Directed Graphical Models
+
Undirected Graphical Models
Reminders

• Lecture 5.5: required video lecture
• Homework 2: Learning to Search for RNNs
  – Out: Sun, Sep 18
  – Due: Thu, Sep 29 at 11:59pm
• Poll Questions 0a and 0b about HW1
INTUITION FOR FACTOR GRAPHS

Representation of both directed and undirected graphical models
Joint Modeling

After we come up with a way to decompose our structure into variables, what comes next?

• We can define a joint model over those variables
• The joint model defines a score for each possible structure allowed by our decomposition
• The model should give high scores to “good” structures and low scores to “bad” structures
  – in probability terms: high scores for likely structures and low scores for unlikely structures
  – “likely structures” could be defined as those appearing in your training dataset
• (Hopefully, the joint model is also able to capture interesting interactions between pairs, triples, quadruples, ... of variables)
How do we write down a joint model?

(Factor Graphs)
Three Types of Graphical Models

- Directed Graphical Model
- Undirected Graphical Model
- Factor Graph
Factor Graphs

Factor Graph (bipartite graph)
- variables (circles)
- factors (squares)
Factor Graphs

Factor Graph (bipartite graph)
- variables (circles)
- factors (squares)

Each random variable can be assigned a value

The collection of values for all the random variables is called an assignment.
Factor Graph (bipartite graph)
- variables (circles)
- factors (squares)

Factors have local opinions about the assignments of their neighboring variables
Factor Graphs

Factor Graph
(bipartite graph)
- variables (circles)
- factors (squares)

Factors have local opinions about the assignments of their neighboring variables
Factor Graphs

**Factor Graph**
(bipartite graph)
- variables (circles)
- factors (squares)

Factors have local opinions about the assignments of their neighboring variables
Factor Graphs

\[ P(\text{tuna, ice cream}) = ? \]

Those opinions are expressed through potential tables
Factor Graphs

$$P(tuna, ice cream) = \frac{1}{Z} (6 * 7 * 0.1)$$

Uh-oh! The probabilities of the various assignments sum up to $$Z > 1$$.
So divide them all by $$Z$$.

The combined potential tables of all factors defines the probability of an assignment.
Bayesian Networks

DIRECTED GRAPHICAL MODELS
Three Types of Graphical Models

Directed Graphical Model

Undirected Graphical Model

Factor Graph
Bayesian Network

\[ p(X_1, X_2, X_3, X_4, X_5) = p(X_5|X_3)p(X_4|X_2, X_3) \]
\[ p(X_3)p(X_2|X_1)p(X_1) \]
Bayesian Network

Definition:

\[ P(X_1 \ldots X_n) = \prod_{i=1}^{n} P(X_i \mid \text{parents}(X_i)) \]

- A Bayesian Network is a **directed graphical model**
- It consists of a directed acyclic graph (DAG) \( G \) and the conditional probabilities \( P \)
- These two parts full specify the distribution:
  - Qualitative Specification: \( G \)
  - Quantitative Specification: \( P \)
Suppose we have an arbitrary directed graph $G$ over $T$ variables $X_i$ and define the following product:

$$P_{\text{fact}}(X) = \prod_{i=1}^{T} P(X_i|\text{parents}(X_i))$$

- **Proposition:** The function $P_{\text{fact}}(X)$ is a valid joint distribution when $G$ is a DAG.

- **Proof:** Let $X_s$ be a leaf node. By our factorization we have that,

$$P_{\text{fact}}(X) = P(X_s|\text{parents}(X_s))P_{\text{fact}}(\text{parents}(X_s))$$

By induction, if $P_{\text{fact}}(\text{parents}(X_s))$ is a valid joint distribution then $P_{\text{fact}}(X)$ is a valid joint distribution.
Qualitative Specification

• Where does the qualitative specification come from?
  
  – Prior knowledge of causal relationships
  – Prior knowledge of modular relationships
  – Assessment from experts
  – Learning from data (i.e. structure learning)
  – We simply prefer a certain architecture (e.g. a layered graph)
  – ...

