CS281A/Stat241A: Statistical Learning Theory

Directed and Undirected Graphical Models (9/7/04)

Lecturer: Michael I. Jordan *Scribes: Bonnie Kirkpatrick and Pei-Chen Liu*

1 Review of Three Canonical Graphs

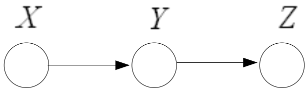

Figure 1: **Markov Chain**

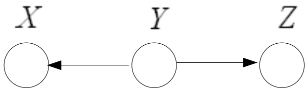

Figure 2: **Hidden Cause**

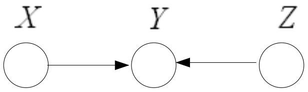

Figure 3: **Explaining Away**

As we learned last Thursday, there are three canonical graphs that can be used to describe the conditional independence of all graphs. They are the Markov chain (Fig. 1), the hidden cause graph (Fig. 2), and the explaining away graph (Fig. 3).

The Markov chain of these has a joint probability distribution that factorizes in the following way: $p(x, y, z) = p(x)p(y \mid x)p(z \mid y)$. It is associated with the conditional independence statement $X \perp\!\!\!\perp Z \mid Y$.

The hidden cause graph has a joint probability that factorizes in the following way: $p(x, y, z) = p(y)p(x \mid y)p(z \mid y)$. It is associated with the conditional independence statement $X \perp\!\!\!\perp Z \mid Y$.

Finally, the explaining away case has a joint probability that factorizes in the following way: $p(x, y, z) = p(x)p(z)p(y \mid x, y)$. It is associated with the conditional independence statement $X \perp\!\!\!\perp Z$.

2 Bayes Ball Algorithm and d-Separation

Our intent now is to answer conditional independence queries such as $X_A \perp\!\!\!\perp X_B \mid X_C$ by using d-separation. First, shade the nodes, C, that are conditioned on. Second, start the ball in the nodes in set A and bounce it around the graph according to the conditional independence rules we will discuss. Finally, evaluate the results of the algorithm:

If the ball can reach a node in set B, then the nodes in A and B are *not necessarily* conditionally independent.

If the ball cannot reach B, then the nodes in A and B must be conditionally independent.

The rules that govern the bouncing of the ball can be described using the three canonical graphs that we saw earlier. Each graph has two sets of rules. When moving from X to Z (or Z to X) on the Markov chain, the ball will pass through Y to Z (or X) if we do *not* condition on Y. On the other hand, if we condition on Y, then X and Z are rendered independent, and the ball bounces off of Y when it tries to travel to Z (Fig. 4). When the ball moves from X to Z (or Z to X) on the hidden cause graph, the rules are similar to those on the Markov chain (Fig. 5). If we condition on Y, then the ball bounces off of Y, otherwise, it passes through Y to Z (or X). The last canonical graph, known as explaining away, represents different conditional independence statements than the other two graphs. The ball is blocked when it attempts to move from X to Z (or Z to X) when we do not condition on Y. But conditioning on variable Y opens a path for the ball to pass from X to Z (Fig. 6).

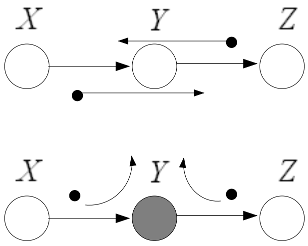

Figure 4: **Markov Chain**

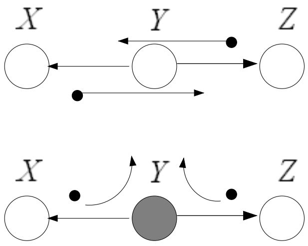

Figure 5: **Hidden Cause**

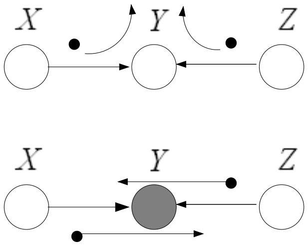

Figure 6: **Explaining Away**

This last graph, explaining away, has some interesting properties when additional nodes are appended to the Y node (Fig 7). In the case where nothing is known about the new node W or Y, then the ball cannot get from X to Z. But, if the graph is conditioned on node W, then a path opens up for the ball to move from X to Z. This means that the bouncing ball may have to search an arbitrary number of nodes (attached in a chain to Y) before deciding whether X and Z are independent.

For intuition into this case, an example can be given by assigning *burglary* to X, *earthquake* to Z, *alarm* to Y, and *friend's report* to W. In general, the chances of a burglary or an earthquake are independent, but if an alarm goes off in Bob's building, then his suspicions of the cause (either burglary or earthquake) are highly dependent on each other. This is the case where we condition on Y. Consider a scenario where Bob does not hear the alarm, but a friend tells him that the alarm went off. In this case, we condition on W, and again, the events of a burglary or an earthquake are no longer independent.

