4 Graphical Model Evaluation, Construction, and Conversion

4.1 Recap: Comparison between Directed and Undirected Graphical Models

So far we've introduced both directed and undirected graphical models as ways to capture conditional independence relations satisfied by families of probability distributions. Let us briefly review and compare the two:

- *Factorization:* Directed graphical models factorize as $p_{\mathsf{x}}(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid x_{\pi_i})$, where π_i is the set of parents of i in the given DAG. Undirected graphical models factorize as $p_{\mathsf{x}}(\mathbf{x}) = \frac{1}{Z} \prod_{\mathcal{C} \in \mathrm{cl}^*(\mathcal{G})} \psi_{\mathcal{C}}(x_{\mathcal{C}})$, where $\mathrm{cl}^*(\mathcal{G})$ is the set of maximal cliques, $\psi_{\mathcal{C}}$ is some nonnegative function, and Z is a normalizing constant.

- *Global Markov Property:* Both directed and undirected graphical models have a version of the global Markov property in order to relate graph separation to conditional independencies. In directed graphs, a distribution satisfies the global Markov property: $\mathsf{x}_{\mathcal{A}} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$ whenever $\mathcal{A}$ and $\mathcal{B}$ are d-separated in the DAG with respect to $\mathcal{C}$. In undirected graphs, the global Markov property condition is much simpler: $\mathsf{x}_{\mathcal{A}} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$ whenever $\mathcal{C}$ separates $\mathcal{A}$ and $\mathcal{B}$ in $\mathcal{G}$.

- *Local/Pairwise Markov properties:* We also discussed weaker conditional independence assumptions for directed and undirected graphs, the local and pairwise Markov properties respectively. The local Markov property for a directed graph is simply the condition that $\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_{\mathrm{nd}(i) \setminus \pi_i} \mid \mathsf{x}_{\pi_i}$, i.e. that i is conditionally independent of its non-descendants (except its parents) given its parents. Meanwhile, the pairwise Markov property for undirected graphs is the condition that $\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_j \mid \mathsf{x}_{V \setminus \{i,j\}}$ for all $(i,j) \notin E$; i.e. that for vertices i and j not connected by an edge, x_i and x_j are conditionally independent given all other nodes in the graph. We can similarly derive a local Markov property for undirected models, or a pairwise Markov property for directed graphs - see the following table for those statements.

	Directed	Undirected
Factorization	$p_{\mathsf{x}}(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid x_{\pi_i})$	$p_{\mathsf{x}}(\mathbf{x}) = \frac{1}{Z} \prod_{\mathcal{C} \in \mathrm{cl}^*(\mathcal{G})} \psi_{\mathcal{C}}(x_{\mathcal{C}})$
Global Markov Property	$\mathsf{x}_{\mathcal{A}} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$ $\forall \mathcal{A}, \mathcal{B}$ d-separated w.r.t $\mathcal{C}$	$\mathsf{x}_{\mathcal{A}} \perp\!\!\!\perp \mathsf{x}_{\mathcal{B}} \mid \mathsf{x}_{\mathcal{C}}$ $\forall \mathcal{A}, \mathcal{B}$ separated by $\mathcal{C}$
Local Markov Property	$\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_{\mathrm{nd}(i) \setminus \pi_i} \mid \mathsf{x}_{\pi_i}$	$\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_{V \setminus (\mathcal{N}(i) \cup \{i\})} \mid \mathsf{x}_{\mathcal{N}(i)}$
Pairwise Markov Property	$\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_j \mid \mathsf{x}_{\mathrm{nd}(i) \setminus j}$ $\forall\ j \in \mathrm{nd}(i) \setminus \pi_i$	$\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_j \mid \mathsf{x}_{V \setminus \{i,j\}}$ $\forall\ (i,j) \notin E$

Here, $\mathcal{N}(i)$ refers to all the neighbors of node i in an undirected graph. We previously showed that for directed graphs factorization, the global Markov property, and the local Markov property are all equivalent. It is also straightforward to show that in undirected graphs, factorization implies the global Markov property, which implies the pairwise Markov property. For undirected graphs, however, we require certain stronger conditions (strict positivity) in order to go in the other direction. The Hammersley-Clifford Theorem, for instance, tells us that in the case of strictly positive density $p(x) > 0$ everywhere, the global Markov property indeed implies factorization. In fact, when $p(x) > 0$, these 4 conditions are all equivalent. This is true in the directed setting as well, where we require $p(x) > 0$ for the pairwise Markov property to imply the other 3 properties.

4.2 Evaluating Graphical Model Representations

We've seen that many graphical models can be used to represent the same distribution. Next, we will begin to develop how to quantify the concept of efficiency of representation of a distribution by a graphical model.