© Eric Xing @ CMU, 2006-2011
Quantitative Specification

Example: Conditional probability tables (CPTs) for discrete random variables

\[
P(a,b,c,d) = P(a)P(b)P(c|a,b)P(d|c)
\]

<table>
<thead>
<tr>
<th></th>
<th>a^0</th>
<th>b^0</th>
<th>c^0</th>
<th>a^1</th>
<th>b^1</th>
<th>c^1</th>
</tr>
</thead>
<tbody>
<tr>
<td>a^0</td>
<td>0.75</td>
<td>0.33</td>
<td>0.45</td>
<td>1</td>
<td>0.9</td>
<td>0.7</td>
</tr>
<tr>
<td>a^1</td>
<td>0.25</td>
<td>0.67</td>
<td>0.55</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>c^0</th>
<th>c^1</th>
</tr>
</thead>
<tbody>
<tr>
<td>d^0</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>d^1</td>
<td>0.7</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Quantitative Specification

Example: Conditional probability density functions (CPDs) for continuous random variables

\[ A \sim \mathcal{N}(\mu_a, \Sigma_a) \quad B \sim \mathcal{N}(\mu_b, \Sigma_b) \]

\[ C \sim \mathcal{N}(A+B, \Sigma_c) \]

\[ D \sim \mathcal{N}(\mu_d+C, \Sigma_d) \]

\[ P(a,b,c,d) = P(a)P(b)P(c|a,b)P(d|c) \]
Quantitative Specification

Example: Combination of CPTs and CPDs for a mix of discrete and continuous variables

\[
P(a,b,c,d) = P(a)P(b)P(c|a,b)P(d|c)
\]

| a \( ^0 \) | 0.75 |
| a \( ^1 \) | 0.25 |
| b \( ^0 \) | 0.33 |
| b \( ^1 \) | 0.67 |

\( C \sim N(A+B, \Sigma_c) \)

\( D \sim N(\mu_d+C, \Sigma_d) \)
Compactness of a BayesNet

Consider random variables $X_1, X_2, \ldots, X_T$ where $X_i \in \mathcal{X}$, where $|\mathcal{X}| = R$

• To represent an arbitrary distribution $P(\mathbf{X})$ via a single joint probability table requires $R^T - 1$ values

• If the distribution factors according to a graph $G$ and $\max_{X_i} |\text{parents}(X_i)| \leq D$

then each $P(X_i \mid \text{parents}(X_i))$ needs only $R^D(R - 1)$ values for a total of only $T(R^D(R - 1))$ values

Observed Variables

• In a graphical model, shaded nodes are “observed”, i.e. their values are given

Example:

\[ P(X_2, X_5 \mid X_1 = 0, X_3 = 1, X_4 = 1) \]
Familiar Models as BayesNets

**Question:** Describe in words the directed graphical model that you would draw to represent an RNN-LM.

**Answer:**

**Question:** Describe in words the directed graphical model that you would draw to represent a seq2seq model.

**Answer:**
UNDIRECTED GRAPHICAL MODELS
Three Types of Graphical Models

Directied Graphical Model

Undirected Graphical Model

Factor Graph
Undirected Graphical Models

Undirected Graph Terminology

• **Definition**: a **clique** is a set of fully connected nodes (e.g. \{X_1, X_2\} or \{X_1, X_2, X_3\})

• **Definition**: a **maximal clique** is a clique to which adding any node makes it no longer a clique (e.g. \{X_1, X_2, X_3\} but not \{X_1, X_2\})

• **Definition**: a set of nodes \(X_C\) **separates** sets \(X_A\) and \(X_B\) if removing \(X_C\) leaves no path from a node in \(X_A\) to one in \(X_B\). (e.g. \{X_4, X_7\} separates \{X_1, X_2, X_3\} and \{X_5, X_6\})

**Notation**: Let \(X_S\) denote all the variables with indices in the set \(S \subseteq \mathbb{Z}^+\)
**Def:** an **undirected graphical model (UGM)** consists of a graph $G$ (qualitative specification) and potential functions $\psi$ (quantitative specification)

- The graph $G$ is an undirected graph over random variables $X_1, \ldots, X_T$ and cycles are permitted
- The potential functions $\psi$, also called “factors”, are used to define the joint probability

also called a Markov Random Field (MRF)
Def: Joint probability of a UGM

\[ p(x_1, \ldots, x_T) = \frac{1}{Z} \prod_{C \in C} \psi_C(x_C) \]

where \( C \) is the set of all cliques and \( C \in C \) is an index set \( \Rightarrow C \subseteq \{1, \ldots, T\} \)

1. we have one potential function (aka. factor) per clique
2. potential functions must be non-negative
   \[ \psi_C(x_C) \geq 0, \forall C, x_C \]
3. \( Z \) is the partition function \( \Rightarrow \) globally normalized model
   \[ Z = \sum_{x \in X} \prod_{C \in C} \psi_C(X_C) = \sum_{x \in X} s(x) \]
Undirected Graphical Models

