer is a common cause of the two symptoms, a positive X-ray result and dyspnoea. Common causes (or common ancestors) give rise to the same conditional independence structure as chains:

$$P(C|A \wedge B) = P(C|B) \equiv A \perp\!\!\!\perp C|B$$

If there is no evidence or information about cancer, then learning that one symptom is present will increase the chances of cancer which in turn will increase the probability of the other symptom. However, if we already know about cancer, then an additional positive X-ray won't tell us anything new about the chances of dyspnoea.

2.4.4.3 Common effects

A common effect is represented by a network v-structure, as in Figure 2.4(c). This represents the situation where a node (the effect) has two causes. Common effects (or their descendants) produce the exact opposite conditional independence structure to that of chains and common causes. That is, the parents are marginally independent ($A \perp\!\!\!\perp C$), but become dependent given information about the common effect (i.e., they are **conditionally dependent**):

$$P(A|C \wedge B) \neq P(A|C) \equiv \neg(A \perp\!\!\!\perp C|B)$$

Thus, if we observe the effect (e.g., cancer), and then, say, we find out that one of the causes is absent (e.g., the patient does not smoke), this *raises* the probability of the other cause (e.g., that he lives in a polluted area) — which is just the inverse of explaining away.

Compactness again

So we can now see *why* building networks with an order violating causal order can, and generally will, lead to additional complexity in the form of extra arcs. Consider just the subnetwork { *Pollution, Smoker, Cancer* } of Figure 2.1. If we build the subnetwork in that order we get the simple v-structure *Pollution* $\rightarrow$ *Smoker* $\leftarrow$ *Cancer*. However, if we build it in the order < *Cancer, Pollution, Smoker* >, we will first get *Cancer* $\rightarrow$ *Pollution*, because they are dependent. When we add *Smoker*, it will be dependent upon *Cancer*, because in reality there is a direct dependency there. But we shall also have to add a spurious arc to *Pollution*, because otherwise *Cancer* will act as a common cause, inducing a spurious dependency between *Smoker* and *Pollution*; the extra arc is necessary to reestablish marginal independence between the two.

2.4.5 d-separation

We have seen how Bayesian networks represent conditional independencies and how these independencies affect belief change during updating. The conditional independence in $A \perp\!\!\!\perp C|B$ means that knowing the value of B **blocks** information about C being relevant to A, and vice versa. Or, in the case of Figure 2.4(c), *lack* of information about B blocks the relevance of C to A, whereas learning about B **activates** the relation between C and A.

These concepts apply not only between pairs of nodes, but also between sets of nodes. More generally, given the Markov property, it is possible to determine whether a set of nodes **X** is independent of another set **Y**, given a set of evidence nodes **E**. To do this, we introduce the notion of **d-separation** (from **direction-dependent separation**).

Definition 2.1 Path (Undirected Path) *A* **path** *between two sets of nodes* **X** *and* **Y** *is any sequence of nodes between a member of* **X** *and a member of* **Y** *such that every adjacent pair of nodes is connected by an arc (regardless of direction) and no node appears in the sequence twice.*

Definition 2.2 Blocked path *A path is* **blocked**, *given a set of nodes* **E**, *if there is a node Z on the path for which at least one of three conditions holds:*

1. *Z is in* **E** *and Z has one arc on the path leading in and one arc out (chain).*
2. *Z is in* **E** *and Z has both path arcs leading out (common cause).*
3. *Neither Z nor any descendant of Z is in* **E**, *and both path arcs lead in to Z (common effect).*

Definition 2.3 d-separation *A set of nodes* **E d-separates** *two other sets of nodes* **X** *and* **Y** *if every path from a node in* **X** *to a node in* **Y** *is* **blocked** *given* **E**.

If **X** and **Y** are **d-separated** by **E**, then **X** and **Y** are **conditionally independent** given **E** (given the Markov property). Examples of these three blocking situations are shown in Figure 2.5. Note that we have simplified by using single nodes rather than sets of nodes; also note that the evidence nodes **E** are shaded.

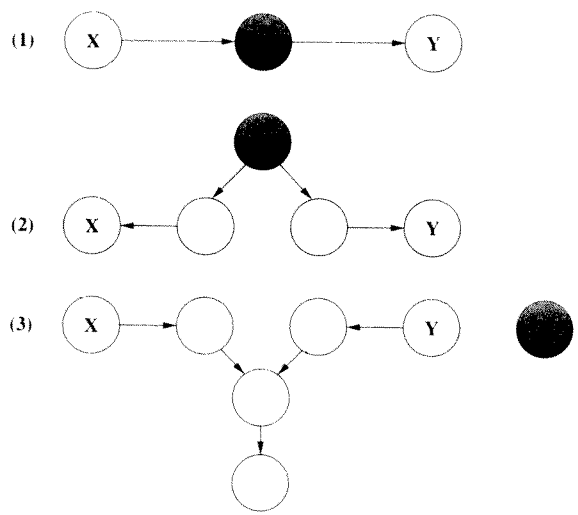

FIGURE 2.5

Examples of the three types of situations in which the path from X to Y can be blocked, given evidence E. In each case, X and Y are **d-separated** by E.

Consider d-separation in our cancer diagnosis example of Figure 2.1. Suppose an observation of the Cancer node is our evidence. Then:

1. P is d-separated from X and D. Likewise, S is d-separated from X and D (blocking condition 1).

2. While X is d-separated from D (condition 2).
3. However, if C had not been observed (and also not X or D), then S would have been d-separated from P (condition 3).

2.5 More examples

In this section we present further simple examples of BN modeling from the literature. We encourage the reader to work through these examples using BN software (see §B.4).

2.5.1 Earthquake

Example statement: *You have a new burglar alarm installed. It reliably detects burglary, but also responds to minor earthquakes. Two neighbors, John and Mary, promise to call the police when they hear the alarm. John always calls when he hears the alarm, but sometimes confuses the alarm with the phone ringing and calls then also. On the other hand, Mary likes loud music and sometimes doesn't hear the alarm. Given evidence about who has and hasn't called, you'd like to estimate the probability of a burglary (from [217]).*

A BN representation of this example is shown in Figure 2.6. All the nodes in this BN are Boolean, representing the true/false alternatives for the corresponding propositions. This BN models the assumptions that John and Mary do not perceive a burglary directly and they do not feel minor earthquakes. There is no explicit representation of loud music preventing Mary from hearing the alarm, nor of John's confusion of alarms and telephones; this information is summarized in the probabilities in the arcs from *Alarm* to *JohnCalls* and *MaryCalls*.

2.5.2 Metastatic cancer

Example statement: *Metastatic cancer is a possible cause of brain tumors and is also an explanation for increased total serum calcium. In turn, either of these could explain a patient falling into a coma. Severe headache is also associated with brain tumors. (This example has a long history in the literature [51, 217, 262].)*

A BN representation of this metastatic cancer example is shown in Figure 2.7. All the nodes are Booleans. Note that this is a *graph*, not a tree, in that there is more than one path between the two nodes M and C (via S and B).

2.5.3 Asia

Example Statement: *Suppose that we wanted to expand our original medical diagnosis example to represent explicitly some other possible causes of shortness of*