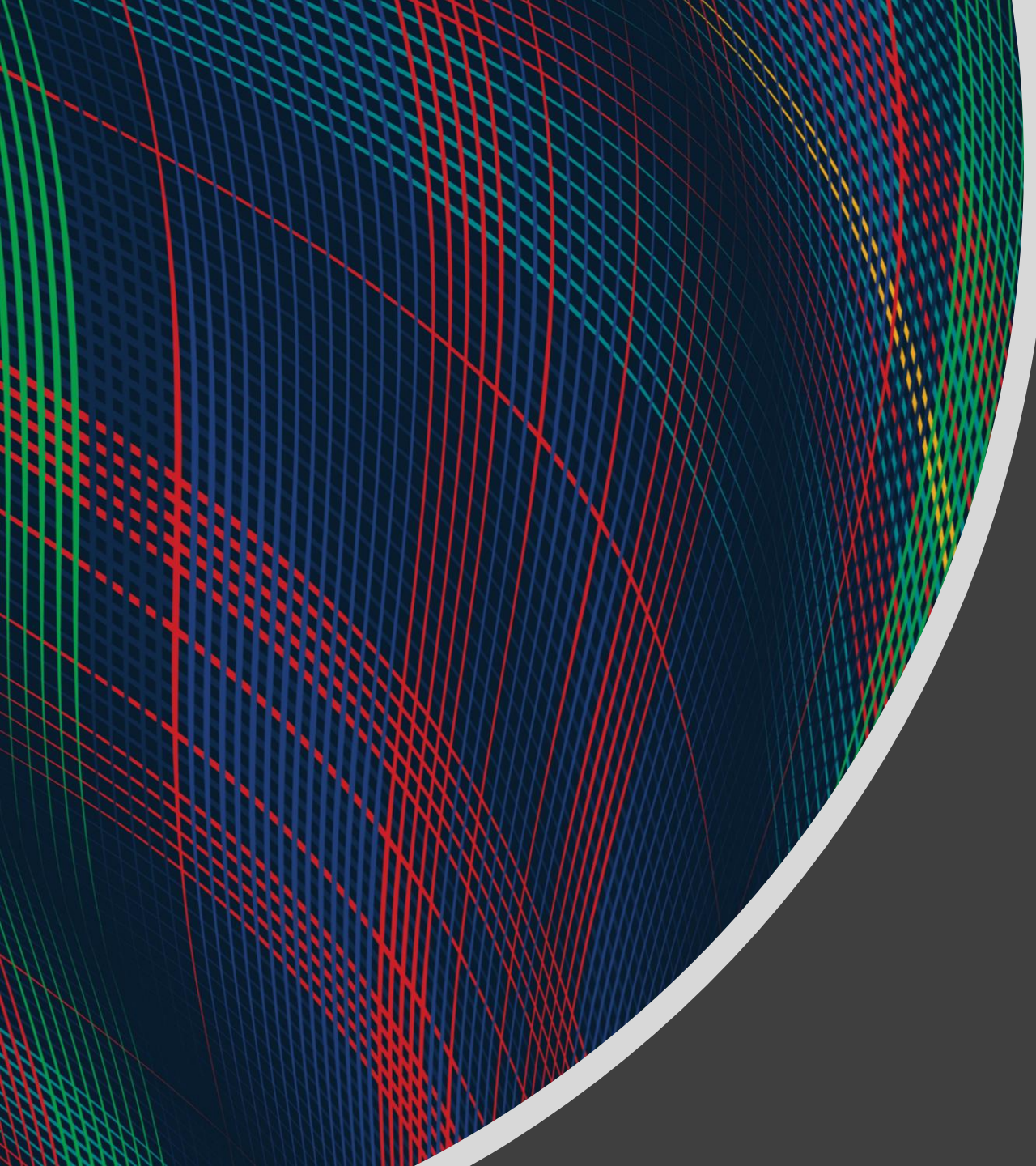


An abstract graphic on the left side of the slide, featuring a sphere-like shape composed of a dense grid of intersecting red, green, and blue lines. The lines are curved and follow the contours of the sphere, creating a complex, woven pattern. The colors are vibrant against a dark background.

10-607 Computational Foundations for Machine Learning

Graphs and Bayes Nets

Instructor: Pat Virtue

Plan

Wrap up trees

Graphs

- Examples
- Implementations
- Dense vs sparse structures

Probabilistic Graphical Models

- Bayes Nets
- Factors
- Variable Elimination

Trees

Previous lecture slides

Plan

Wrap up trees

Graphs

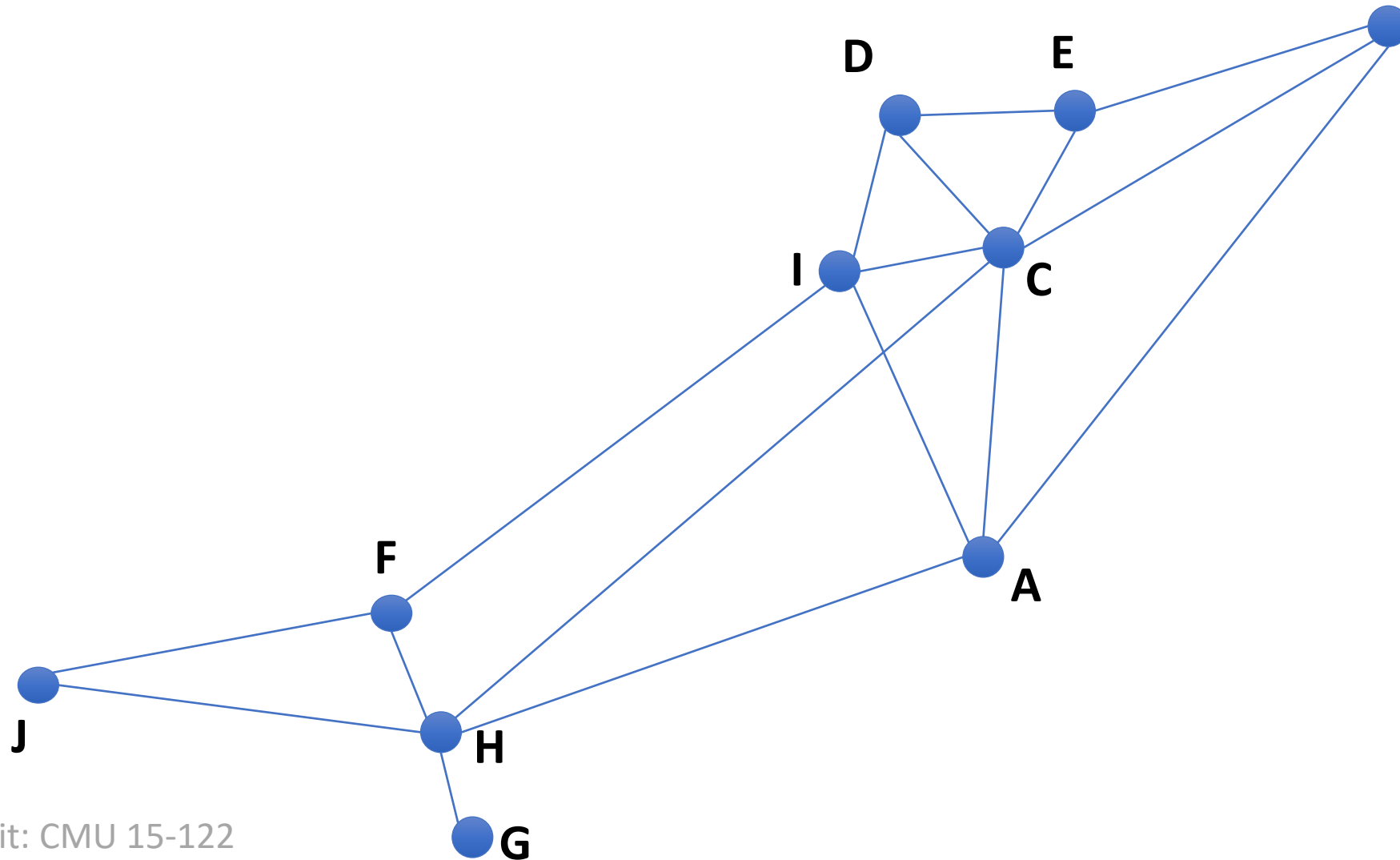
- Examples
- Implementations
- Dense vs sparse structures

Probabilistic Graphical Models

- Bayes Nets
- Factors
- Variable Elimination

Graphs

Defined by vertices and edges



Vertices (or nodes):

A, B, C, D, E, F, G, H, I, J

Edges:

(A,B) (A,C)

(A,H) (B,C)

(B,E) (C,D)

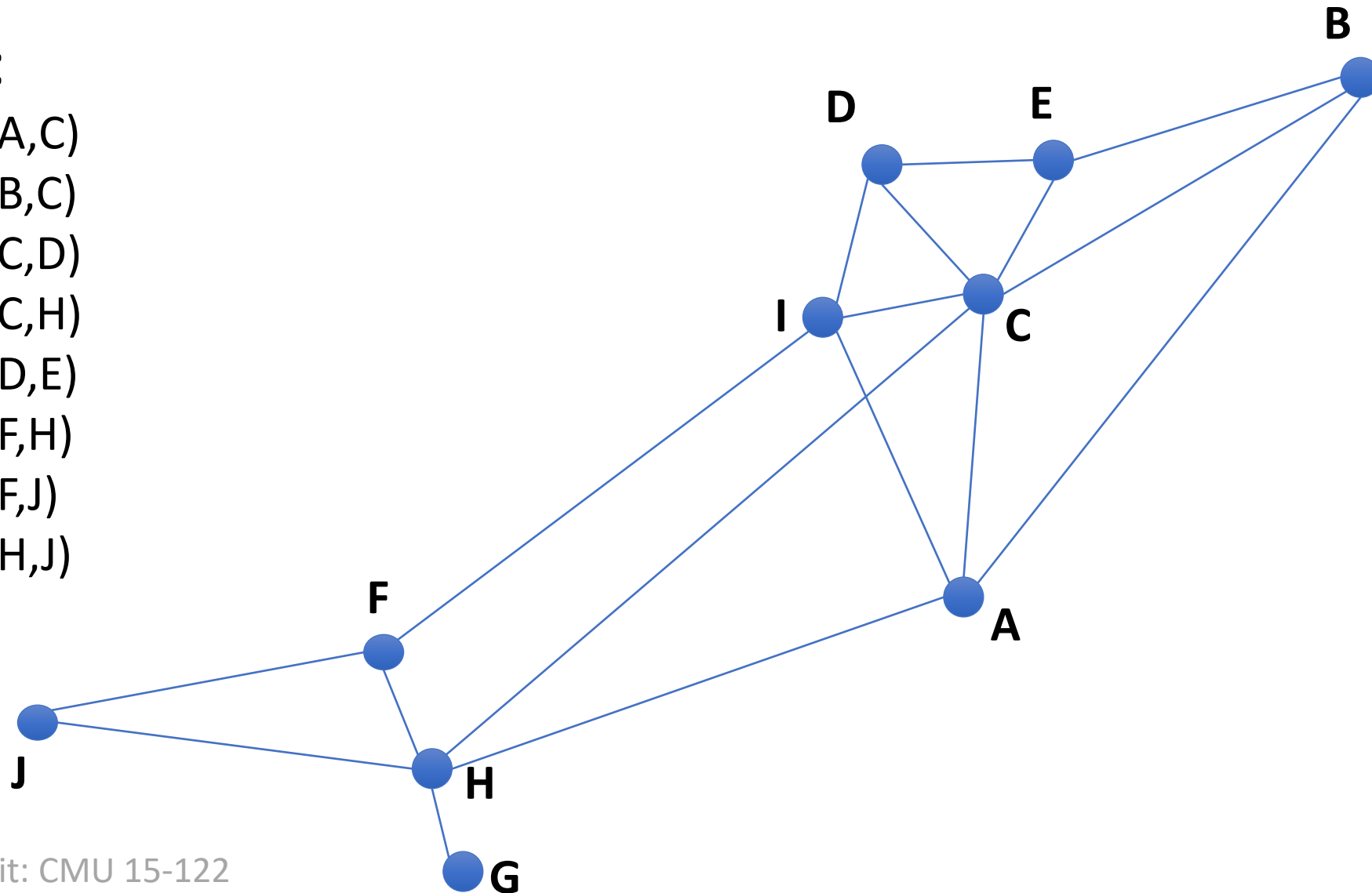
(C,E) (C,H)