As our examples have illustrated, one natural way to think of the efficiency with which a given graphical model represents a given distribution in terms of the richness (i.e., size) of the family represented by the graph. More specifically, suppose we have two graphical models, $\mathcal{G}_1$ and $\mathcal{G}_2$. Associated with these graphs are two families of distributions, $\mathcal{P}(\mathcal{G}_1)$ and $\mathcal{P}(\mathcal{G}_2)$, respectively, corresponding to the factorization structure expressed by the graph. Then if a given distribution p of interest is representable by both graphs, i.e., $p \in \mathcal{P}(\mathcal{G}_1)$ and $p \in \mathcal{P}(\mathcal{G}_2)$, then we say that $\mathcal{G}_1$ is a more efficient representation for p than $\mathcal{G}_2$ if $\mathcal{P}(\mathcal{G}_1) \subset \mathcal{P}(\mathcal{G}_2)$ (strictly).

When the factorization structure corresponds to conditional independence structure, as in the case of directed and undirected graphs, then these relationships can be equivalently expressed in terms of conditional independencies. In particular, let $\mathcal{I}(\mathcal{G}_1)$ and $\mathcal{I}(\mathcal{G}_2)$ denote the collections of conditional independencies asserted by graphs $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. Then when p is, again, representable by both graphs, we say that $\mathcal{G}_1$ is the more efficient representation if $\mathcal{I}(\mathcal{G}_1) \supset \mathcal{I}(\mathcal{G}_2)$, i.e., if $\mathcal{G}_1$ corresponds to the more constrained family of distributions.

Let us now develop some of the formal language used to express these kinds of relationships. We begin with the following basic concept.

Definition 1 (I-map)**.** *A graph $\mathcal{G}$ is an* independence map *or* I-map *for a given distribution p if the set of conditional independencies $\mathcal{I}(\mathcal{G})$ expressed by $\mathcal{G}$ are among the set of conditional independencies $\mathcal{I}(p)$ satisfied by p, i.e., $\mathcal{I}(\mathcal{G}) \subset \mathcal{I}(p)$.*

From Definition 1 we conclude that that statement that $\mathcal{G}$ is an I-map for p is equivalent to the statement that p is one of the distributions in the family $\mathcal{P}(\mathcal{G})$ that can be represented by $\mathcal{G}$, i.e., $p \in \mathcal{P}(\mathcal{G})$.

Figure 1: Two graphs that are I-maps for a distribution with factorization $p_{\mathsf{x},\mathsf{y}}(x,y) = p_{\mathsf{x}}(x)\, p_{\mathsf{y}}(y)$.

Example 1. Consider a distribution $p_{\mathsf{x},\mathsf{y}}$ such that $\mathsf{x} \perp\!\!\!\perp \mathsf{y}$, i.e., one with the factorization structure $p_{\mathsf{x},\mathsf{y}}(x,y) = p_{\mathsf{x}}(x)\, p_{\mathsf{y}}(y)$. Then both graphs in Fig. 1 are I-maps for $p_{\mathsf{x},\mathsf{y}}$.

As the previous example illustrates, a complete (i.e., fully connected) graph is always an I-map for any given distribution because it implies no conditional independencies and thus corresponds to the family of all possible distributions.

As the preceding observation makes clear, the notion of an I-map pertains to whether some distribution of interest is representable by a given graph. As we discussed at the outset, we are more interested in is whether some distribution is efficiently represented by a given graph.

From this perspective, the following definition develops the notion of what it means for a graph to be a maximally efficient representation for some distribution of interest, i.e., for the graph to capture all structure in the distribution.

Definition 2 (P-map)**.** *The graph $\mathcal{G}$ is a* perfect map *or* P-map *for p if $\mathcal{I}(\mathcal{G}) = \mathcal{I}(p)$, i.e., if every conditional independence expressed by $\mathcal{G}$ is satisfied by p and vice versa.*

In essence, this means that if $\mathcal{G}$ is a P-map, the family of distributions represented by $\mathcal{G}$ includes no distributions with factors other than those present in p.

Example 2. Consider distributions p_1, p_2, and p_3 that factor as follows

$$
\begin{aligned}
p_1(x,y,z) &= p_{\mathsf{x}}(x)\, p_{\mathsf{y}}(y)\, p_{\mathsf{z}}(z) \\
p_2(x,y,z) &= p_{\mathsf{z}|\mathsf{x},\mathsf{y}}(z|x,y)\, p_{\mathsf{x}}(x)\, p_{\mathsf{y}}(y) \\
p_3(x,y,z) &= p_{\mathsf{z}|\mathsf{x},\mathsf{y}}(z|x,y)\, p_{\mathsf{x}|\mathsf{y}}(x|y)\, p_{\mathsf{y}}(y).
\end{aligned}
$$

Then the directed graph with V-structure in Fig. 2, which represents the independence relation $\mathsf{x} \perp\!\!\!\perp \mathsf{y}$ is an I-map for p_1, a P-map for p_2, and neither for p_3.

Example 3. Consider the graph in Fig. 3, which expresses the conditional independence relation $\mathsf{x} \perp\!\!\!\perp \mathsf{y} \mid \{\mathsf{z}, \mathsf{w}\}$. Then since

$$
p_{\mathsf{x},\mathsf{y},\mathsf{w},\mathsf{z}}(x,y,w,z) = \frac{1}{Z}\, f_1(x,w)\, f_2(w,y)\, f_3(y,z)\, f_4(z,x) \tag{1}
$$

corresponds to a factorization over the cliques in this graph, the Hammersley-Clifford theorem establishes that the graph of Fig. 3 is an I-map for the distribution (1). It is

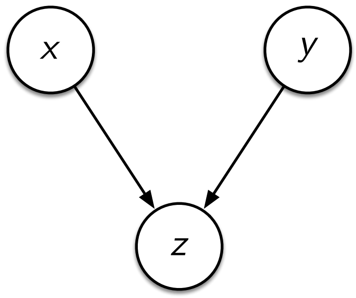


Figure 2: A directed graph expressing the independence $x \perp\!\!\!\perp y$.