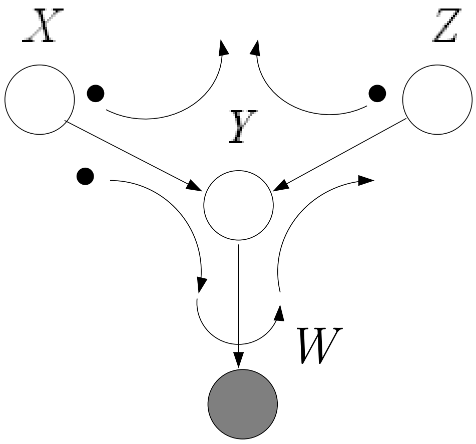

Figure 7: **Augmented Explaining Away**

For more information on the Bayes ball algorithm, see Ross Shachter's paper showing that there is an implementation which is linear in the number of edges in the graph.

3 Graphical Models and Probability Theory

Now, we will turn our attention back to the sets D_1 and D_2:

$D_1 = \{p(x) = \prod_{i=1}^{n} p(x_i \mid x_{\pi_i})\}$

and

$D_2 = \{p(x_v) : p(x_v) satisfies\ all\ conditional\ independence\ statements\ associated\ with\ the\ graph\}$

Theorem: $D_1 = D_2$
Proof: Given later in the book.

At the moment we can conceptualize the proof by using the Bayes ball algorithm to investigate all possible sets A, B, C to answer the query: $X_A \perp\!\!\!\perp X_B \mid X_C$?

The set D_2 containing all the conditional independence (CI) statements from a given graph is useful for describing the family of probability distributions that is associated with the graph. We can use these CI statements to carve up the space of all probability distributions. For example, consider the two graphs in Figures 8, 9. It is clear that the distribution $p(x_v) = p(x_1)p(x_2)p(x_3)p(x_4)p(x_5)$ is in the family of distributions satisfying the graph in Figure 9. Is this same distribution in the family of distributions that satisfies the graph in Figure 8? This is true because the factorization does not imply any dependencies at all. Recall that it is the set of conditional independence statements that define the family of distributions corresponding to the graph. Since the above distribution satisfies all the CI statements for figure 8, then it is in the family of distributions for the graph.

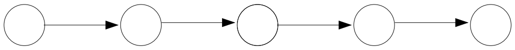

Figure 8: **Markov Chain**

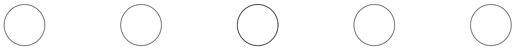

Figure 9: **Edge-less Graph**

Figure 10 shows an example of a fully-connected directed graph. Even when conditioned on X_2, X_1 and X_4 are not conditionally independent because there is an edge connecting X_1 and X_4 directly. From this example we can see that when there are more edges, there are fewer conditional independencies associated with a graph.

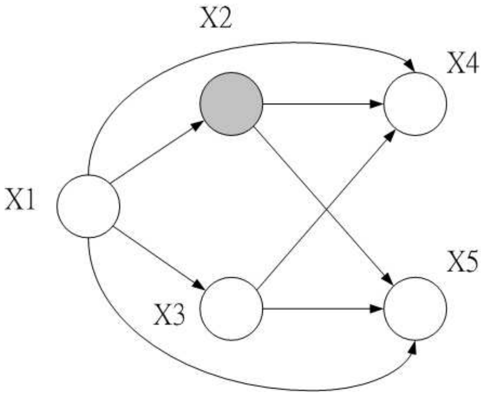

Figure 10: **Fully-connected Directed Graph**

Figure 11 shows another example of a simple Markov chain with four nodes. We can get the following basic conditional independencies from this graph:

$\{X_1, X_2\} \perp\!\!\!\perp X_4 \mid X_3$
$X_1 \perp\!\!\!\perp \{X_3, X_4\} \mid X_2$
This example illustrates that the missing variables in the local conditional probability functions correspond to missing edges in the underlying graph.

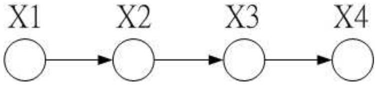

Figure 11: **Markov Chain**

4 Undirected Graphical Models

There are two major classes of graphical models: those based on directed graphs and those based on undirected graphs. The algorithm to answer the conditional independence queries for undirected graphs is more straightforward than that for the directed graphs. In fact, for undirected graphs, the problem of determination of conditional independence is a "reachability" problem in classical graph theory. That is, starting from X_A if you could not reach X_B through any path in the graph, then X_A and X_B are conditionally independent. This "reachability" problem can be solved by standard search algorithms.