(C,I) (D,E)

(D,I) (F,H)

(F,I) (F,J)

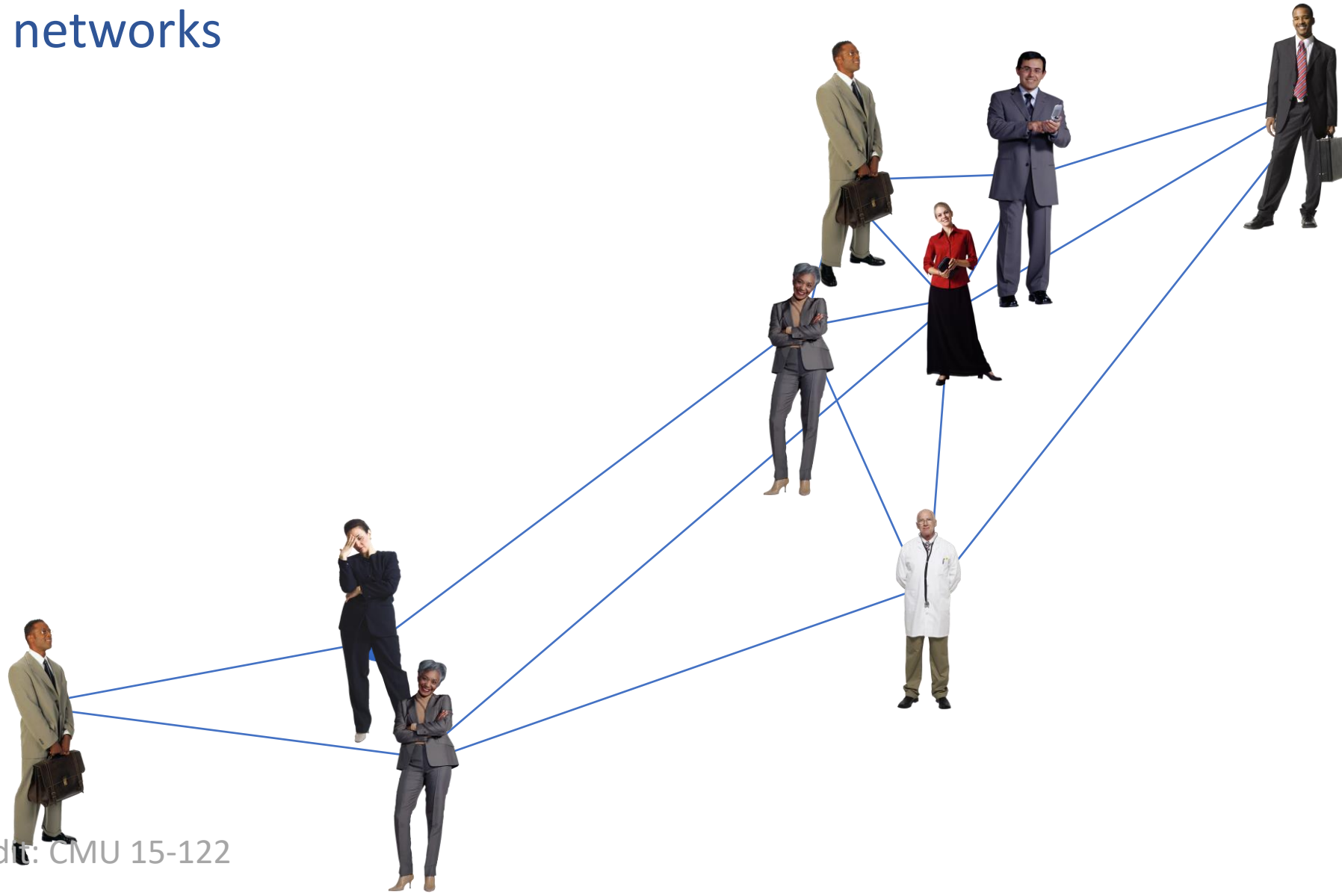
(G,H) (H,J)