**Def**: A distribution is said to factor according to the graph $G$ if it can be written as

$$p(x_1, \ldots, x_T) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

where $\mathcal{C}$ is the set of all cliques and $C \in \mathcal{C}$ is an index set $\Rightarrow C \subseteq \{1, \ldots, T\}$
Undirected Graphical Models

Ex: Joint probability of UGM

\[ p(\mathbf{x}) = \frac{1}{Z} \left( \psi_{1,2,3}(x_1, x_2, x_3) \psi_{2,3,4}(x_2, x_3, x_4) \right. \]

\[ \left. \psi_{4,5}(x_4, x_5) \psi_{5,6}(x_5, x_6) \psi_{4,7}(x_4, x_7) \right) \]
Potential Functions for UGM

How should we interpret the potential functions in a UGM?

• **Idea #1**: Maybe as **marginals** of the distribution? In general, no.

\[
p(x) \neq \frac{1}{Z} \left( p(x_1, x_2, x_3) p(x_2, x_3, x_4) \\
p(x_4, x_5) p(x_5, x_6) p(x_4, x_7) \right)
\]

• **Idea #2**: Maybe as **conditionals** of the distribution? In general, no.

\[
p(x) \neq \frac{1}{Z} \left( p(x_1 | x_2, x_3) p(x_2, x_3 | x_4) \\
p(x_4 | x_5) p(x_5 | x_6) p(x_7 | x_4) \right)
\]
Potential Functions for UGM

Whiteboard

– Simple example of potential functions as tables
Undirected Graphical Models

Whiteboard

– Alternate definition using maximal cliques
– Pairwise Markov Random Field (MRF)
Compactness of a UGM

Consider random variables $X_1, X_2, \ldots, X_T$ where $X_i \in \mathcal{X}$, where $|\mathcal{X}| = R$

- To represent an arbitrary distribution $P(\mathbf{X})$ via a single joint probability table requires $R^T - 1$ values.
- If the distribution factors according to a graph $G$ and $\max_{C \in \mathcal{C}} |C| \leq D$

then each $\psi_C(X_C)$ needs only $R^D$ values for a total of only $O(T(R^D))$ values.
CONDITIONAL INDEPENDENCIES OF DGMS
What Independencies does a Bayes Net Model?

• In order for a Bayesian network to model a probability distribution, the following must be true:

  Each variable is conditionally independent of all its non-descendants in the graph given the value of all its parents.

• This follows from

\[
P(X_1 \ldots X_n) = \prod_{i=1}^{n} P(X_i \mid \text{parents}(X_i))
\]

\[
= \prod_{i=1}^{n} P(X_i \mid X_1 \ldots X_{i-1})
\]

• But what else does it imply?
What Independencies does a Bayes Net Model?

Three cases of interest...

**Cascade**

```
Z
 |   
Y-----X
```

**Common Parent**

```
Y
 |   
X-----Z
```

**V-Structure**

```
X
 |   
Y-----Z
```
What Independencies does a Bayes Net Model?

Three cases of interest...

Cascade

Knowing Y **decouples** X and Z

Common Parent

Knowing Y **couples** X and Z

V-Structure
Proof of conditional independence

\( X \perp Z \mid Y \)

(The other two cases can be shown just as easily.)
The “Burglar Alarm” example

- Your house has a twitchy burglar alarm that is also sometimes triggered by earthquakes.
- Earth arguably doesn’t care whether your house is currently being burgled.
- While you are on vacation, one of your neighbors calls and tells you your home’s burglar alarm is ringing. Uh oh!

Quiz: True or False?

\[ \text{Burglar} \perp \text{Earthquake} \mid \text{PhoneCall} \]
**Markov Blanket (Directed)**

**Def:** the **co-parents** of a node are the parents of its children.

**Def:** the **Markov Blanket** of a node in a directed graphical model is the set containing the node’s parents, children, and co-parents.
**Markov Blanket (Directed)**

**Def:** the co-parents of a node are the parents of its children.

**Def:** the Markov Blanket of a node in a directed graphical model is the set containing the node’s parents, children, and co-parents.

