Algorithm 3.2 Procedure to build a minimal I-map given an ordering

Procedure Build-Minimal-I-Map (
  $X_1, \ldots, X_n$ // an ordering of random variables in $\mathcal{X}$
  $\mathcal{I}$ // Set of independencies
)
  Set $\mathcal{G}$ to an empty graph over $\mathcal{X}$
  for $i = 1, \ldots, n$
    $\boldsymbol{U} \leftarrow \{X_1, \ldots, X_{i-1}\}$ // $\boldsymbol{U}$ is the current candidate for parents of X_i
    for $\boldsymbol{U}' \subseteq \{X_1, \ldots, X_{i-1}\}$
      if $\boldsymbol{U}' \subset \boldsymbol{U}$ and $(X_i \perp \{X_1, \ldots, X_{i-1}\} - \boldsymbol{U}' \mid \boldsymbol{U}') \in \mathcal{I}$ **then**
        $\boldsymbol{U} \leftarrow \boldsymbol{U}'$
      // At this stage $\boldsymbol{U}$ is a minimal set satisfying $(X_i \perp \{X_1, \ldots, X_{i-1}\} - \boldsymbol{U} \mid \boldsymbol{U})$
      // Now set $\boldsymbol{U}$ to be the parents of X_i
    for $X_j \in \boldsymbol{U}$
      Add $X_j \rightarrow X_i$ to $\mathcal{G}$
  return $\mathcal{G}$

complete graph is an I-map for any distribution, yet it does not reveal any of the independence structure in the distribution. However, examples such as this one are not very interesting. The graph that we used as an I-map is clearly and trivially unrepresentative of the distribution, in that there are edges that are obviously redundant. This intuition leads to the following definition, which we also define more broadly:

Definition 3.13
minimal I-map

A graph $\mathcal{K}$ is a minimal I-map *for a set of independencies $\mathcal{I}$ if it is an I-map for $\mathcal{I}$, and if the removal of even a single edge from $\mathcal{K}$ renders it not an I-map.* ■

This notion of an I-map applies to multiple types of graphs, both Bayesian networks and other types of graphs that we will encounter later on. Moreover, because it refers to a set of independencies $\mathcal{I}$, it can be used to define an I-map for a distribution P, by taking $\mathcal{I} = \mathcal{I}(P)$, or to another graph $\mathcal{K}'$, by taking $\mathcal{I} = \mathcal{I}(\mathcal{K}')$.

Recall that definition 3.5 defines a Bayesian network to be a distribution P that factorizes over $\mathcal{G}$, thereby implying that $\mathcal{G}$ is an I-map for P. It is standard to restrict the definition even further, by requiring that $\mathcal{G}$ be a minimal I-map for P.

variable ordering

How do we obtain a minimal I-map for the set of independencies induced by a given distribution P? The proof of the factorization theorem (theorem 3.1) gives us a procedure, which is shown in algorithm 3.2. We assume we are given a predetermined *variable ordering*, say, $\{X_1, \ldots, X_n\}$. We now examine each variable X_i, $i = 1, \ldots, n$ in turn. For each X_i, we pick some minimal subset $\boldsymbol{U}$ of $\{X_1, \ldots, X_{i-1}\}$ to be X_i's parents in $\mathcal{G}$. More precisely, we require that $\boldsymbol{U}$ satisfy $(X_i \perp \{X_1, \ldots, X_{i-1}\} - \boldsymbol{U} \mid \boldsymbol{U})$, and that no node can be removed from $\boldsymbol{U}$ without violating this property. We then set $\boldsymbol{U}$ to be the parents of X_i.

The proof of theorem 3.1 tells us that, if each node X_i is independent of $X_1, \ldots, X_{i-1}$ given its parents in $\mathcal{G}$, then P factorizes over $\mathcal{G}$. We can then conclude from theorem 3.2 that $\mathcal{G}$ is an I-map for P. By construction, $\mathcal{G}$ is minimal, so that $\mathcal{G}$ is a minimal I-map for P.