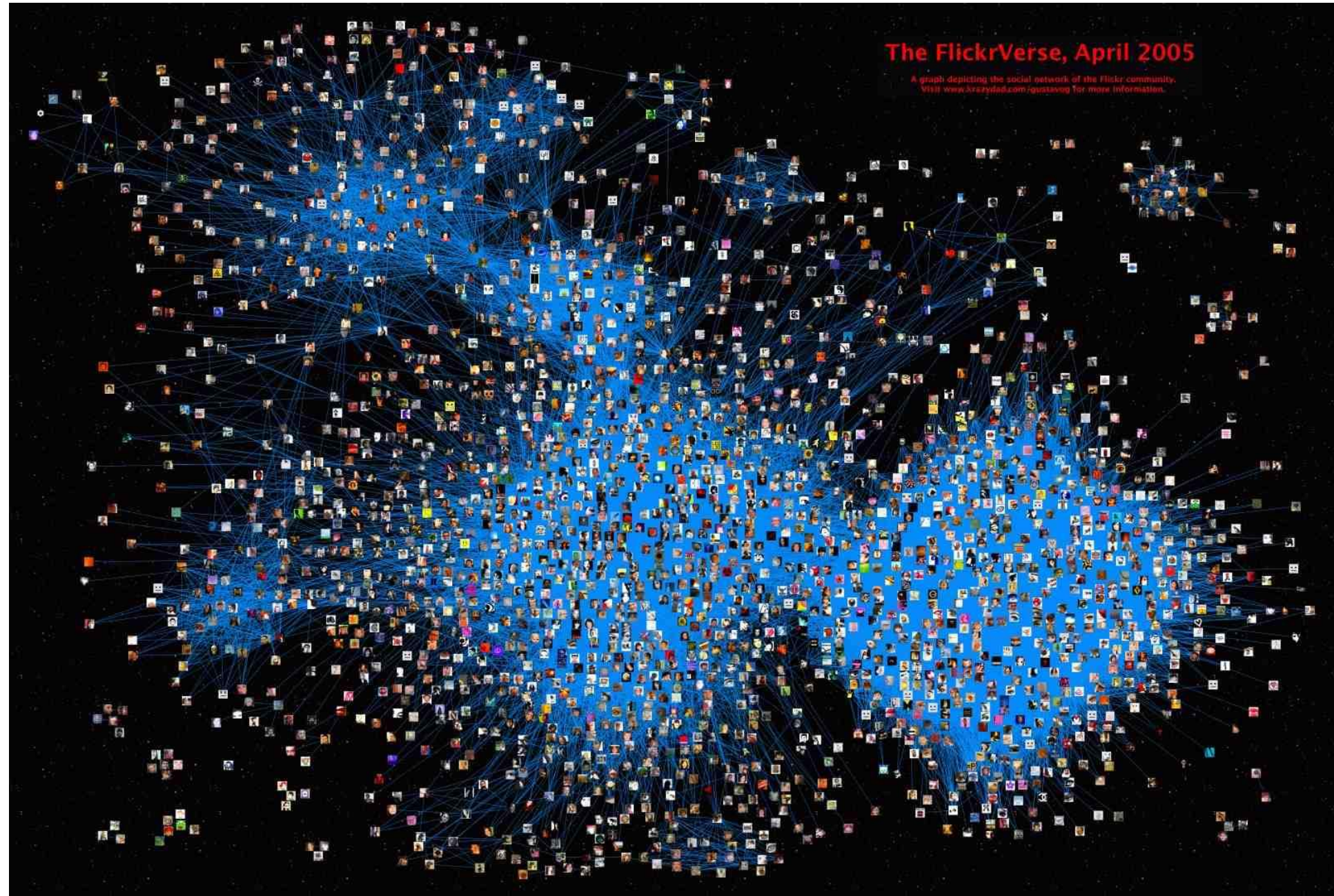


Graphs

Social networks

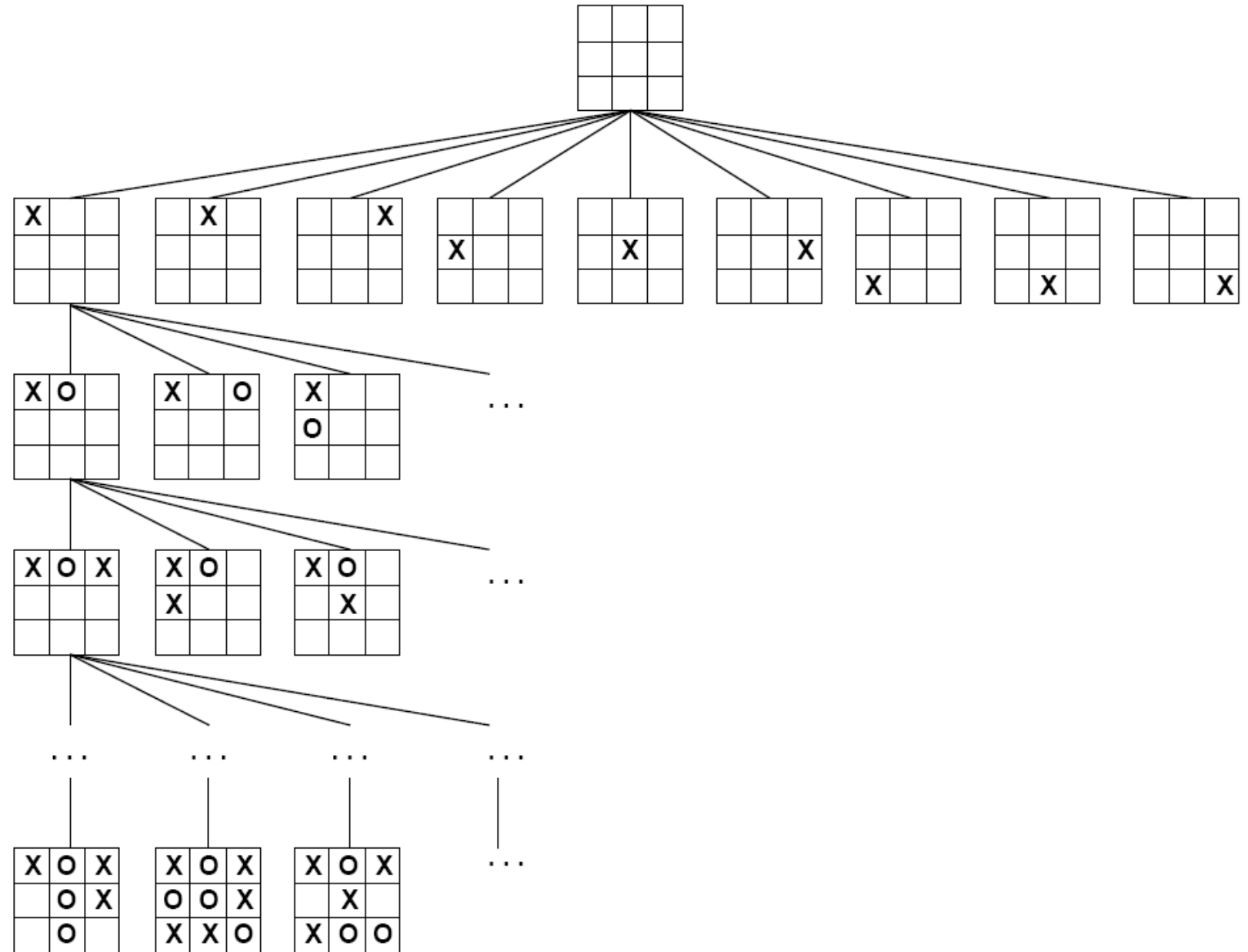


Flickr, 2005



Graphs

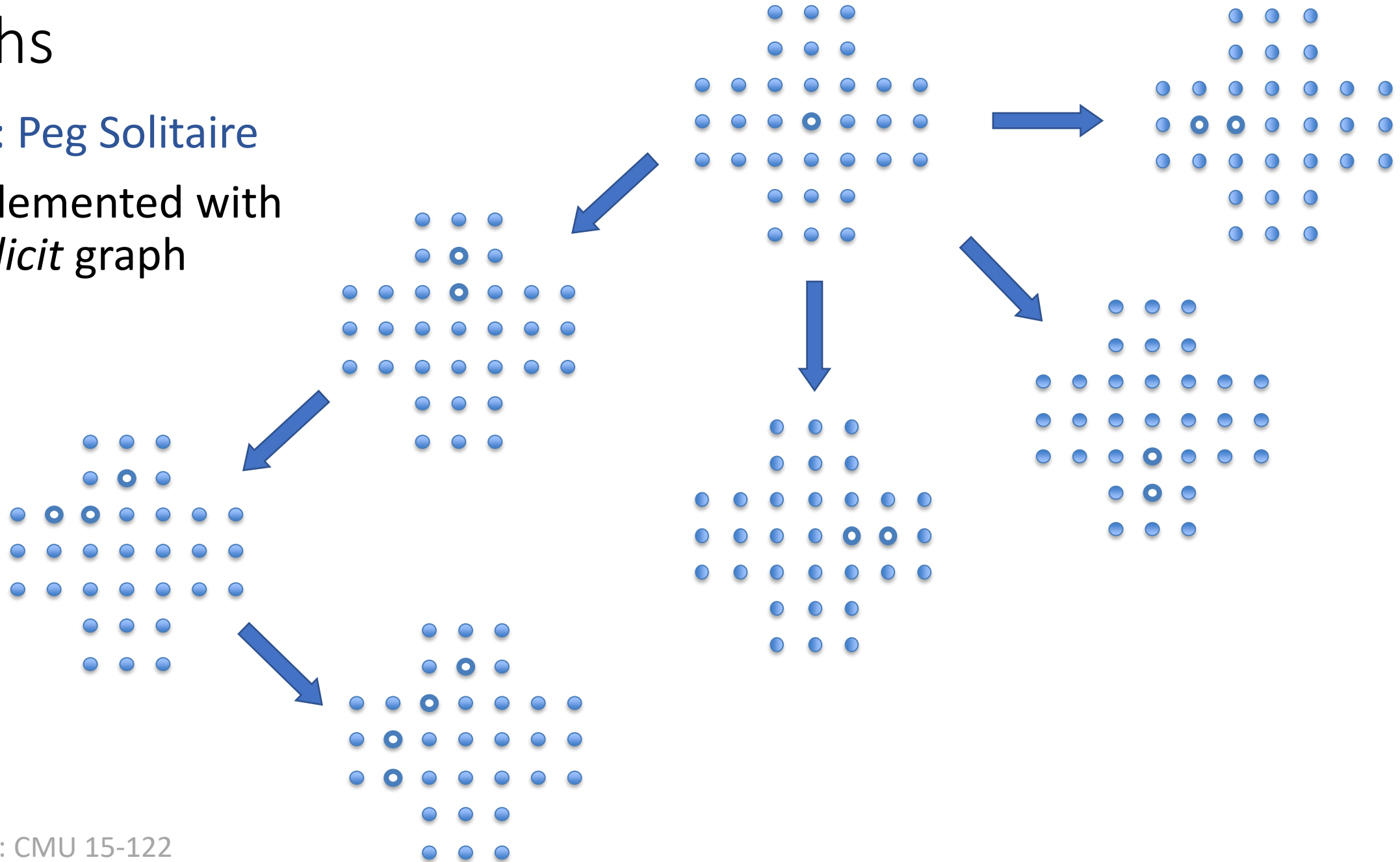
Games: Tic-tac-toe



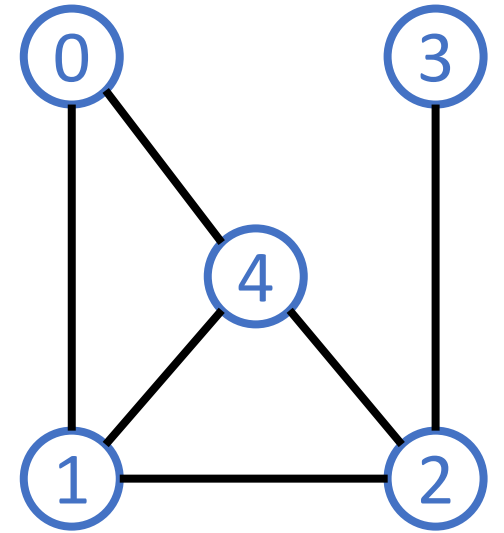
Graphs

Games: Peg Solitaire

- Implemented with *implicit* graph

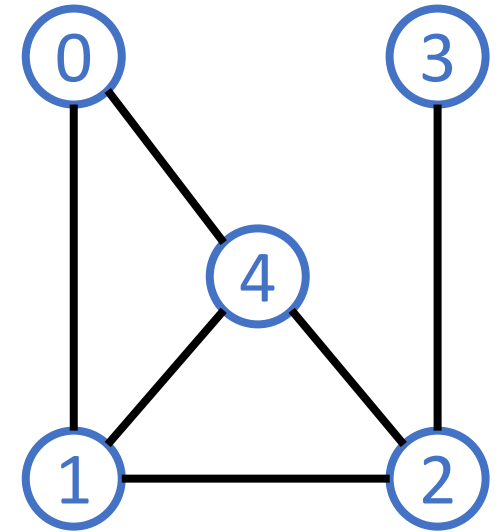


Question: How could you implement a graph?

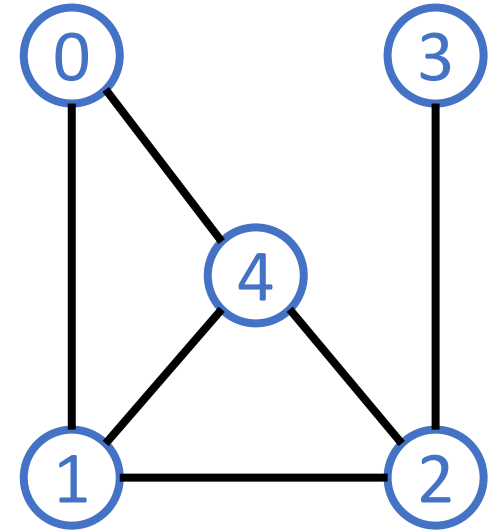


Graph as an adjacency matrix (AM)

	0	1	2	3	4
0					
1					
2					
3					
4					

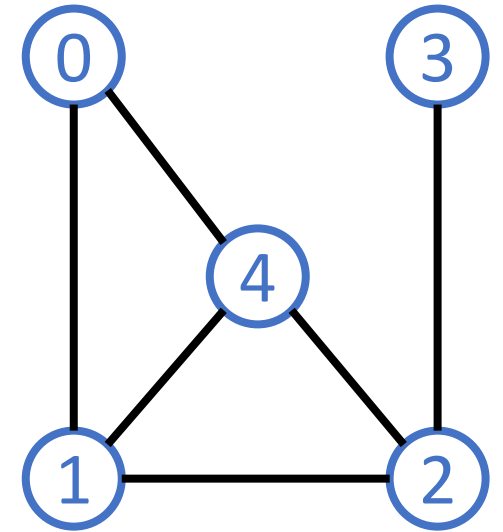


Graph as an adjacency list (AL)

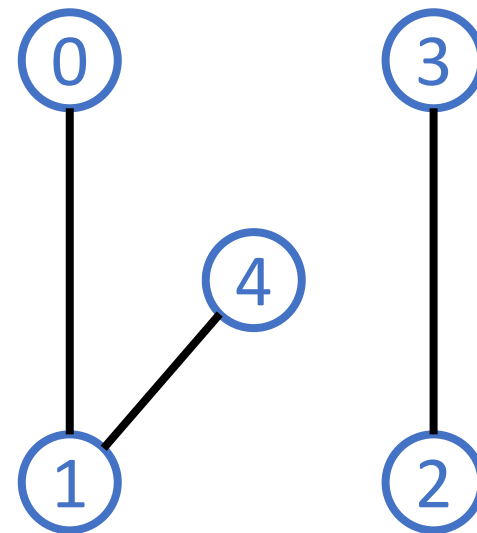
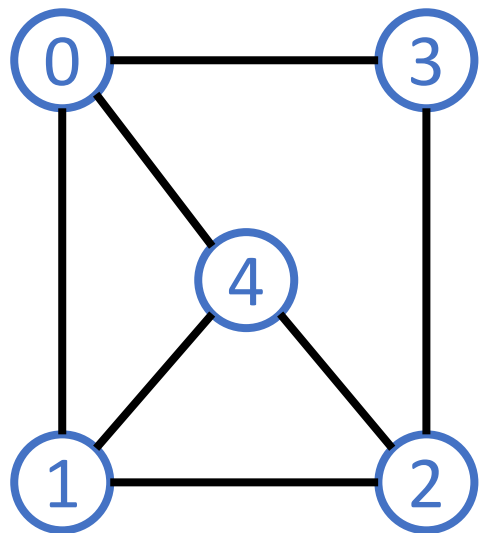


Graph as an adjacency list (AL)

0	1	4		
1	0	2	4	
2	1	4	3	
3	2			
4	0	1	2	



Dense vs Sparse



Which implementation do we use?

