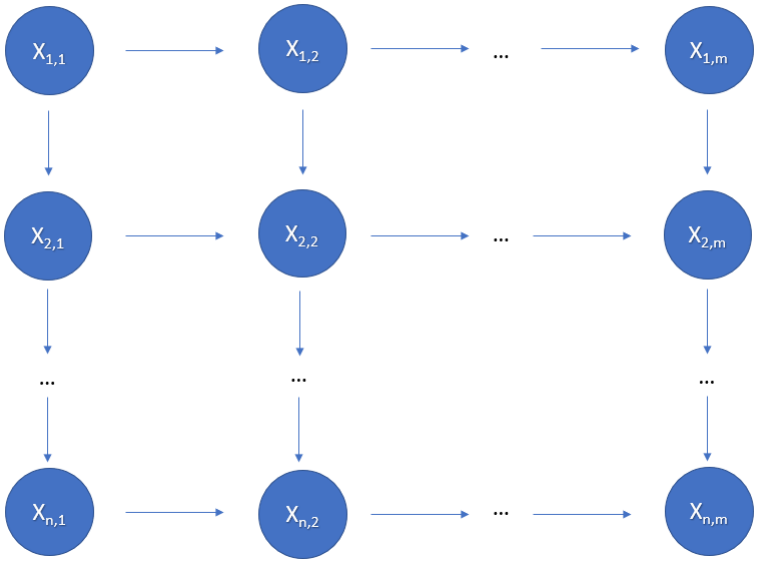
CSC412 Notes Week 4

Jerry Zheng

April 2021

1 Problems with Directed Graphical Models

For some problems, directionality for the edges in our DAGMs really hinders us.
for example, when we process an image we know that a pixel depends on its neighbours.
Lets Say pixel 2 depends on pixel 1 and pixel 3 depends on pixel 2.
We can extrapolate this into a Markov mesh.

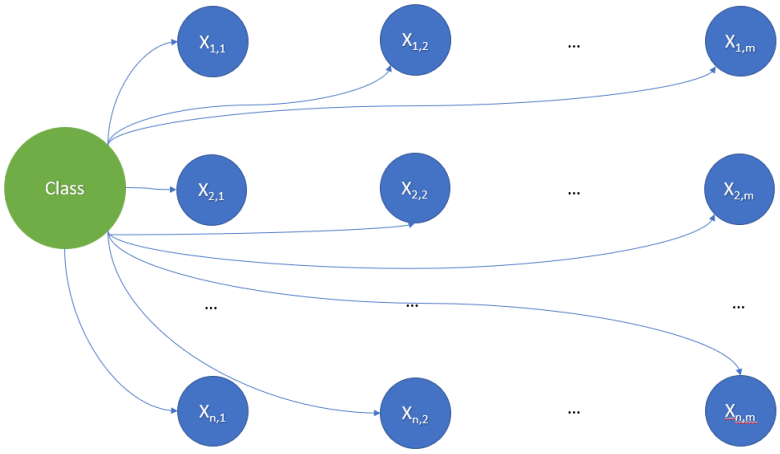


Of course, this model isnt very good because dependencies are directional and only go down and to the right.
Also, if we observe some pixels, then pixels nearby can be arbitrarily conditionally independent!

Alternatively we can have a Naive Bayes model by introducing a hidden class variable.

$$p(X) = \sum_z p(X, z)$$

$$p(X) = \sum_z \prod_{x_i \in X} p(x_i|z)$$



However there are issues with this too.
The top left and bot right pixels are dependent on each other which might not be desired.
Also, if we know what the class is then all the pixels are conditionally independent.

An alternative to DAGMs, is undirected graphical models (UGMs).

2 Undirected Graphical Models

In UGMs, we have edges that captures relation between variables rather than defining them as parent and child.

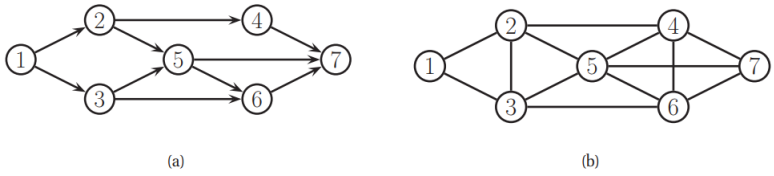
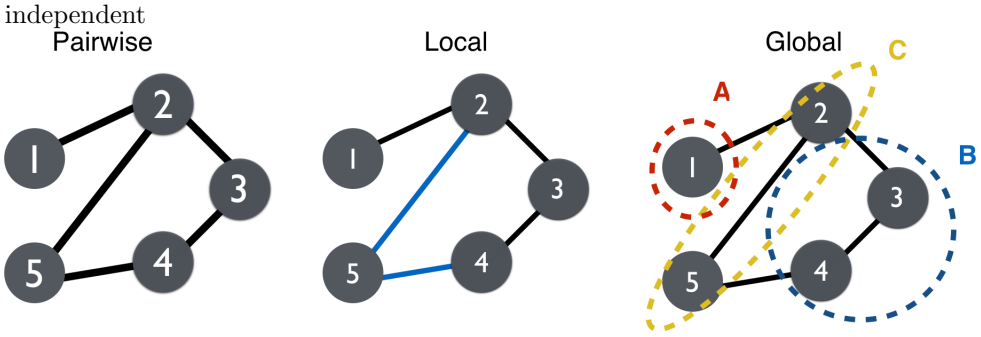