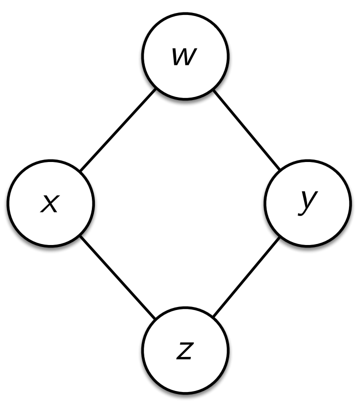


Figure 3: An undirected graph expressing the conditional independence $x \perp\!\!\!\perp y \mid \{z, w\}$.

a useful exercise to check that there is a distribution factorizing as in (1), for which the graph in Fig. 3 is a P-map, which means that the distribution does not satisfy any CI statements beyond those implied by the graph.

Unfortunately, for a given distribution p of interest, it is not always possible to find a graphical model of any particular type that is a P-map. In such cases, the best we could hope for is a graphical model for which the associated family of distributions is as small as possible while still containing p. In general, finding such a graph can be a difficult global optimization, so it is typical to resort to a local optimization, which the following definition reflects.

Definition 3 (Minimal I-Map)**.** *For a distribution p, a graph is* minimal I-map *if it is an I-map with the property that removing any single edge from the graph would cause it to no longer be an I-map.*

4.3 Constructing Good Graphical Model Representations

Now we shall discuss ways to construct an I-map for a distribution of interest.

4.3.1 Directed Minimal I-Maps

Suppose $\mathcal{V} = \{1, \ldots, N\}$ so our variables are $\mathsf{x}_1, \ldots, \mathsf{x}_N$. Then we can always write our target distribution via the chain rule as

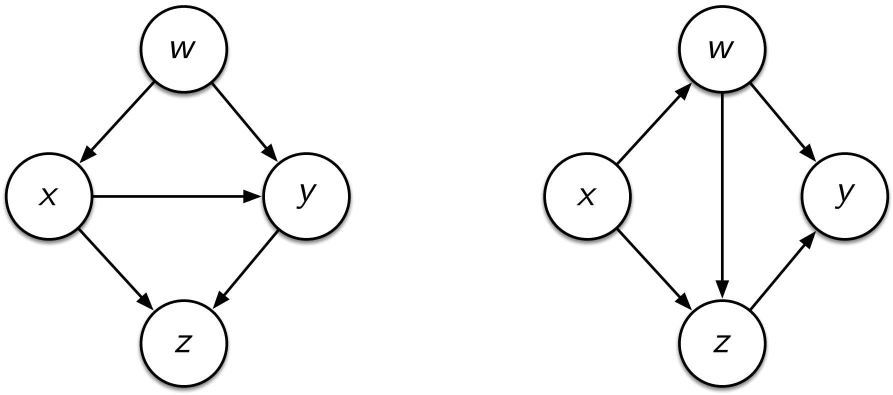
$$p_{\mathsf{x}_1,\ldots,\mathsf{x}_N}(x_1, \ldots, x_N) = p_{\mathsf{x}_1}(x_1)\, p_{\mathsf{x}_2|\mathsf{x}_1}(x_2|x_1)\, \ldots\, p_{\mathsf{x}_N|\mathsf{x}_{N-1},\ldots,\mathsf{x}_1}(x_N|x_{N-1}, \ldots, x_1), \tag{2}$$

where we have chosen the particular chain rule corresponding to the numerical ordering of the variables. For our graph, we choose the topological ordering of nodes to correspond to the numerical ordering of variables, so that the parents of a node always have earlier indices. We then draw the complete (fully-connected) directed acyclic graph corresponding to (2) with this ordering.

Next, for each conditional probability factor in (2), remove as many conditioning variables from the factor as conditional independencies of the distribution allow. In the graph this corresponds to generating the smallest number of parents for each node. Specifically, for node i, we choose the smallest set of parents $\pi_i \subset \{1, \ldots, i-1\}$ such that $\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_{\{1,\ldots,i-1\}\setminus\pi_i} \mid \mathsf{x}_{\pi_i}$ holds for the target distribution.

Clearly, this greedy algorithm produces a directed minimal I-map since by definition we cannot remove any additional edges without introducing a conditional independence that the target distribution does not satisfy.

It should be stressed that we could equally well choose any other ordering of the variables, in which case the corresponding chain rule and parent-pruning procedure would lead to potentially a different minimal I-map.



(a) The minimal I-map that expresses (4a).

(b) The minimal I-map that expresses (4b).

Figure 4: Two directed minimal I-maps for the distribution (3).