Adjacency matrix vs adjacency list

Considerations

- Density
- Storage/Memory
- Checking for edge
- Adding an edge
- Getting neighbors of vertex
- Target algorithm

Plan

Wrap up trees

Graphs

- Examples
- Implementations
- Dense vs sparse structures

Probabilistic Graphical Models

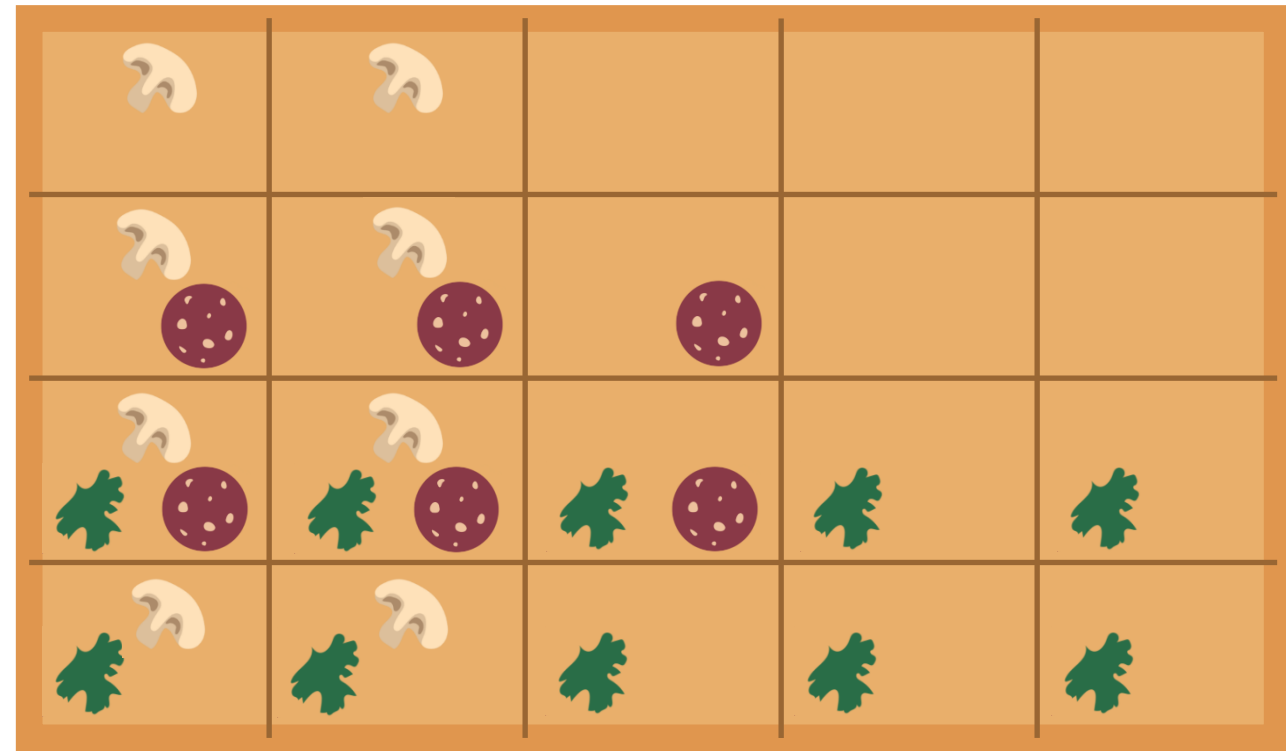
- Bayes Nets
- Factors
- Variable Elimination

Markov Chains

Answer Any Query from Joint Distribution

What is the probability of getting a slice with:

- 1) No mushrooms
 - 2) Spinach and no mushrooms
 - 3) Spinach, when asking for slice with no mushrooms
- Mushrooms
 - Spinach
 - No spinach
 - No spinach and mushrooms
 - No spinach when asking for no mushrooms
 - No spinach when asking for mushrooms
 - Spinach when asking for mushrooms
 - No mushrooms and no spinach



Icons: CC, <https://openclipart.org/detail/296791/pizza-slice>

Answer Any Query from Joint Distribution

You can answer all of these questions:

$P(M)$	
m_1	12/20
m_2	

$P(S)$	
s_1	
s_2	

$P(M, S)$		
m_1	s_1	
m_1	s_2	6/20
m_2	s_1	
m_2	s_2	

$P(M s_1)$	
m_1	
m_2	

$P(M s_2)$	
m_1	
m_2	

$P(S m_1)$	
s_1	
s_2	6/12

$P(S m_2)$	
s_1	
s_2	

Answer Any Query from Joint Distribution

$P(\text{Weather})?$

$P(\text{Weather} \mid \text{winter})?$

$P(\text{Weather} \mid \text{winter, hot})?$

Season	Temp	Weather	$P(S, T, W)$
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Answer Any Query from Joint Distribution

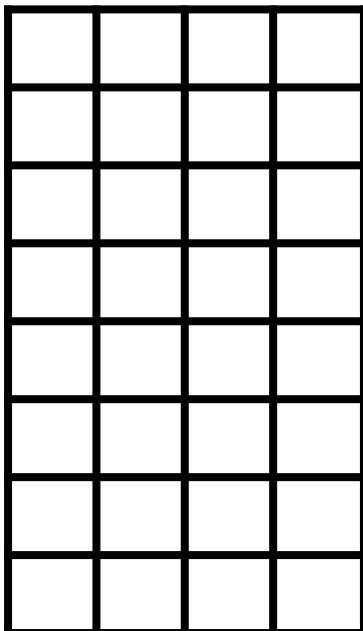
P(Weather)?

Season	Temp	Weather	P(S, T, W)
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Answer Any Query from Joint Distribution

Joint distributions are the best!

Joint





Query

$$P(q_1, q_2 \mid e_1, e_2, e_3)$$

Answer Any Query from Joint Distribution

Two tools to go from joint to query

1. Definition of conditional probability

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

2. Law of total probability (marginalization, summing out)

$$P(A) = \sum_b P(A, b)$$

$$P(Y \mid U, V) = \sum_x \sum_z P(x, Y, z \mid U, V)$$

Answer Any Query from Joint Distribution

Two tools to go from joint to query

Joint: $P(H_1, H_2, Q, E)$

Query: $P(Q | e)$

1. Definition of conditional probability

$$P(Q|e) = \frac{P(Q, e)}{P(e)}$$

2. Law of total probability (marginalization, summing out)

$$P(Q, e) = \sum_{h_1} \sum_{h_2} P(h_1, h_2, Q, e)$$

$$P(e) = \sum_q \sum_{h_1} \sum_{h_2} P(h_1, h_2, q, e)$$

Answer Any Query from Joint Distribution

$P(\text{Weather})?$

$P(\text{Weather} \mid \text{winter})?$

$P(\text{Weather} \mid \text{winter, hot})?$

Season	Temp	Weather	$P(S, T, W)$
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Answer Any Query from Joint Distribution

Joint distributions are the best!

Problems with joints

- We aren't given the joint table
 - Usually some set of conditional probability tables

Joint

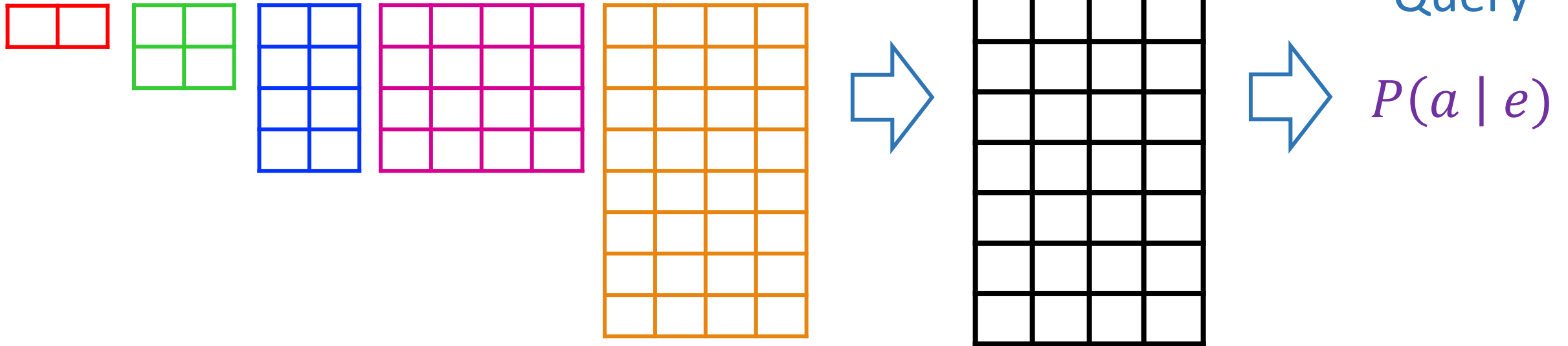


Query

$$P(a | e)$$

Build Joint Distribution Using Chain Rule

Conditional Probability Tables
and Chain Rule



$$P(A) \ P(B|A) \ P(C|A, B) \ P(D|A, B, C) \ P(E|A, B, C, D)$$

Build Joint Distribution Using Chain Rule

Two tools to construct joint distribution

1. Product rule

$$P(A, B) = P(A | B)P(B)$$
$$P(A, B) = P(B | A)P(A)$$

2. Chain rule

$$P(X_1, X_2, \dots, X_n) = \prod_i P(X_i | X_1, \dots, X_{i-1})$$

$$P(A, B, C) = P(A)P(B | A)P(C | A, B) \quad \text{for ordering } A, B, C$$

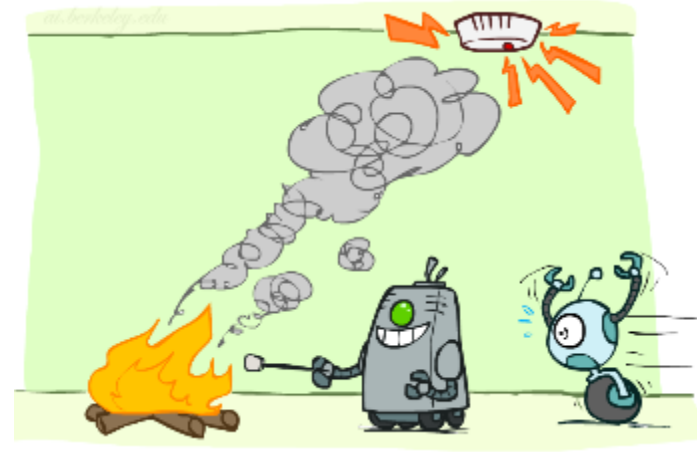
$$P(A, B, C) = P(A)P(C | A)P(B | A, C) \quad \text{for ordering } A, C, B$$

$$P(A, B, C) = P(C)P(B | C)P(A | C, B) \quad \text{for ordering } C, B, A$$

Build Joint Distribution Using Chain Rule

Binary random variables

- Fire
- Smoke
- Alarm



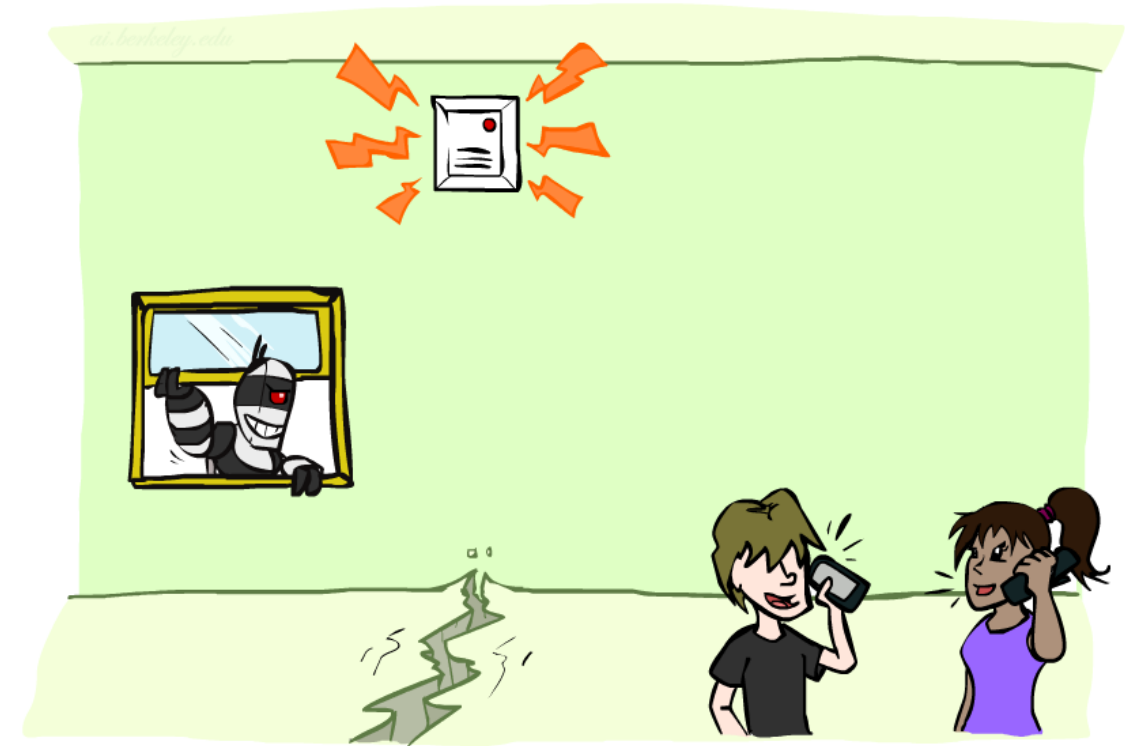
Poll 3

Variables

- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!