**Example:** The Markov Blanket of $X_6$ is \{$X_3, X_4, X_5, X_8, X_9, X_{10}$\}

![Diagram of directed graphical model with nodes labeled $X_1$ to $X_{13}$ and categories Parents, Co-parents, and Children indicated with orange labels.](image-url)
Markov Blanket (Directed)

**Def:** the co-parents of a node are the parents of its children

**Def:** the Markov Blanket of a node in a directed graphical model is the set containing the node’s parents, children, and co-parents.

**Theorem:** a node is conditionally independent of every other node in the graph given its Markov blanket

**Example:** The Markov Blanket of $X_6$ is \{X_3, X_4, X_5, X_8, X_9, X_{10}\}
D-Separation

If variables $X$ and $Z$ are **d-separated** given a set of variables $E$
Then $X$ and $Z$ are **conditionally independent** given the set $E$

**Definition #1:**
Variables $X$ and $Z$ are **d-separated** given a set of evidence variables $E$
iff every path from $X$ to $Z$ is “blocked”.

A path is “blocked” whenever:

1. $\exists Y$ on path s.t. $Y \in E$ and $Y$ is a “common parent”

   ![Diagram 1](image1)

2. $\exists Y$ on path s.t. $Y \in E$ and $Y$ is in a “cascade”

   ![Diagram 2](image2)

3. $\exists Y$ on path s.t. $\{Y, \text{descendants}(Y)\} \notin E$ and $Y$ is in a “v-structure”

   ![Diagram 3](image3)
D-Separation

If variables X and Z are \textit{d-separated} given a \textit{set} of variables E
Then X and Z are \textit{conditionally independent} given the \textit{set} E

\textbf{Definition \#2:}
Variables X and Z are \textit{d-separated} given a \textit{set} of evidence variables E iff there does not exist a path in the \textit{undirected ancestral moral} graph with E removed.

1. \textbf{Ancestral graph:} keep only X, Z, E and their ancestors
2. \textbf{Moral graph:} add undirected edge between all pairs of each node’s parents
3. \textbf{Undirected graph:} convert all directed edges to undirected
4. \textbf{Givens Removed:} delete any nodes in E

\textbf{Example Query:} A ⊥ B | \{D, E\}

Original: \hspace{1cm} Ancestral: \hspace{1cm} Moral: \hspace{1cm} Undirected: \hspace{1cm} Givens Removed:

\[ \Rightarrow \text{A and B connected} \]
\[ \Rightarrow \text{not d-separated} \]
CONDITIONAL INDEPENDENCIES OF UGMS
**Undirected Graphical Models**

**Conditional Independence Semantics**

Consider a distribution over r.v.s $X_1, \ldots, X_T$

For a UGM and any disjoint index sets $A, B, C$, (i.e., $A \subseteq \{1, \ldots, T\}$, $B \subseteq \{1, \ldots, T\}$, $C \subseteq \{1, \ldots, T\}$)

$X_A$ is **conditionally independent** of $X_B$ given $X_C$ iff $X_C$ separates sets $X_A$ and $X_B$
**Markov Blanket (Directed)**

**Def:** the co-parents of a node are the parents of its children.

**Def:** the Markov Blanket of a node in a directed graphical model is the set containing the node’s parents, children, and co-parents.

**Theorem:** a node is conditionally independent of every other node in the graph given its Markov blanket.

**Example:** The Markov Blanket of $X_6$ is \{X_3, X_4, X_5, X_8, X_9, X_{10}\}
Markov Blanket (Undirected)

Def: the Markov Blanket of a node in an undirected graphical model is the set containing the node’s neighbors.

Theorem: a node is conditionally independent of every other node in the graph given its Markov blanket.

Example: The Markov Blanket of $X_6$ is $\{X_3, X_4, X_9, X_{10}\}$.
Undirected Graphical Models

Whiteboard

– Proof of independence by separation (simple case)
Non-equality of Directed / Undirected Graphical Models

There does **not** exist an **undirected** graphical model that can capture the conditional independence assumptions of this **directed** graphical model:

![Directed Graphical Model](image1)

There does **not** exist a **directed** graphical model that can capture the conditional independence assumptions of this **undirected** graphical model:

![Undirected Graphical Model](image2)
FACTOR GRAPHS

Representation of both directed and undirected graphical models
Three Types of Graphical Models

Directed Graphical Model

Undirected Graphical Model

Factor Graph
A joint distribution defines a probability $p(x)$ for each assignment of values $x$ to variables $X$. This gives the proportion of samples that will equal $x$. 