Example 4. Consider, once again, a distribution of the form (1):

$$p_{\mathsf{x},\mathsf{y},\mathsf{w},\mathsf{z}}(x, y, w, z) = \frac{1}{Z} f_1(x, w)\, f_2(w, y)\, f_3(y, z)\, f_4(z, x) \tag{3}$$

Recall that its corresponding undirected graph is depicted in Fig. 3. It follows that the conditional independencies corresponding to this family of distributions are

$$\mathsf{w} \perp\!\!\!\perp \mathsf{z} \mid \{\mathsf{x}, \mathsf{y}\} \tag{4a}$$

$$\mathsf{x} \perp\!\!\!\perp \mathsf{y} \mid \{\mathsf{z}, \mathsf{w}\} \tag{4b}$$

To obtain one directed minimal I-map for this distribution, choose the topological ordering $\mathsf{w}, \mathsf{x}, \mathsf{y}, \mathsf{z}$, consider the chain rule decomposition

$$p_{\mathsf{x},\mathsf{y},\mathsf{w},\mathsf{z}}(x, y, w, z) = \underbrace{p_{\mathsf{z}|\mathsf{y},\mathsf{x},\mathsf{w}}(z|y, x, w)}_{=p_{\mathsf{z}|\mathsf{y},\mathsf{x}}(z|y,x)}\, p_{\mathsf{y}|\mathsf{x},\mathsf{w}}(y|x, w)\, p_{\mathsf{x}|\mathsf{w}}(x|w)\, p_{\mathsf{w}}(w),$$

and note that, as indicated, (4a) allows the first term to be replaced with $p_{\mathsf{z}|\mathsf{y},\mathsf{x}}(z|y, x)$, but that no other terms can have parents removed without introducing a conditional independency that is not one of (4). As a result, the corresponding directed graph depicted in Fig. 4(a) is a minimal I-map for the distribution.

However, to obtain a different directed minimal I-map for this distribution, choose the topological ordering $\mathsf{x}, \mathsf{w}, \mathsf{z}, \mathsf{y}$, consider the chain rule decomposition

$$p_{\mathsf{x},\mathsf{y},\mathsf{w},\mathsf{z}}(x, y, w, z) = \underbrace{p_{\mathsf{y}|\mathsf{z},\mathsf{w},\mathsf{x}}(y|z, w, x)}_{=p_{\mathsf{y}|\mathsf{z},\mathsf{w}}(y|z,w)}\, p_{\mathsf{z}|\mathsf{w},\mathsf{x}}(z|w, x)\, p_{\mathsf{w}|\mathsf{x}}(w|x)\, p_{\mathsf{x}}(x),$$

and note that now (4b) allows the first term to be replaced with $p_{\mathsf{y}|\mathsf{z},\mathsf{w}}(y|z,w)$, but that no other terms can have parents removed without introducing a conditional independency that is not one of (4). Hence, in this case, the corresponding directed graph depicted in Fig. 4(b) is a minimal I-map for the distribution.

4.3.2 Undirected Minimal I-Maps

We shall describe a procedure to obtain a minimal undirected graphical model I-map for a given distribution p. We shall assume that the distribution is strictly positive over the domain of interest.

Start with a fully-connected undirected graph with nodes $\mathcal{V}$. Next, process each of the $N(N-1)/2$ edges in this graph in an arbitrary order, as follows. For the edge between nodes $i \in \mathcal{V}$ and $j \in \mathcal{V}$, remove it if the target distribution satisfies the conditional independence relation

$$\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_j \mid \mathsf{x}_{\mathcal{V}\setminus\{i,j\}}.$$

Otherwise, retain the edge. After testing all edges and removing those that meet the criterion, we claim that the resulting graph is an undirected minimal I-map for the target distribution.

To see that the resulting graph is indeed a minimal I-map, suppose to contrary. That is, suppose there exist a graphical model $\mathcal{G}$ that is a minimal I-map of distribution p but it either (a) has an edge between i, j even though $\mathsf{x}_i \perp\!\!\!\perp \mathsf{x}_j \mid \mathsf{x}_{\mathcal{V}\setminus\{i,j\}}$, or (b) does not have an edge between i, j but x_i is not independent of x_j given $\mathsf{x}_{\mathcal{V}\setminus\{i,j\}}$. We shall argue that neither (a) nor (b) is possible.

First we argue that (a) is not possible. By Hammersley-Clifford Theorem, the resulting distribution has factorization over maximal cliques $\mathrm{cl}^*(\mathcal{G})$ so that

$$p_{\mathsf{x}_\mathcal{V}}(x_\mathcal{V}) \propto \prod_{\mathcal{C}\in\mathrm{cl}^*(\mathcal{G})} \psi_\mathcal{C}(x_\mathcal{C}). \tag{5}$$

Further, there exists cliques $\mathcal{C} \in \mathrm{cl}^*(\mathcal{G})$ that contain both i and j. Due to minimality of this graph, it must be that these factors are non-trivial functions in all of its arguments. Therefore, if we fix assignment for $\mathsf{x}_{\mathcal{V}\setminus\{i,j\}} = x_{\mathcal{V}\setminus\{i,j\}}$, the right-hand side of (5) will be effectively a non-trivial function of x_i and x_j. That is, x_i and x_j will not be independent of each other given $\mathsf{x}_{\mathcal{V}\setminus\{i,j\}}$. This is a contradiction to our hypothesis behind removing edge between i and j. That is, all the minimal I-map undirected graphical models $\mathcal{G}$ will not have edge between i and j. This establishes the validity of our procedure for finding minimal I-map.