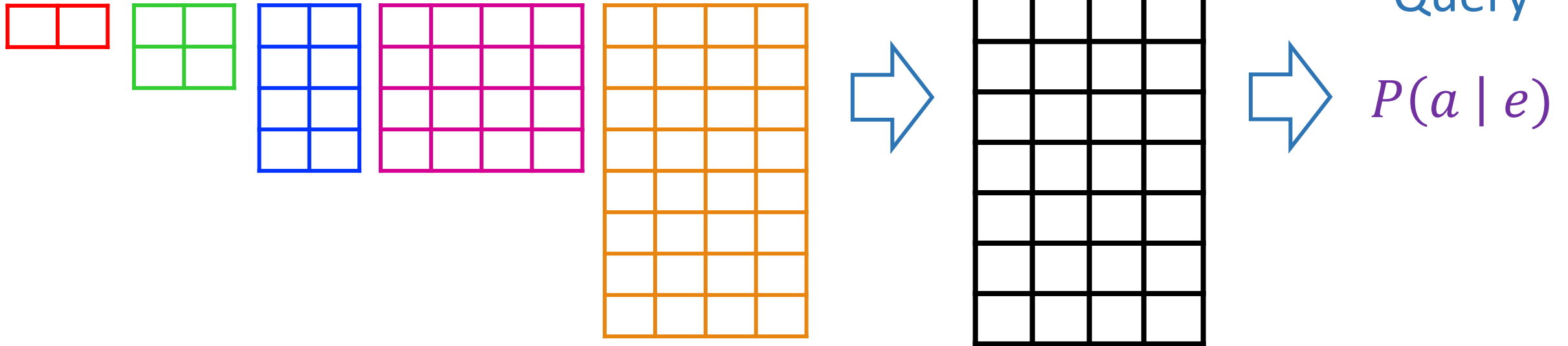
How many different ways can we write the chain rule?

- A.* 1
- B.* 5
- C.* 5 choose 5
- D.* 5!
- E.* 5^5



Build Joint Distribution Using Chain Rule

Conditional Probability Tables
and Chain Rule



$$P(A) \ P(B|A) \ P(C|A, B) \ P(D|A, B, C) \ P(E|A, B, C, D)$$

Answer Any Query from Condition Probability Tables

Process to go from (specific) conditional probability tables to query

1. Construct the joint distribution
 1. Product Rule or Chain Rule
2. Answer query from joint
 1. Definition of conditional probability
 2. Law of total probability (marginalization, summing out)

Answer Any Query from Condition Probability Tables

Bayes' rule as an example

Given: $P(E|Q)$, $P(Q)$ Query: $P(Q | e)$

1. Construct the **joint** distribution

1. Product Rule or Chain Rule

$$P(E, Q) = P(E|Q)P(Q)$$

2. Answer **query** from **joint**

1. Definition of conditional probability

$$P(Q | e) = \frac{P(e, Q)}{P(e)}$$

2. Law of total probability (marginalization, summing out)

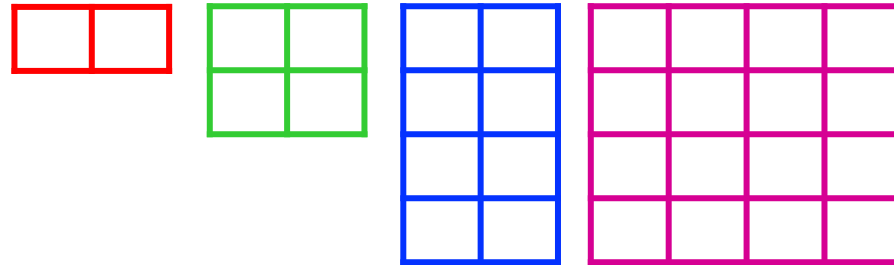
$$P(Q | e) = \frac{P(e, Q)}{\sum_q P(e, q)}$$

Bayesian Networks

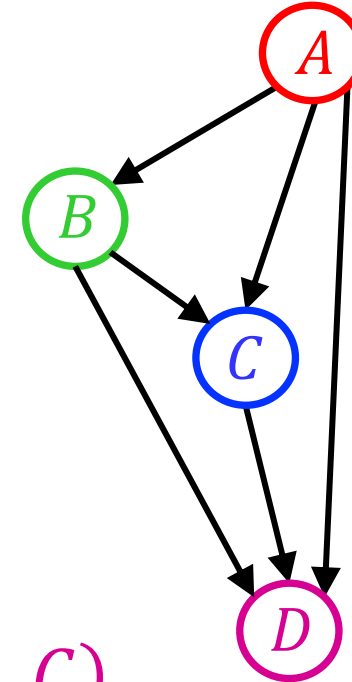
One node per random variable

DAG

One CPT per node: $P(\text{node} \mid \text{Parents}(\text{node}))$



Bayes net



$$P(A, B, C, D) = P(A) P(B|A) P(C|A, B) P(D|A, B, C)$$

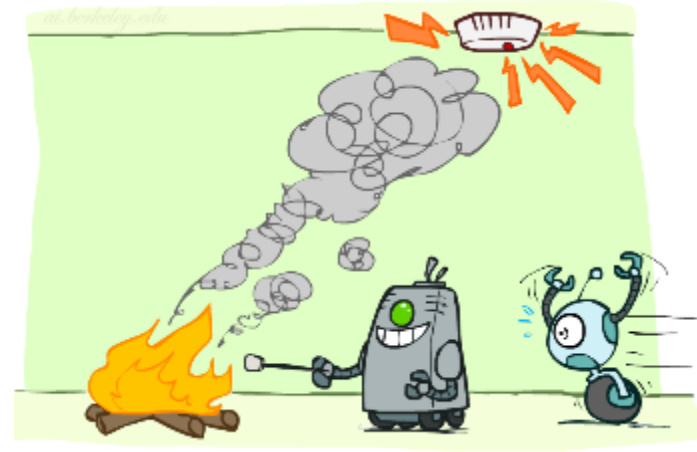
Encode joint distributions as product of conditional distributions on each variable

$$P(X_1, \dots, X_N) = \prod_i P(X_i \mid \text{Parents}(X_i))$$

Build Bayes Net Using Chain Rule

Binary random variables

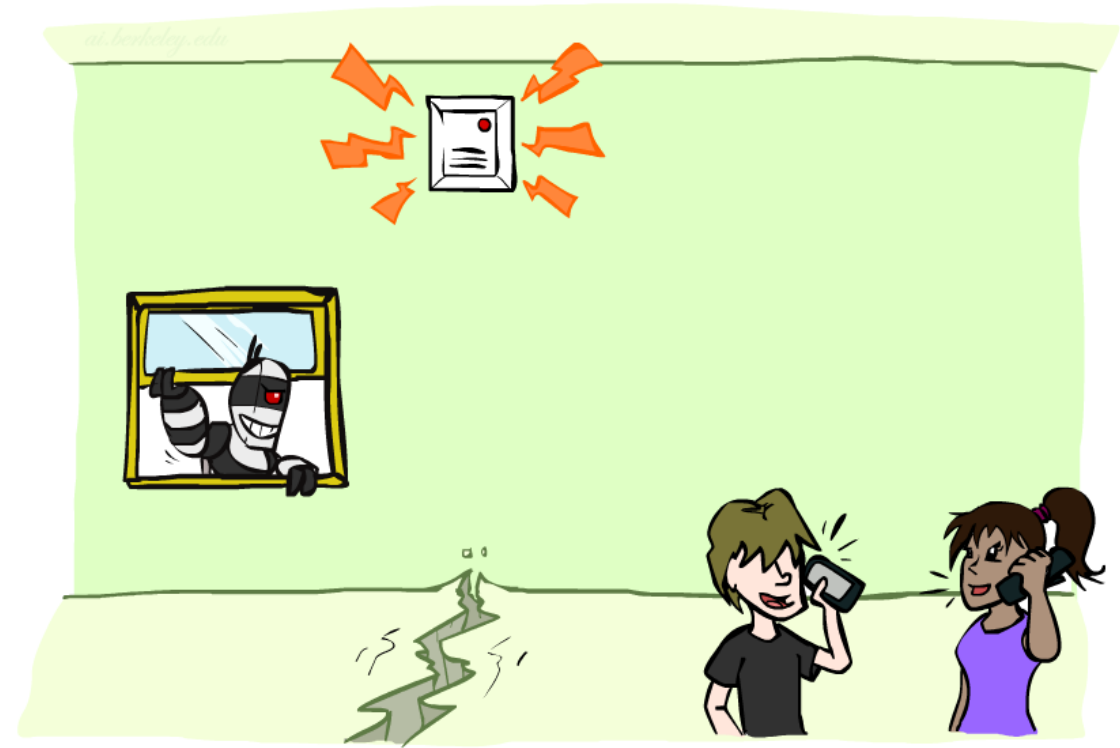
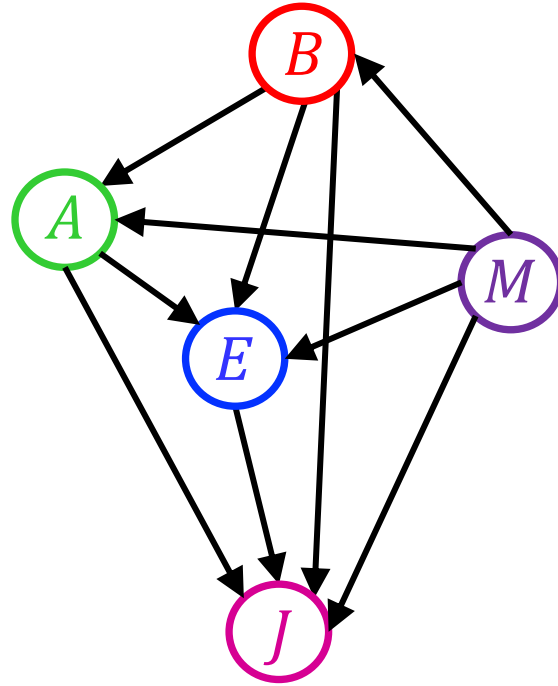
- Fire
- Smoke
- Alarm