Sample 1:  
Sample 2:  
Sample 3:  
Sample 4:  
Sample 5:  
Sample 6:  

$X_0$  
$X_1$  
$X_2$  
$X_3$  
$X_4$  
$X_5$  

<START>  

time  
flies  
like  
an  
arow
A joint distribution defines a probability \( p(x) \) for each assignment of values \( x \) to variables \( X \). This gives the proportion of samples that will equal \( x \).
Sampling from a Joint Distribution

A joint distribution defines a probability $p(x)$ for each assignment of values $x$ to variables $X$. This gives the proportion of samples that will equal $x$. 

Sample 1:
- Time
- Flies
- Like
- An
- Arrow

Sample 2:
- Time
- Flies
- Like
- An
- Arrow

Sample 3:
- Flies
- Fly
- With
- Their
- Wings

Sample 4:
- With
- Time
- You
- Will
- See

A joint distribution defines a probability $p(x)$ for each assignment of values $x$ to variables $X$. This gives the proportion of samples that will equal $x$. 

Sample 1:
- Time
- Flies
- Like
- An
- Arrow

Sample 2:
- Time
- Flies
- Like
- An
- Arrow

Sample 3:
- Flies
- Fly
- With
- Their
- Wings

Sample 4:
- With
- Time
- You
- Will
- See
Factors have local opinions ($\geq 0$)

Each black box looks at some of the tags $X_i$ and words $W_i$

Note: We chose to reuse the same factors at different positions in the sentence.
Factors have local opinions (≥ 0)

Each black box looks at some of the tags $X_i$ and words $W_i$

$$p(n, v, p, d, n, \text{time}, \text{flies}, \text{like}, \text{an}, \text{arrow}) = ?$$
Global probability = product of local opinions

Each black box looks at some of the tags $X_i$ and words $W_i$

$$p(n, v, p, d, n, \text{time}, \text{flies}, \text{like}, \text{an}, \text{arrow}) = \frac{1}{Z} (4 \times 8 \times 5 \times 3 \times \ldots)$$

Uh-oh! The probabilities of the various assignments sum up to $Z > 1$. So divide them all by $Z$. 
Markov Random Field (MRF)

Joint distribution over tags $X_i$ and words $W_i$
The individual factors aren’t necessarily probabilities.

$$p(\text{time}, \text{flies}, \text{like}, \text{an}, \text{arrow}) = \frac{1}{\mathcal{Z}} (4 \times 8 \times 5 \times 3 \times \ldots)$$
Bayesian Networks

But sometimes we choose to make them probabilities. Constrain each row of a factor to sum to one. Now $Z = 1$.

$$p(n, v, p, d, n, \text{time, flies, like, an, arrow}) = \frac{1}{Z}(0.3 \times 0.8 \times 0.2 \times 0.5 \times \ldots)$$
Markov Random Field (MRF)

Joint distribution over tags $X_i$ and words $W_i$

$$p(n, v, p, d, n, \text{time, flies, like, an, arrow}) = \frac{1}{Z}(4 \times 8 \times 5 \times 3 \times \ldots)$$
Conditional Random Field (CRF)

Conditional distribution over tags $X_i$ given words $w_i$. The factors and $Z$ are now specific to the sentence $w$.

$$p(n, v, p, d, n \mid \text{time, flies, like, an, arrow}) = \frac{1}{Z} (4 \times 8 \times 5 \times 3 \times \ldots)$$
How General Are Factor Graphs?

• Factor graphs can be used to describe
  – Markov Random Fields (undirected graphical models)
    • i.e., log-linear models over a tuple of variables
  – Conditional Random Fields
  – Bayesian Networks (directed graphical models)

• Inference treats all of these interchangeably.
  – Convert your model to a factor graph first.
  – Pearl (1988) gave key strategies for exact inference:
    • Belief propagation, for inference on acyclic graphs
    • Junction tree algorithm, for making any graph acyclic
      (by merging variables and factors: blows up the runtime)
Factor Graph Notation

- Variables:
  \[ \mathcal{X} = \{X_1, \ldots, X_i, \ldots, X_n\} \]

- Factors:
  \[ \psi_\alpha, \psi_\beta, \psi_\gamma, \ldots \]
  where \( \alpha, \beta, \gamma, \ldots \subseteq \{1, \ldots, n\} \)

Joint Distribution

\[ p(x) = \frac{1}{Z} \prod_\alpha \psi_\alpha(x_\alpha) \]
Factors are Tensors

- **Factors:** \( \psi_\alpha, \psi_\beta, \psi_\gamma, \ldots \)
Converting to Factor Graphs

Each conditional and marginal distribution in a **directed GM** becomes a factor

Each maximal clique in an **undirected GM** becomes a factor
Equivalence of directed and undirected trees

- Any undirected tree can be converted to a directed tree by choosing a root node and directing all edges away from it.