To see (b) is not possible, we can again appeal to Hammersley-Clifford Theorem: if i and j do not have edge, there is a factorization of distribution where i and j are not part of the same factor (or clique). Using this factorization, it can be easily argued that x_i and x_j are independent given $\mathsf{x}_{\mathcal{V}\setminus\{i,j\}}$. This complete the proof of the claim that the above procedure produces a minimal undirected graphical model I-map.

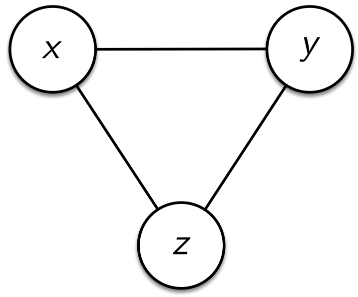


Figure 5: The unique undirected minimal I-map for the distribution (6).

Now, by construction, this is the unique minimal I-map for p since the edge removal rule does not depend on the order in which edges are removed. For some distributions, the undirected minimal I-map obtained from this procedure will turn out to be a P-map.

Example 5. Revisiting Example 2, consider a distribution of the form

$$p(x, y, z) = p_{\mathsf{z}|\mathsf{x},\mathsf{y}}(z|x, y)\, p_{\mathsf{x}}(x)\, p_{\mathsf{y}}(y), \tag{6}$$

which expresses the independence relation $\mathsf{x} \perp\!\!\!\perp \mathsf{y}$. As we've discussed, the directed graph Fig. 2 is a P-map for this distribution. However, to obtain an undirected minimal I-map for it, we start with the fully connected graph of Fig. 5, but note that that attempting to eliminate any edge leads to a conditional independency not satisfied by our distribution. For instance, removing the edge between x and y implies that $\mathsf{x} \perp\!\!\!\perp \mathsf{y} \mid \mathsf{z}$, which is not satisfied by (6). By analogous reasoning, we cannot remove either of the other two edges as well. Hence, we conclude that the graph of Fig. 5 is the unique undirected I-map for (6).

4.4 Converting Graphical Model Representations

In this section, we more generally discuss good ways of converting graphs of one type into graphs of another. Our goal in such mappings is to ensure that: 1) every probability distribution represented by the source graph is also represented by the destination graph; and 2) the destination graph represents as few distributions as possible.

When the factorization structure is associated with conditional independencies, as in the case of directed and undirected graphical models, we can express our goal as seeking a destination graph that is a minimal I-map for every distribution for which source graph is a P-map. Equivalently, we seek a destination graph that expresses the largest possible subset of the conditional independencies expressed by the source graph.

4.4.1 From Directed to Undirected Graphical Models: Moralization

The procedure for generating an undirected graphical model from a directed one is based on the method of generating an undirected minimal I-map as discussed in Section 4.3.2. In particular, in Section 4.3.2 we begin with a *distribution*, from which we extract conditional independence relations, then use these to generate an undirected minimal I-map. By contrast, here, analogous to the process above, we begin with a *directed graph*, describing the collection of conditional independence relations implied by graph separation, then from this point use these relations in the same manner. We likewise note that Example 5 can be interpreted as carrying out such a conversion, since the distribution (6) is the factorization corresponding to the directed graph of Fig. 2.

Note that the resulting process of converting directed to undirected graphs can be equivalently be described as one of *moralization*, defined as follows.

Definition 4 (Moralization)**.** *An undirected graph is the* moralization *of a directed graph if the undirected graph has an (undirected) edge between every pair of nodes that the directed graph does, and, in addition, (undirected) edges between every pair nodes in the set of parents π_i, for each node $i \in \mathcal{V}$.*[1]

From the perspective of Definition 4, Example 5 is the simplest example of moralization that leads to an extra edge being added to the undirected graph. In summary, we have the following claim about moralized graphs.

Theorem 1. *An undirected graph that is the moralization of a given directed graph is a minimal I-map for the family of distributions represented by the directed graph.*

Proof. We briefly sketch the proof here.

Consider the directed graph and the family of distributions that are represented by it. Consider a node i and its parents π_i. As per the factorization of the directed graphical model, $p_{x_i|x_{\pi_i}}$ is non-trivial function of all variables associated with i and π_i. We claim that, in any undirected graph that is an I-map of this family of distribution, edges between i and each of its parents as well as edges between any two parents of node i must exist. This will establish the statement of Theorem.

