Even in a locally structured domain, we will get a compact Bayesian network only if we choose the node ordering well. What happens if we happen to choose the wrong order? Consider the burglary example again. Suppose we decide to add the nodes in the order \textit{MaryCalls, JohnCalls, Alarm, Burglary, Earthquake}. We then get the somewhat more complicated network shown in Figure 14.3(a). The process goes as follows:

- **Adding \textit{MaryCalls}**: No parents.
- **Adding \textit{JohnCalls}**: If Mary calls, that probably means the alarm has gone off, which of course would make it more likely that John calls. Therefore, \textit{JohnCalls} needs \textit{MaryCalls} as a parent.
- **Adding \textit{Alarm}**: Clearly, if both call, it is more likely that the alarm has gone off than if just one or neither calls, so we need both \textit{MaryCalls} and \textit{JohnCalls} as parents.
- **Adding \textit{Burglary}**: If we know the alarm state, then the call from John or Mary might give us information about our phone ringing or Mary’s music, but not about burglary:
  \[
  P(\text{Burglary} \mid \text{Alarm, John Calls, Mary Calls}) = P(\text{Burglary} \mid \text{Alarm}).
  \]
  Hence we need just \textit{Alarm} as parent.
- **Adding \textit{Earthquake}**: If the alarm is on, it is more likely that there has been an earthquake. (The alarm is an earthquake detector of sorts.) But if we know that there has been a burglary, then that explains the alarm, and the probability of an earthquake would be only slightly above normal. Hence, we need both \textit{Alarm} and \textit{Burglary} as parents.

The resulting network has two more links than the original network in Figure 14.2 and requires three more probabilities to be specified. What’s worse, some of the links represent tenuous relationships that require difficult and unnatural probability judgments, such as as-
Section 14.2, The Semantics of Bayesian Networks 517

sessing the probability of Earthquake, given Burglary and Alarm. This phenomenon is quite general and is related to the distinction between causal and diagnostic models introduced in Section 13.5.1 (see also Exercise 8.13). If we try to build a diagnostic model with links from symptoms to causes (as from MaryCalls to Alarm or Alarm to Burglary), we end up having to specify additional dependencies between otherwise independent causes (and often between separately occurring symptoms as well). If we stick to a causal model, we end up having to specify fewer numbers, and the numbers will often be easier to come up with. In the domain of medicine, for example, it has been shown by Tversky and Kahneman (1982) that expert physicians prefer to give probability judgments for causal rules rather than for diagnostic ones.

Figure 14.3(b) shows a very had node ordering MaryCalls, JohnCalls, Earthquake, Burglary, Alarm. This network requires 31 distinct probabilities to be specified—exactly the same number as the full joint distribution. It is important to realize, however, that any of the three networks can represent exactly the same joint distribution. The last two versions simply fail to represent all the conditional independence relationships and hence end up specifying a lot of unnecessary numbers instead.

14.2.2 Conditional independence relations in Bayesian networks

We have provided a "numerical" semantics for Bayesian networks in terms of the representation of the full joint distribution, as in Equation (14.2). Using this semantics to derive a method for constructing Bayesian networks, we were led to the consequence that a node is conditionally independent of its other predecessors, given its parents. It turns out that we can also go in the other direction. We can start from a "topological" semantics that specifies the conditional independence relationships encoded by the graph structure, and from this we can derive the "numerical" semantics. The topological semantics specifies that each variable is conditionally independent of its non-descendants, given its parents. For example, in Figure 14.2, JohnCalls is independent of Burglary, Earthquake, and Mary Calls given the value of Alarm. The definition is illustrated in Figure 14.4(a). From these conditional independence assertions and the interpretation of the network parameters \( \theta(X_i \mid \text{Parents}(X_i)) \) as specifications of conditional probabilities \( P(X_i \mid \text{Parents}(X_i)) \), the full joint distribution given in Equation (14.2) can be reconstructed. In this sense, the "numerical" semantics and the "topological" semantics are equivalent.

Another important independence property is implied by the topological semantics: a node is conditionally independent of all other nodes in the network, given its parents, children, and children's parents—that is, given its Markov blanket (Exercise 14.7 asks you to prove this.) For example, Burglary is independent of JohnCalls and MaryCalls, given Alarm and Earthquake. This property is illustrated in Figure 14.4(b).

There is also a general topological criterion called d-separation for deciding whether a set of nodes \( X \) is conditionally independent of another set \( Y \), given a third set \( Z \). The criterion is rather complicated and is not needed for deriving the algorithms in this chapter, so we omit it. Details may be found in Pearl (1988) or Darwiche (2009). Shachter (1998) gives a more intuitive method of ascertaining d-separation.