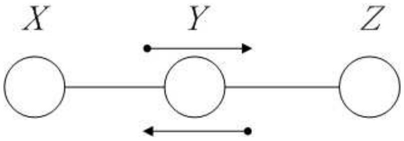

Figure 12: **Undirected Graph**

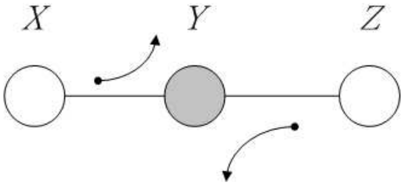

Figure 13: **Undirected Graph with a Blocking Node**

This means that, for undirected graphs, the Bayes ball algorithm only uses two rules to answer the conditional independence queries. When moving from X to Z (or Z to X), the ball will pass through Y to Z (or X) (Fig. 12). In this case, X and Z are not conditionally independent. On the other hand, with node Y shaded (i.e. conditioning on Y), when moving from X to Z (or Z to X), the ball bounces off of Y when it tries to travel to Z (or X) (Fig. 13). In this case, X and Z are independent when conditioned on Y. In another example (Fig. 14), with X_2 and X_4 shaded, i.e. conditioned on X_2 and X_4, X_1 cannot get to X_5, and therefore, $X_1 \perp\!\!\!\perp X_5 \mid \{X_2, X_4\}$. In the notion of graph theory, a shaded node represents that the network flow is blocked at that node. Hence, the Bayes ball algorithm for undirected graphs is essentially the reachability algorithm under the notion of naive graph separation.

A *clique* is a subset of nodes in the graph that are fully-connected. For example, in Figure 14, $\{X_1, X_3\}$ is a clique, $\{X_1\}$ is a clique, $\{X_2, X_5\}$ is also a clique, but $\{X_1, X_2, X_3\}$ is not a clique, whereas $\{X_2, X_4, X_5\}$ is a *maximal clique.* The *maximal cliques* of a graph are the cliques that cannot have more nodes added and remain a valid clique. Now let $\mathcal{C}$ = set of all cliques in $\mathcal{G}$, and $\mathcal{C} = \{C_1, C_2, \ldots\}$. To each clique C_i we associate a potential function $\psi_{C_i}(X_{C_i})$, which is nonnegative and real-valued, but otherwise arbitrary.

For a given undirected graph, we define a family of probability distributions, $\mathcal{U}_1$, by ranging over all possible choices of positive potential functions on the maximal cliques of the graph. $\mathcal{U}_1$ is therefore a parametric description of joint probability distribution. We define a second family of probability distributions,

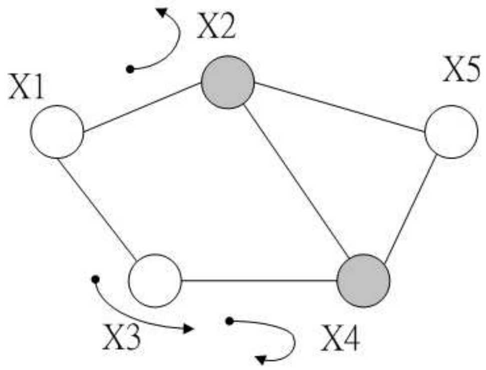

Figure 14: **Sample Undirected Graph**

$\mathcal{U}_2$, via the conditional independence assertions associated with the graph. The sets $\mathcal{U}_1$ and $\mathcal{U}_2$ are:

$$\mathcal{U}_1 = p(x_v) = \frac{1}{Z(\psi)} \prod_{C_i \in \mathcal{C}} \psi_{C_i}(X_{C_i}), \text{ where the normalization factor } Z(\psi) = \sum_{X_v} \prod_{C_i \in \mathcal{C}} \psi_{C_i}(X_{C_i})$$

and

$$\mathcal{U}_2 = \{p(x_v) : p(x_v) \; satisfies\ all\ conditional\ independence\ statements\ associated\ with\ the\ graph\ \mathcal{G}\}$$