To see this claim, note that the factorization $p_{x_i|x_{\pi_i}}$ is a non-trivial function of values associated with i and any of its parent. Therefore, conditioning on values assigned to all nodes but i and one of its parent, say j, the above factor function becomes non-trivial function of i and j. That is, x_i and x_j are not independent given $x_{\mathcal{V}\setminus\{i,j\}}$. Therefore, per procedure described earlier for finding undirected I-map, edge between i and j must be present. Similarly, we can argue that if $|\pi_i| > 1$, then for any $j, k \in \pi_i$, edge between them must exists in the corresponding minimal I-map undirected graph. An alternative way to see this is, since j and k are not d-separated

[1] Moralization is so named because it "marries" the parents by connecting them together.

when i is in the conditiong set, $x_j \not\perp\!\!\!\perp x_k | x_{\mathcal{V}\setminus\{j,k\}}$. The above argument proves that all the edges in a moralized graph are necessarily present in a minimal undirected I-map.

We now show that these edges are also sufficient. That is, we want to prove that the moralized graph is an I-map for the original DAG. For this it is sufficient to show that for all nodes a, b and any set of nodes $\mathcal{C}$, if a and b have an unblocked path given $\mathcal{C}$ in the DAG, then a and b also have an unblocked path given $\mathcal{C}$ in the moralized graph. Note that for the DAG, the notion of blocking is given by d-separation, whereas for the moralized graph it is given by Markov separation. Hence assume that such a path exists in the DAG. If the path has no head-to-head nodes which are in $\mathcal{C}$, then the same path is unblocking for the moralized graph. If the path has head-to-head nodes which are in $\mathcal{C}$, we can bypass them using edges added during moralization, giving us a new unblocked path. This completes the proof of the claim and the theorem.

□

4.4.2 From Undirected to Directed Graphical Models: Chordalization

The procedure for generating a directed graphical model from an undirected one is based on the method of generating a directed minimal I-map as discussed in Section 4.3.1. In particular, in Section 4.3.1 we begin with a *distribution*, from which we extract its conditional independence relations, then use these to generate a directed minimal I-map. By contrast, here we begin with an *undirected graph*, describing the collection of conditional independence relations. But once we have these relations we proceed in exactly the same manner. Indeed, Example 4 can be intrepreted as carrying out precisely such a conversion from an undirected to a directed graph, since the distribution (3) is the factorization corresponding to the undirected graph of Fig. 3.

The process above for converting an undirected graph to a directed one can equivalently be interpreted as one of *chordalization*, defined as follows.

Definition 5 (Chordalization)**.** *For any loop (cycle) in a given graph, a "chord" is an edge connecting two non-consecutive nodes in the loop. Furthermore, a graph is called* chordal *if every loop in the graph of length at least four has a chord.*

In the language of this definition, to convert an undirected graph to a directed graph involves: 1) placing an edge in the directed graph whenever the corresponding undirected graph has an edge; and 2) adding edges so that the resulting graph is chordal. Recall that this is precisely what happened in Example 4.

It is important to emphasize that chordalization, while well defined, can be a bit subtle to visualize and in particular does not always correspond to what might be envisioned as simple "triangularization." As an illustration of this, Fig. 6 depicts an (undirected) chordalization of a simple two-dimensional Ising model.

The details of the general procedure, including how to choose the orientation of the edges, and the location of the chords, will be developed later in the notes, since it arises naturally when we begin to analyze inference on graphs.

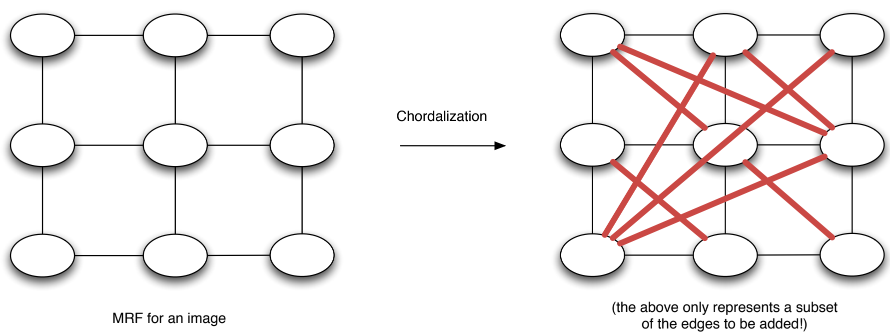


Figure 6: (Undirected) chordalization of a simple two-dimensional Ising model.

4.5 Distributions with both Directed and Undirected P-Maps

In this section, we begin to examine important classes of distributions that have *both* directed *and* undirected P-maps. Such distributions (and the associated graphical models) are particularly special, and allow for highly efficient inference, as we will see.

While the complete picture will take additional lectures to develop, initial insights come from the following claims, which arise out of the preceding development.

Claim 1. *An undirected graph $\mathcal{G}$ has a directed P-map if and only if $\mathcal{G}$ is chordal.*

Claim 2. *A directed graph $\mathcal{G}$ has an undirected P-map if and only if moralization of $\mathcal{G}$ does not add any edges.*

4.5.1 Mutually Independent Variables

The simplest possible class of distributions with both directed and undirected P-maps are those corresponding to mutually independent variables. Indeed, these are represented by directed and undirected graphs with no edges whatsoever.

4.5.2 Markov Chains

A still simple but surprising rich class of distributions with both directed and undirected P-maps are those corresponding to Markov chains, an example of which is depicted in Fig. 7. For Markov chains, the directed graphical model and undirected graphical representations have exactly the same "structure" (ignoring directionality of edges).