Question

Variables

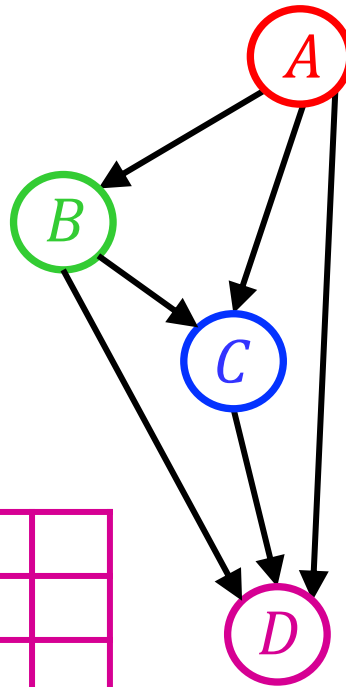
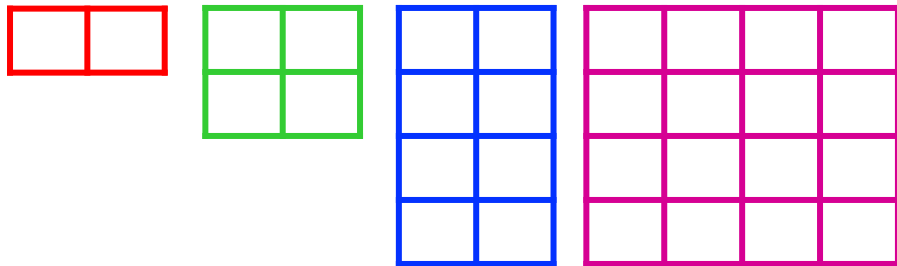
- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!



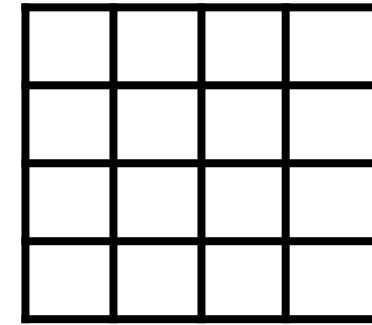
Given the Bayes net, write the joint distribution?

Answer Any Query from Bayes Net

Bayes Net and
Conditional
Probability Tables



Joint

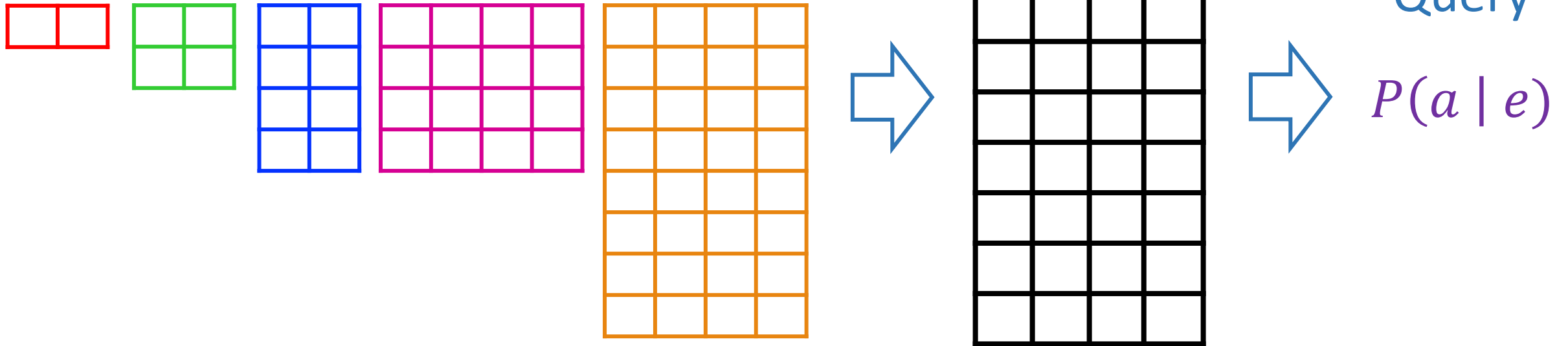


Query

$$P(a | e)$$

Answer Any Query from Condition Probability Tables

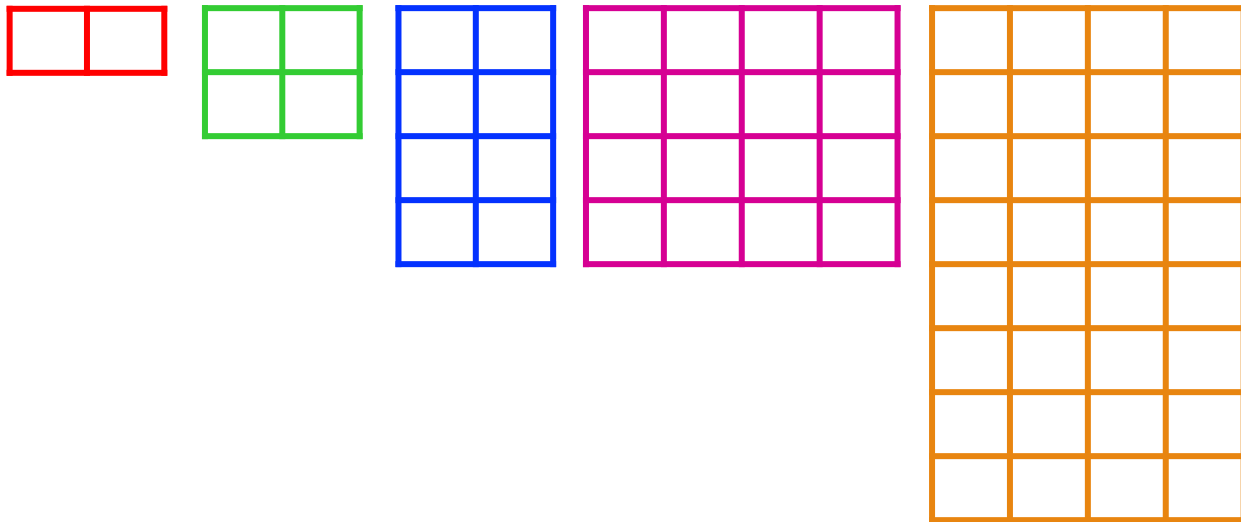
Conditional Probability Tables
and Chain Rule



$$P(A) \ P(B|A) \ P(C|A, B) \ P(D|A, B, C) \ P(E|A, B, C, D)$$

Answer Any Query from Condition Probability Tables

Conditional Probability Tables and Chain Rule



$$P(A) \ P(B|A) \ P(C|A, B) \ P(D|A, B, C) \ P(E|A, B, C, D)$$

Problems

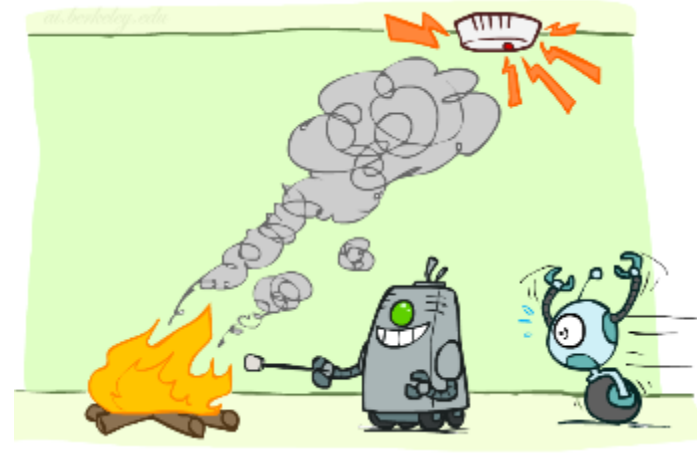
- Huge
 - n variables with d values
 - d^n entries
- We aren't given the right tables

Danielle Belgrave, Microsoft Research

Do We Need the Full Chain Rule?

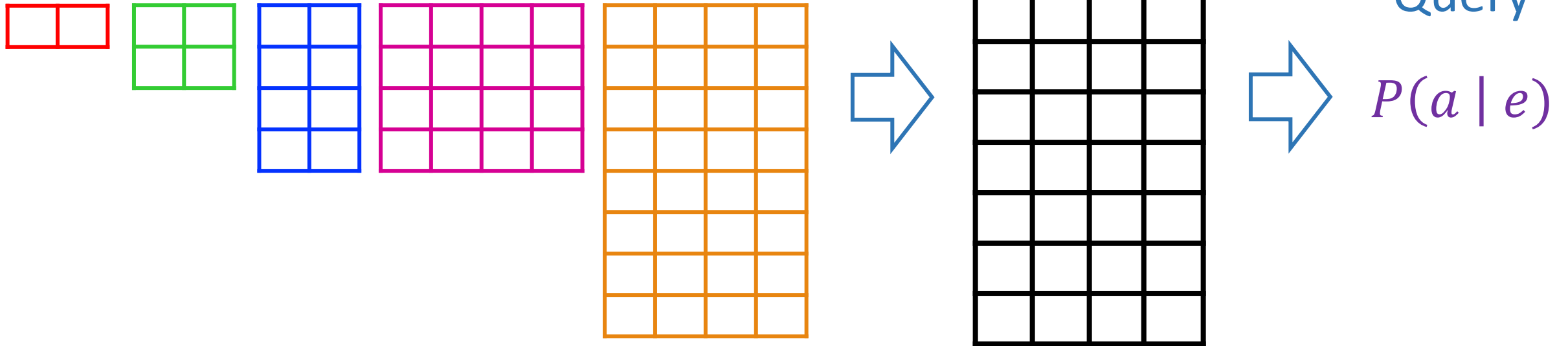
Binary random variables

- Fire
- Smoke
- Alarm



Answer Any Query from Condition Probability Tables

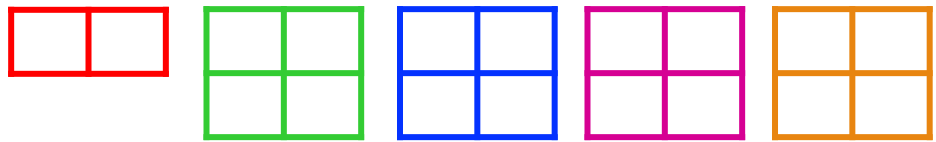
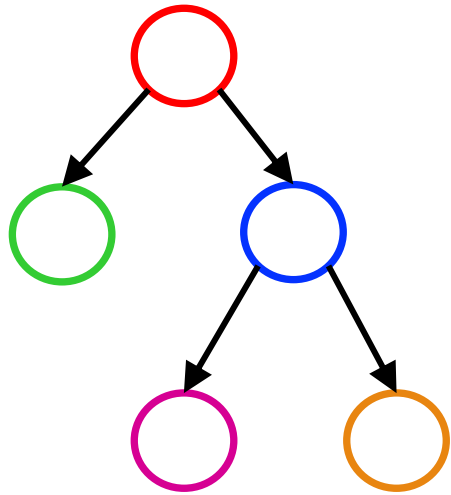
Conditional Probability Tables
and Chain Rule