- A directed tree and the corresponding undirected tree make the same conditional independence assertions.

- Parameterizations are essentially the same.

  - Undirected tree:
    \[ p(x) = \frac{1}{Z} \left( \prod_{i \in V} \psi(x_i) \prod_{(i,j) \in E} \psi(x_i, x_j) \right) \]

  - Directed tree:
    \[ p(x) = p(x_r) \prod_{(i,j) \in E} p(x_j | x_i) \]

  - Equivalence:
    \[ \psi(x_r) = p(x_r); \quad \psi(x_i, x_j) = p(x_j | x_i); \]
    \[ Z = 1, \quad \psi(x_i) = 1 \]
Factor Graph Examples

• Example 1
Factor Graph Examples

• Example 2

\[ \psi(x_1, x_2, x_3) = f_a(x_1, x_2)f_b(x_2, x_3)f_c(x_3, x_1) \]

• Example 3

\[ \psi(x_1, x_2, x_3) = f_a(x_1, x_2, x_3) \]
Tree-like Undirected GMs to Factor Trees

• Example 4
Poly-trees to Factor trees

- Example 5
Why factor graphs?

• Because FG turns tree-like graphs to factor trees,
• Trees are a data-structure that guarantees correctness of BP!
MRF VS. CRF
Markov Random Field (MRF):
• just a distribution over variables $y$
• partition function $Z$ is just a function of the parameters

$$p_\theta(y) = \frac{1}{Z(\theta)} \prod_\alpha \psi_\alpha(y_\alpha; \theta)$$

Conditional Random Field (CRF):
• conditions on some additional observed variables $x$
• partition function $Z$ is a function of $x$ as well

$$p_\theta(y \mid x) = \frac{1}{Z(x; \theta)} \prod_\alpha \psi_\alpha(y_\alpha, x; \theta)$$
TYPES OF GRAPHICAL MODELS
Three Types of Graphical Models

Directed Graphical Model

Undirected Graphical Model

Factor Graph
Key Concepts for Graphical Models

Graphical Models in General
1. A graphical model defines a family of probability distributions
2. That family shares in common a set of conditional independence assumptions
3. By choosing a parameterization of the graphical model, we obtain a single model from the family
4. The model may be either locally or globally normalized

Ex: Directed G.M.
1. Family: directed graphs with locally normalized conditional probabilities
2. Conditional Independencies: d-separation, Markov blanket
3. Example parameterization: conditional probability tables (CPTs) for discrete var.s, conditional probability densities for continuous var.s
4. Normalization: locally normalized, partition function is always 1.0
Key Concepts for Graphical Models

Graphical Models in General
1. A graphical model defines a family of probability distributions
2. That family shares in common a set of conditional independence assumptions
3. By choosing a parameterization of the graphical model, we obtain a single model from the family
4. The model may be either locally or globally normalized

Ex: Undirected G.M.
1. Family: undirected graphs with unnormalized potentials
2. Conditional Independencies: independence by separation, Markov blanket
3. Example parameterization: Markov random field (MRF), conditional random field (CRF), neural potentials
4. Normalization: globally normalized
**Key Concepts for Graphical Models**

<table>
<thead>
<tr>
<th>Graphical Models in General</th>
<th>Ex: Factor Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A graphical model defines a family of probability distributions</td>
<td>1. <strong>Family</strong>: bipartite graph over variables and factors</td>
</tr>
<tr>
<td>2. That family shares in common a set of <strong>conditional independence assumptions</strong></td>
<td>2. <strong>Conditional Independencies</strong>: independence by separation, inferable from underlying DGM or UGM</td>
</tr>
<tr>
<td>3. By choosing a <strong>parameterization</strong> of the graphical model, we obtain a single <strong>model</strong> from the family</td>
<td>3. <strong>Example parameterization</strong>: any DGM parameterization, any UGM parameterization</td>
</tr>
<tr>
<td>4. The model may be either <strong>locally or globally normalized</strong></td>
<td>4. <strong>Normalization</strong>: locally normalized if based on DGM, globally normalized if based on UGM</td>
</tr>
</tbody>
</table>