Figure 19.2 (a) A DGM. (b) Its moralized version, represented as a UGM.

2.1 D-Sepearation in Undirected Graphical Models

The following three properties are used to determine if nodes are conditionally independent



$X_5 \perp X_1 | X_{V \setminus \{1,5\}}$

$X_5 \perp X_{V \setminus \text{ne}(5)} | X_{\text{ne}(5)}$, $\text{ne}(5) = \{2, 5\}$

$X_A \perp X_B | X_C$, where $X_A = \{X_a\}_{a \in A}$

Global Markov Property: $X_A \perp X_B | X_C$ iff (X_C) separates (X_A) from (X_B)

Local Markov Property: The set of nodes that renders a node conditionally independent of all the other nodes in the graph

$$X_j \perp X_{V - \{j, neighbour(j)\}} | X_{neighbour(j)}$$

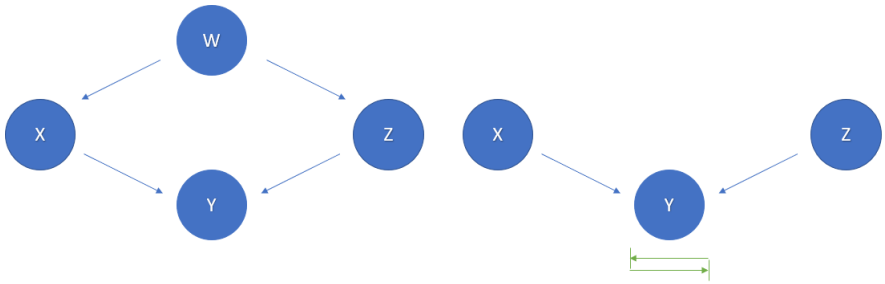
Pairwise Markov Property: The set of nodes that renders two nodes conditionally independent of each other.

$$X_j \perp X_i | X_{V - \{j,i\}}$$

It's obvious that global Markov implies local Markov which implies pairwise Markov.

2.2 limitations of UAGs and DAGs

Note, though we can represent new relations between variables, we can't represent others.

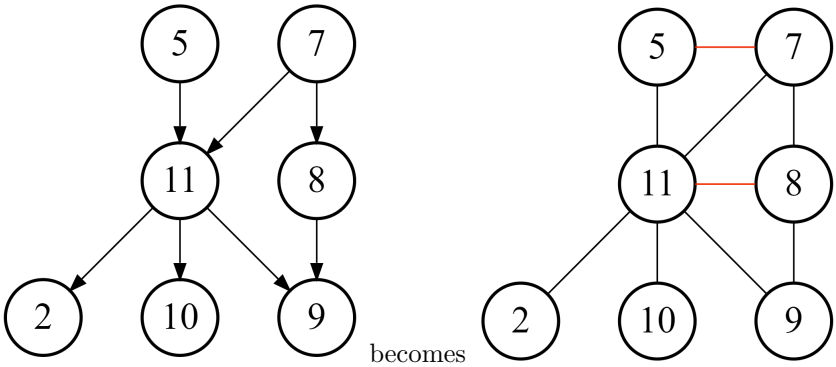


1 2

A DAG cant represent graph 1 where X and Z are conditionally dependent while a UAG can.
But a UAG cannot represent graph 2 where X and Z are conditionally independent without knowing Y.

2.3 Moralization

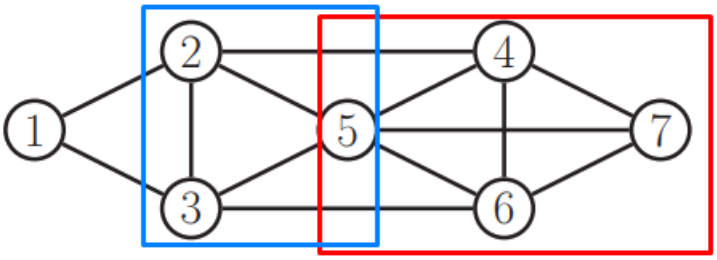
This was only mentioned in passing during lecture but a DAG can be converted to a UAG using Moralization
This is done adding edges between all pairs of non-adjacent nodes that have a common child, then making all edges in the graph undirected.