It is worth emphasizing that Markov chains also have other associated graphical representations with which you may already be familiar: state transition diagrams and trellis diagrams. While these representations, depicted in Fig. 8, continue to be useful, it is important not confuse these with our new graphical model representations

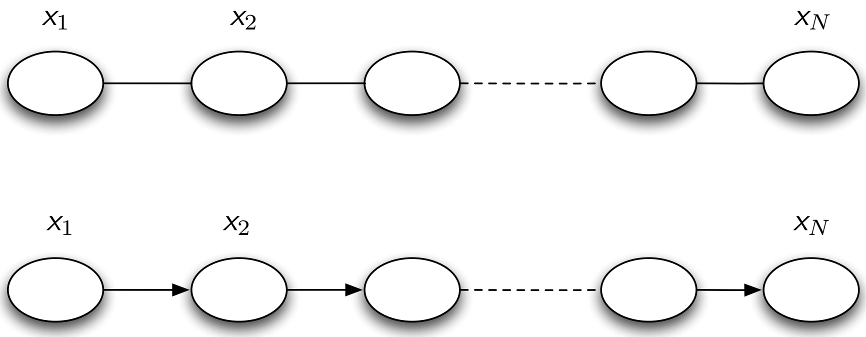


Figure 7: Markov chain in undirected and directed representations.

of Markov chains: the nodes and edges mean very different things in these different representations!

A closely related class are the "hidden" Markov models, an example of which is depicted in Fig. 9. The random variables of interest are $x_1, \ldots, x_N$ and the $y_1, \ldots, y_N$ represent "noisy" observations. The undirected and directed graphical representation for both of these distributions is likewise the same.

4.5.3 Trees

A class of distributions with both directed and undirected P-maps that is slightly richer still are tree models. To make its directed graph counterpart, start by designating *any* node as the "root" node and start propagating directed edges away from the root in all directions in the tree out to the "leaf" nodes to obtain the associated directed graph. An example is depicted in Fig. 10.

Likewise, a closely related class are forests, which as we discussed are collections of disconnected trees.

4.5.4 Chordal Graphs

Ultimately, the largest class of distributions with both directed and undirected P-maps are those corresponding to chordal graphs. Note that all the preceding classes can be viewed as special cases of chordal graphs, because they contain no loops.

As we will see, a consequence of this is that chordal graphs are those for which efficient *exact* inference is possible, in a particular sense that we will make precise later in the notes.

In turn, this means that there are many graphs whose structure it is less clear how to fully exploit. As an illustration, recall that to convert the simple 2-D Ising model to a chordal graph requires adding a suprising number of edges, which corresponds to giving up on trying to exploit important forms of structure in this model. As a result,

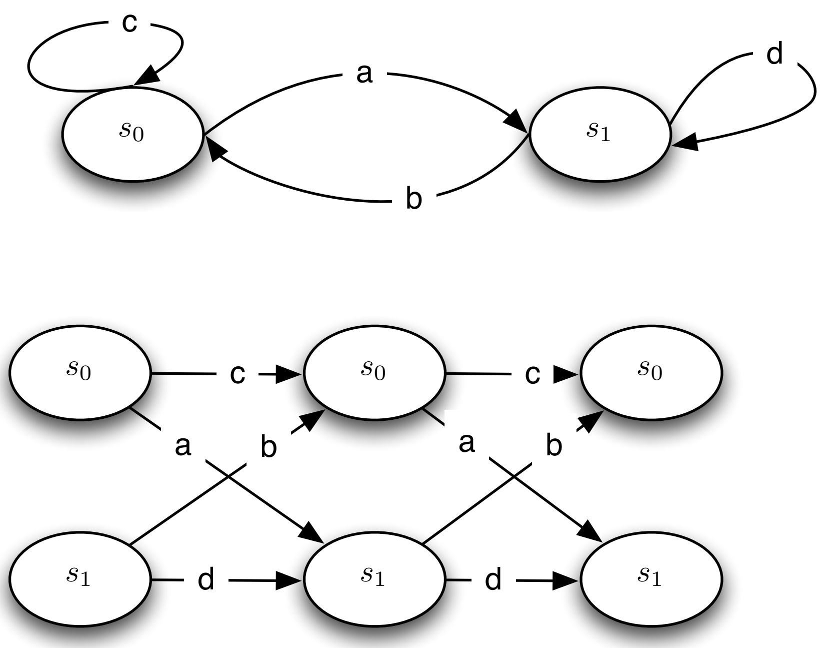


Figure 8: The state transition diagram and trellis diagram for a (homogenous) Markov chain.

