208891 Probabilistic Graphical Models **Spring 2019**

Lecture 15: February 25

Lecturer: Donlapark Pornnopparath

15.1 Markov Random Fields

There are situations in which there are some relations, but no dependency between variables. For example, consider a group of four students voting for a class representative, as shown in Figure 15.1. Even though there is (supposedly) no hierarchy of voting preference, closed friends will most likely vote for the same person. Such influence between friends is indicated by the edge between them.

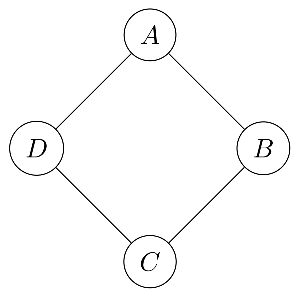

Figure 15.1: A Markov random field

To turn this graph into a probabilistic model, we define a *factor* $\phi(X,Y)$ on each edge XY which returns a high value if $X = Y$. For example, we could simply use $\phi(X,Y) = \mathbb{1}(X = Y)$. We can define a *score* that measures the likelihood of each event by taking a product of all factors.

$$\tilde{p}(A,B,C,D) = \phi(A,B)\phi(B,C)\phi(C,D)\phi(D,A). \tag{15.1}$$

To make the score of all events sum to one, we normalize $\tilde{p}$

$$p(A,B,C,D) = \frac{1}{Z}\tilde{p}(A,B,C,D),$$

where $Z = \sum_{A,B,C,D} \tilde{p}(A,B,C,D)$. This is a simple example of a Markov random fields (MRF).

Definition 15.1. A Markov random field is a probability distribution over variables $X_1, X_2, \ldots, X_n$ defined over an undirected graph G. Its pdf is given by

$$p(x_1,\ldots,x_n) = \frac{1}{Z}\prod_{c\in C}\phi_c(x_c), \tag{15.2}$$

where C is the set of all possible cliques in G and x_c denotes all variables in c. The normalizing factor Z is defined by

$$Z = \sum_{x_1,\ldots,x_n}\prod_{c\in C}\phi_c(x_c).$$

Note that Bayesian networks are MRFs with conditional probabilities as factors. For example, in (15.1), taking $\phi(D, A) = p(D|A)$, $\phi(A, B) = \phi(B|A)$, $\phi(B, C) = \phi(C|B)$ and $\phi(C, D) = p(C|D)$, we obtain the Bayesian network as in Figure 15.2.

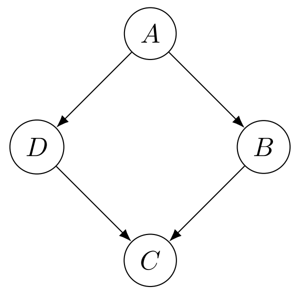

Figure 15.2: A Bayesian network version of the MRF in Figure 15.1

However, when we convert a Bayesian network into an MRF, the joint pdf factorization may lead to additional edges in the MRF. For example, the factorization of the Bayesian network in Figure 15.1 is

$$p(A, B, C, D) = p(D|A)p(B|A)p(C|B, D).$$

By defining $\phi_{DA}(D, A) = p(D|A)$, $\phi_{AB}(A, B) = p(B|A)$ and $\phi_{BCD}(B, C, D) = p(C|B, D)$, we instead obtain the MRF in Figure 15.3.

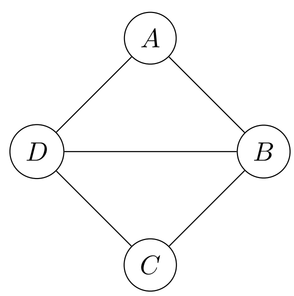

Figure 15.3: An MRF version of the Bayesian network in Figure 15.2

This is because an MRF can also indicate *indirect* dependencies, which are not shown in Beyasian networks, as we can see here; in Figure 15.3, there is an edge between B and D which are not d-separated.

It is easy to determine if two variables X and Y in an MRF are independent; as long as there is a path between X and Y, then they are dependent, otherwise they are independent. This can be shown directly using (15.2); as the sets of cliques that contain X and Y are disjointed, the marginalization in $p(X, Y)$ in (15.2) can be separated as a product of marginals over X and Y.

If we have observed another set of variables $\boldsymbol{Z}$, then $X \perp\!\!\!\perp Y|\boldsymbol{Z}$ if $\boldsymbol{Z}$ "blocks" all the paths from X to Y. For example,every variable in S_1 are conditionally independent from every variable in S_2 given X and Y.

In particular, if V is the set of all nodes and $\mathcal{N}(X)$ is the neighborhood of X then $X \perp\!\!\!\perp V - \{X\} - \mathcal{N}(X)|\mathcal{N}(X)$. We also call $\mathcal{N}(X)$ the *Markov blanket*, which is the minimal set of nodes such that X is independent from any other nodes in the graph.

One advantage of MRFs is that they can express some conditional probability conditions that cannot be satisfied by a Bayesian network. For example, consider the following conditions:

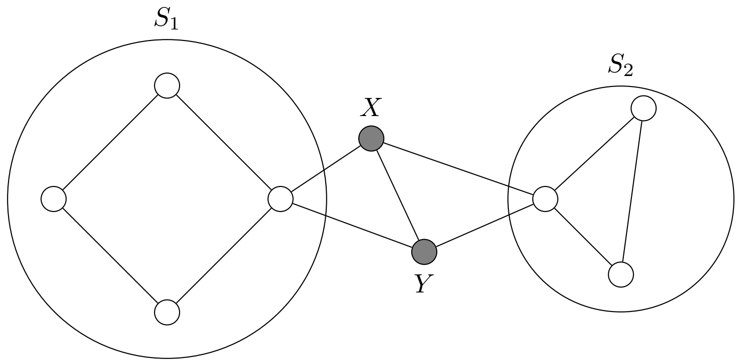

Figure 15.4

$$A \perp\!\!\!\perp C | B, D$$
$$B \perp\!\!\!\perp D | A, C$$

We can check that the MRF in Figure 15.1 satisfies these conditions. However, this is not the case for Bayesian networks; either there would be a cycle or a pair of arrows pointing to a node, creating a v-structure, and one of the independence conditions would not hold.

Nonetheless MRFs have some disadvantages. For large networks, it is harder to make inference and learn the structure in an MRF, mainly because of the difficulty of computing Z. In fact, computing Z is generally NP-hard. It is also harder to generate data from an MRF compared to Bayesian networks.