$$P(A) \ P(B|A) \ P(C|A, B) \ P(D|A, B, C) \ P(E|A, B, C, D)$$

Answer Any Query from Condition Probability Tables

Bayes Net



$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

$$P(X_1, \dots, X_N) = \prod_i P(X_i | \text{Parents}(X_i))$$

Joint

Query

$P(a | e)$

Plan

Wrap up trees

Graphs

- Examples
- Implementations
- Dense vs sparse structures

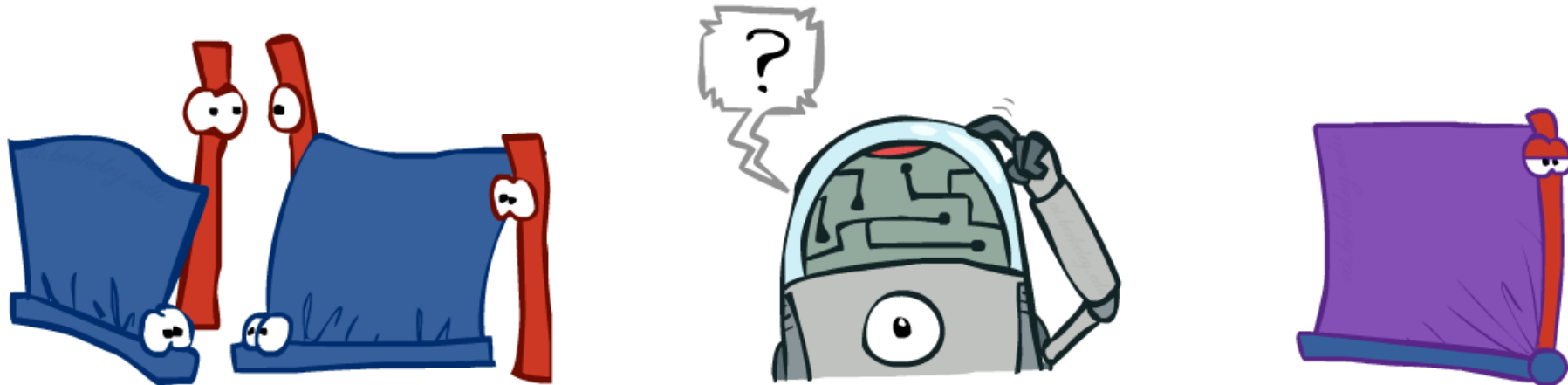
Probabilistic Graphical Models

- Bayes Nets
- Factors
- Variable Elimination

Inference

Inference: calculating some useful quantity from a probability model (joint probability distribution)

- Examples:
 - Posterior marginal probability
 - $P(Q|e_1, \dots, e_k)$
 - e.g., what disease might I have?
 - Most likely explanation:
 - $\operatorname{argmax}_{q,r,s} P(Q=q, R=r, S=s|e_1, \dots, e_k)$
 - e.g., what was just said?



Inference Overview

Given random variables Q, H, E (query, hidden, evidence)

We know how to do inference on a joint distribution

$$\begin{aligned} P(q|e) &= \alpha P(q, e) \\ &= \alpha \sum_{h \in \{h_1, h_2\}} P(q, h, e) \end{aligned}$$

We know Bayes nets can break down joint in to CPT factors

$$\begin{aligned} P(q|e) &= \alpha \sum_{h \in \{h_1, h_2\}} P(h) P(q|h) P(e|q) \\ &= \alpha [P(h_1) P(q|h_1) P(e|q) + P(h_2) P(q|h_2) P(e|q)] \end{aligned}$$



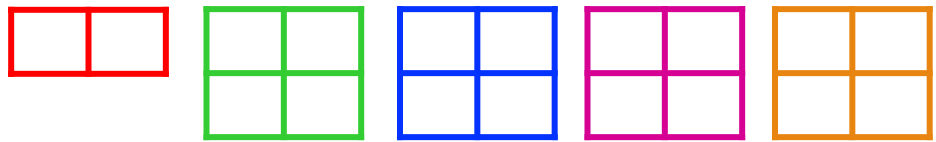
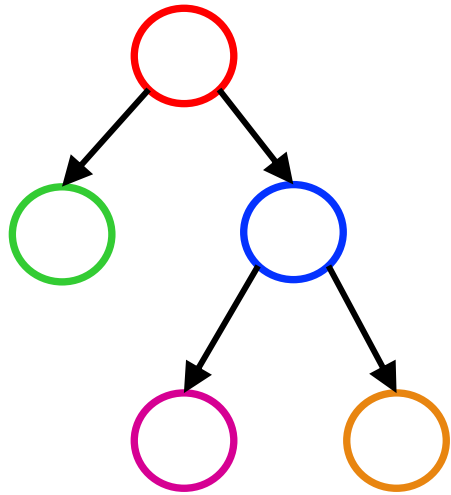
But we can be more efficient

$$\begin{aligned} P(q|e) &= \alpha P(e|q) \sum_{h \in \{h_1, h_2\}} P(h) P(q|h) \\ &= \alpha P(e|q) [P(h_1) P(q|h_1) + P(h_2) P(q|h_2)] \\ &= \alpha P(e|q) P(q) \end{aligned}$$

Now just extend to larger Bayes nets and a variety of queries

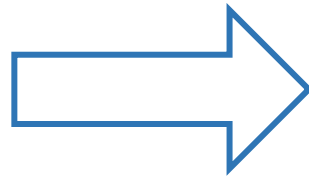
Answer Any Query from Bayes Net

Bayes Net



$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

Joint



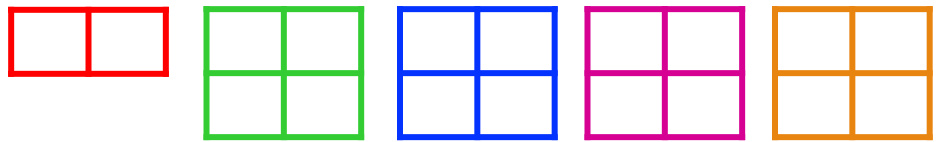
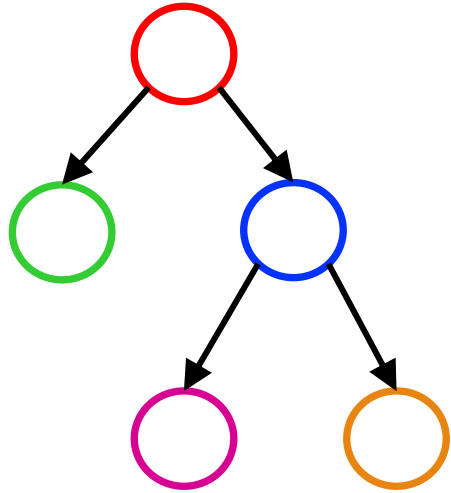


Query

$P(a | e)$

Next: Answer Any Query from Bayes Net

Bayes Net



$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

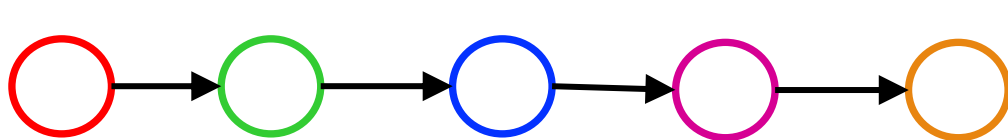
Query



$P(a | e)$

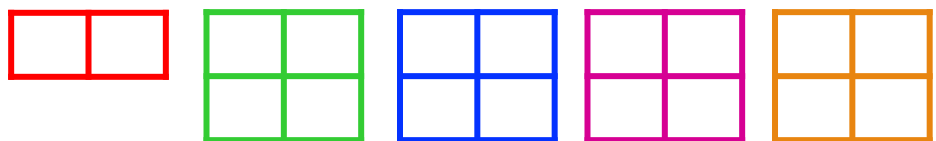
Markov Chain

Bayes Net



Query

$$P(x_3)$$

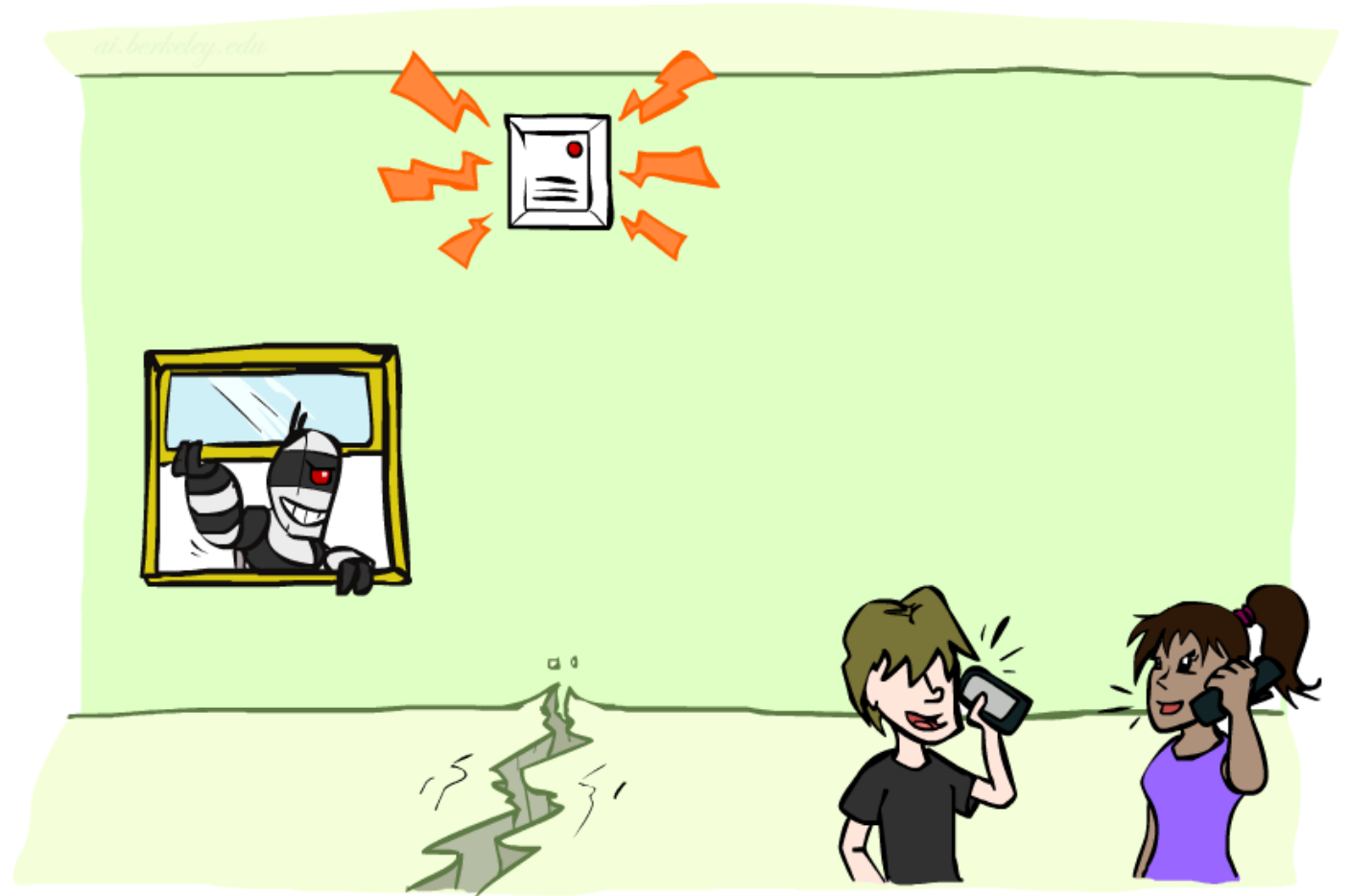


$$P(X_1) \quad P(X_2|X_1) \quad P(X_3|X_2) \quad P(X_4|X_3) \quad P(X_5|X_4)$$

Example: Alarm Network

Variables

- B: Burglary
- A: Alarm goes off
- M: Mary calls
- J: John calls
- E: Earthquake!