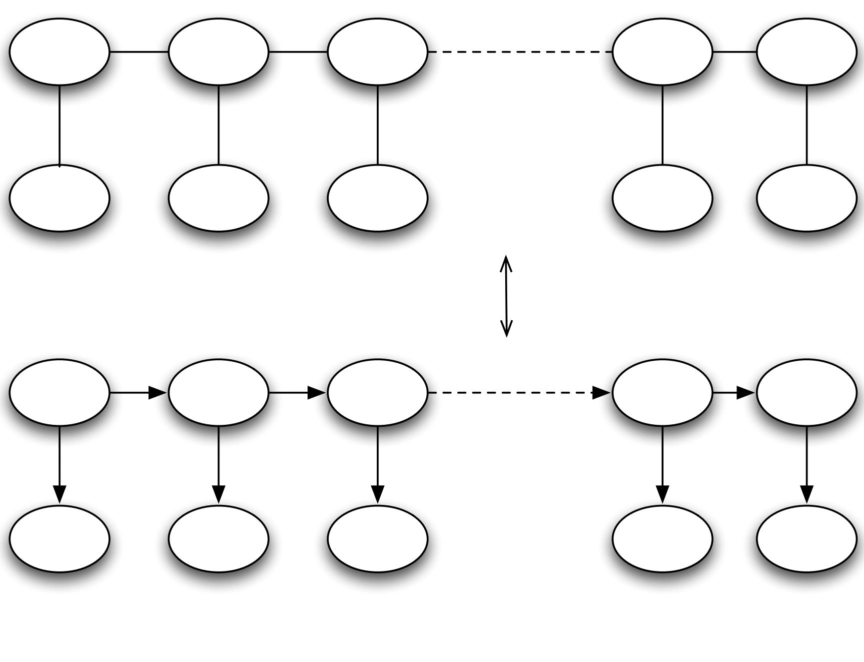


Figure 9: Hidden Markov model in undirected and directed representations.

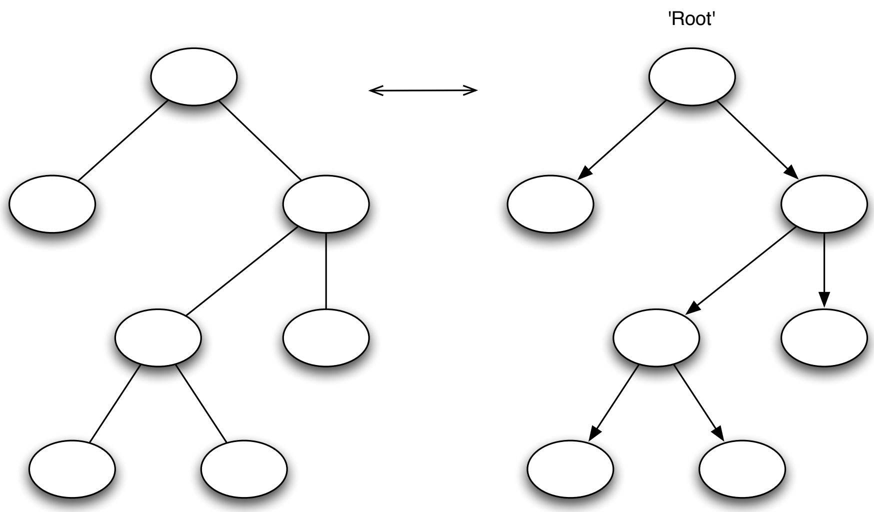


Figure 10: Converting from undirected to a directed tree.

in practice it is common to seek efficient *approximate* inference algorithms that more fully exploit the available structure, which we will also explore.

4.5.5 Summary

As a summary of some of the key concepts developed in these notes, Fig. 11 provides a useful visual categorization of distributions.

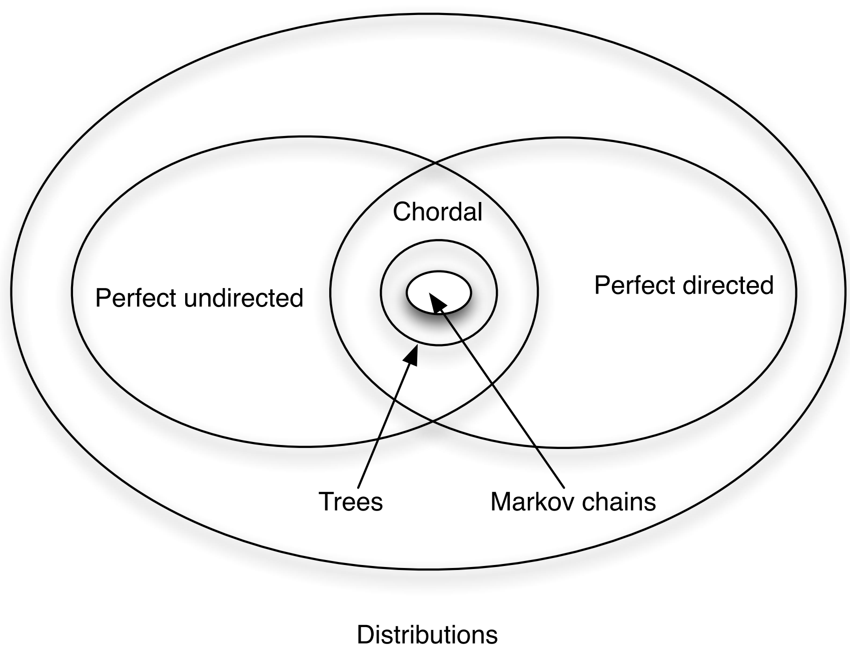


Figure 11: A Venn diagram categorization of distributions.