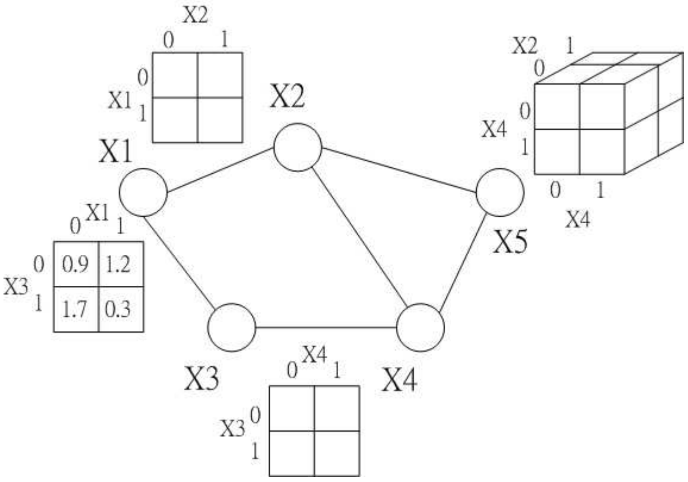

Figure 15: **An example of undirected graphical model**

Figure 15 shows that the maximal cliques in this graph are $\{X_1, X_2\}$, $\{X_1, X_3\}$, $\{X_3, X_4\}$, and $\{X_2, X_4, X_5\}$. Letting all nodes be binary, we represent a joint distribution on the graph via the potential tables that are displayed. Also note that one node is itself a clique.

Theorem: (Hammersley-Clifford) $\mathcal{U}_1 = \mathcal{U}_2$
Proof: Given later in the book.

Here we look at examples of conditional indepence relationships unique to directed or undirected graphs. In Fig 16, we have seen that $X \perp\!\!\!\perp Z$; however, $\neg X \perp\!\!\!\perp Z \mid Y$. Although X and Z are marginally independent, we cannot assert in general that $X \perp\!\!\!\perp Z \mid Y$ because there exists for this graph structure some acceptable probability distribution that violates it. In the undirected graph in fig 17, because the ball can go from X through Y to Z, we can state $\neg X \perp\!\!\!\perp Z$. If, however, we condition on Y, just like in a Markov chain, the ball cannot get through and we state that $X \perp\!\!\!\perp Z \mid Y$. In the fully-connected case as in Fig 18, there are no conditional independence statements associated with this graph. There is no way to capture the same conditional independence statements in the explaining-away directed graph using an undirected graph with three nodes.

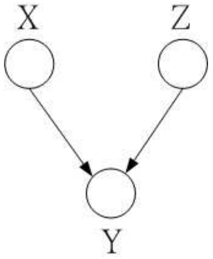

Figure 16: **Directed Graph**

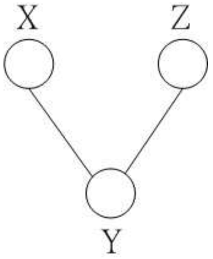

Figure 17: **Undirected Graph**

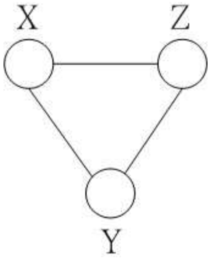

Figure 18: **Fully Connected Undirected Graph**

Consider the four-node graph in Figure 4, which could represent a disease propagation model. Assume X_1 and X_4 are male while X_2 and X_3 are female. There might be diseases that can be transmitted only via the opposite sex. We can get the following relationships from the graph: $X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$, and $X_2 \perp\!\!\!\perp X_3 \mid \{X_1, X_4\}$. We leave it as an exercise to the reader to determine that there are no directed graph with four nodes that can capture the conditional independence statements in this diamond graph.

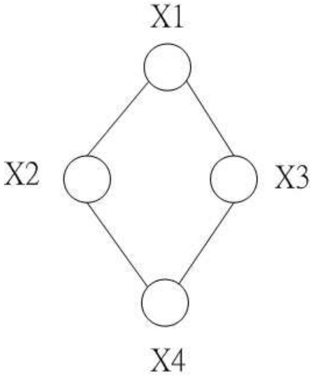

Figure 19: **Undirected Graph**

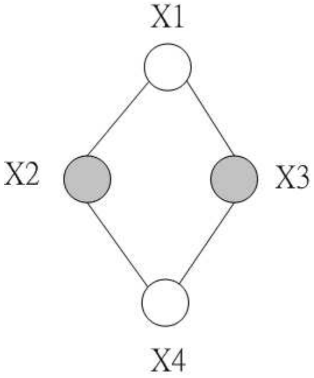

Figure 20: **Directed Graph with Shaded Nodes**

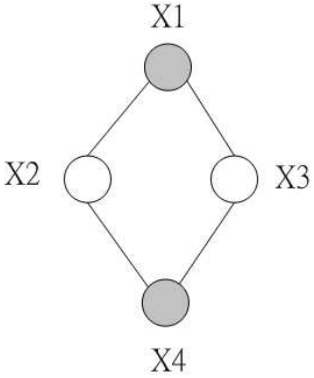

Figure 21: **Directed Graph with Shaded Nodes**