3 Cliques

A clique in an undirected graph is a subset of its vertices such that every two vertices in the subset are connected by an edge.

A maximal clique is a clique that cannot be extended by including one more adjacent vertex.
A maximum clique is a clique of the largest possible size in a graph.



The image shows 2 maximal cliques with the red clique also being the maximum clique

4 Hammersley-Clifford Theorem

Since there is no topological ordering for an undirected graph, we can't use the chain rule to represent p(y). So instead, we associate factors with each maximal clique.
We will denote the potential function for clique with $\psi_c(y_c|\theta_c)$ (back from the last week's notes) this can be any non-negative function.
The joint distribution is then defined to be proportional to the product of clique potentials.

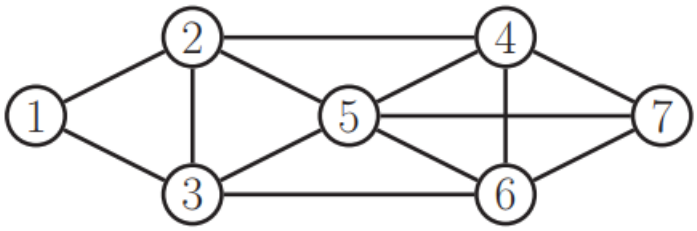
Hammersley-Clifford Theorem: A positive distribution $p(y) > 0$ satisfies the CI properties of an undirected graph G iff p can be represented as a product of factors, one per maximal clique, i.e.,

$$P(y|\theta) = \frac{1}{Z(\theta)} \prod_C \psi_c(y_c|\theta_c)$$

Where $Z(\theta)$ is the sum of all our possible values.

$$P(y|\theta) = \sum_C \psi_c(y_c|\theta_c)$$

going back to our example graph



$$p(y|\theta) = \frac{1}{Z(\theta)}\psi_{123}(y_1, y_2, y_3)\psi_{234}(y_2, y_3, y_4)\psi_{35}(y_3, y_5)$$

$$Z = \sum_y \psi_{123}(y_1, y_2, y_3)\psi_{234}(y_2, y_3, y_4)\psi_{35}(y_3, y_5)$$

this is useful because we can represent terms in terms of cliques instead of edges with reduces the number of terms in variable elimination

5 Energy Based Models

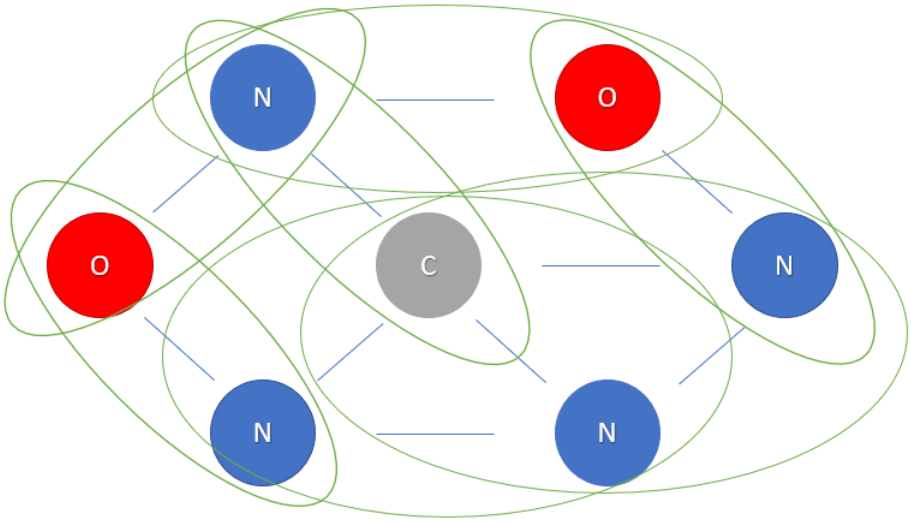
UGMs are very useful in physics. Take for example the Gibbs distribution used for modeling of Gibbs free Energy

$$p(y|) = \frac{1}{Z(\theta)} exp(\sum_c E(y_c|\theta_c))$$

where $E(y_c) > 0$ is the energy associated with the variables in clique c. We can convert this to UGM by defining.

$$\psi_c(y_c|\theta_c) = exp(E(y_c|\theta_c))$$

so we can now model the energy state of say, a protein molecule as a UGM.



not a molecule don't @ me