Example: Alarm Network



Joint distribution factorization example

Generic chain rule

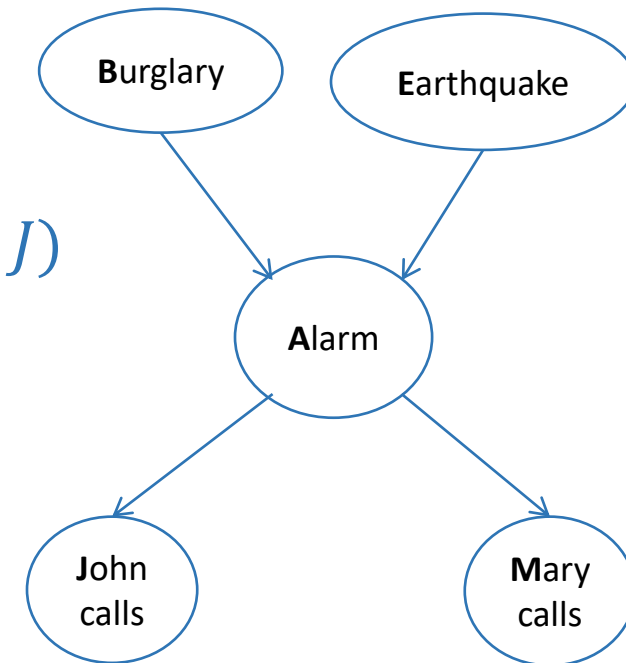
- $P(X_1 \dots X_n) = \prod_i P(X_i | X_1 \dots X_{i-1})$

$$P(B, E, A, J, M) = P(B) P(E|B) P(A|B, E) P(J|B, E, A) P(M|B, E, A, J)$$

$$P(B, E, A, J, M) = P(B) P(E) P(A|B, E) P(J|A) P(M|A)$$

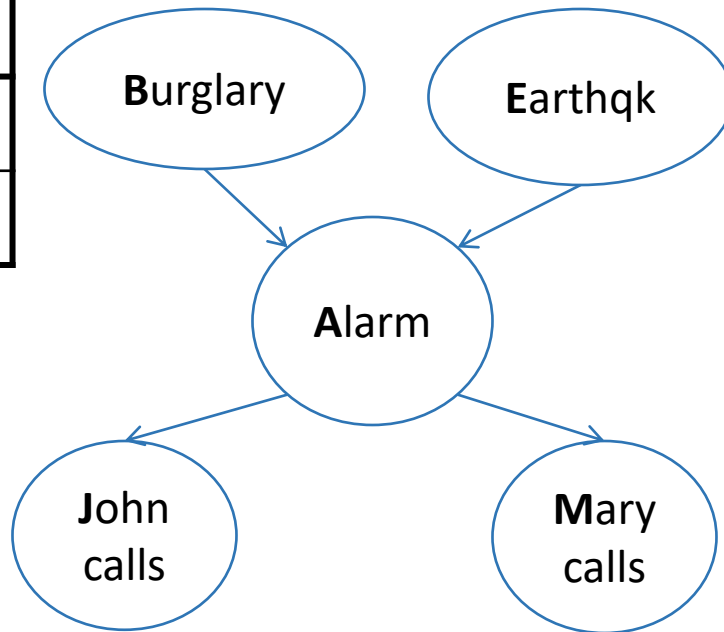
Bayes nets

- $P(X_1 \dots X_n) = \prod_i P(X_i | \text{Parents}(X_i))$



Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999

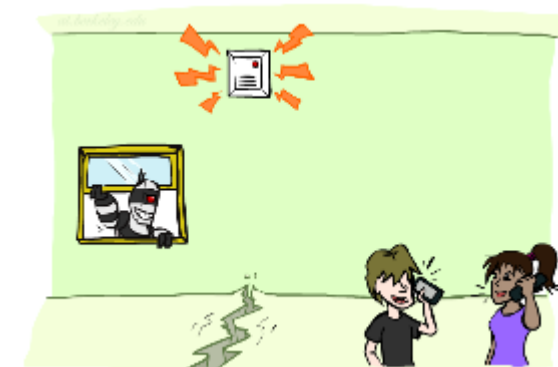


E	P(E)
+e	0.002
-e	0.998

A	J	P(J A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

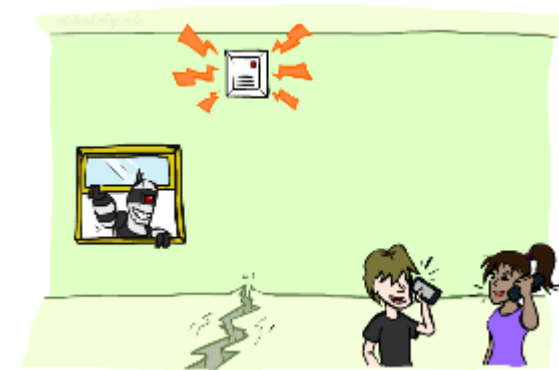
B	E	A	P(A B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999



Example: Alarm Network

$$\begin{aligned}P(+b, -e, -a, -j, -m) &= P(+b) * P(-e) * P(-a|+b, -e) * P(-j|-a) * P(-m|-a) \\&= 0.001 * 0.998 * 0.06 * 0.95 * 0.99\end{aligned}$$

$$\begin{aligned}P(+b, -e, -a, -j, -m) &= P(-e) * P(-a|+b, -e) * P(+b) * P(-j|-a) * P(-m|-a) \\&= 0.998 * 0.06 * 0.001 * 0.95 * 0.99\end{aligned}$$



Example: Alarm Network

$$\begin{aligned} P(+b, -e, -a, -j, -m) &= P(+b) * P(-e) * P(-a|+b, -e) * P(-j|-a) * P(-m|-a) \\ &= 0.001 * 0.998 * 0.06 * 0.95 * 0.99 \end{aligned}$$

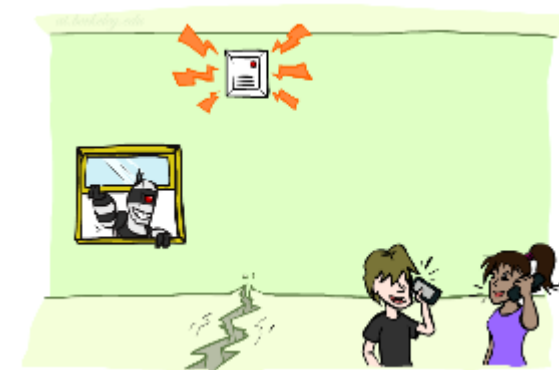
$$\begin{aligned} P(+b, -e, -a, -j, -m) &= P(-e) * P(-a|+b, -e) * P(+b) * P(-j|-a) * P(-m|-a) \\ &= 0.998 * 0.06 * 0.001 * 0.95 * 0.99 \end{aligned}$$

$$\begin{aligned} P(+b, -e, -a, -j, -m) &= P(-e) * P(-a|+b, -e) * P(+b) P(-j|-a) * P(-m|-a) \\ &= 0.998 * 0.06 * 0.0095 * 0.99 \end{aligned}$$

- Multiplication order can change (commutativity)
- Multiplication pairs don't have to make sense (associativity)

Factors

$$\begin{aligned} P(+b, -e, -a, -j, -m) &= P(+b) * P(-e) * P(-a|+b, -e) * P(-j|-a) * P(-m|-a) \\ &= 0.001 * 0.998 * 0.06 * 0.95 * 0.99 \end{aligned}$$



Factors

$$\begin{aligned} P(+b, -e, -a, -j, -m) &= f_1(+b) * f_2(-e) * f_3(-a, +b, -e) * f_4(-j, -a) * f_5(-m, -a) \\ &= 0.001 * 0.998 * 0.06 * 0.95 * 0.99 \end{aligned}$$

$$\begin{aligned} P(+b, -e, -a, -j, -m) &= f_2(-e) * f_3(-a, +b, -e) * f_1(+b) * f_4(-j, -a) * f_5(-m, -a) \\ &= 0.998 * 0.06 * 0.001 * 0.95 * 0.99 \end{aligned}$$

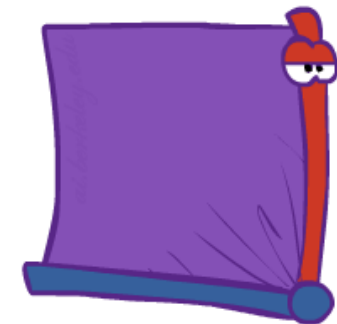
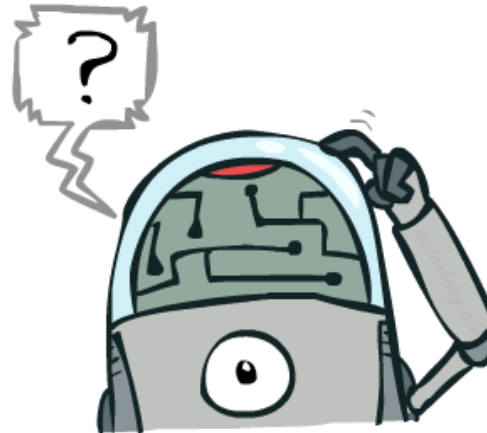
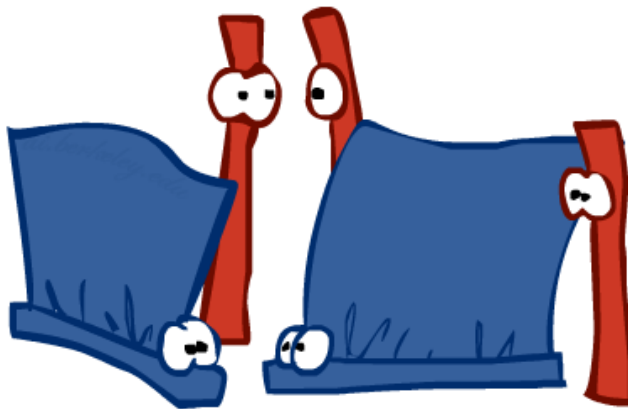
$$\begin{aligned} P(+b, -e, -a, -j, -m) &= f_2(-e) * f_3(-a, +b, -e) * f_6(+b, j, -a) * f_5(-m, -a) \\ &= 0.998 * 0.06 * 0.0095 * 0.99 \end{aligned}$$

- Multiplication order can change (commutativity)
- Multiplication pairs don't have to make sense (associativity)

Factor Tables

$$\begin{aligned} P(+b, -e, -a, -j, -m) &= P(+b) * P(-e) * P(-a|+b, -e) * P(-j|-a) * P(-m|-a) \\ &= 0.001 * 0.998 * 0.06 * 0.95 * 0.99 \end{aligned}$$

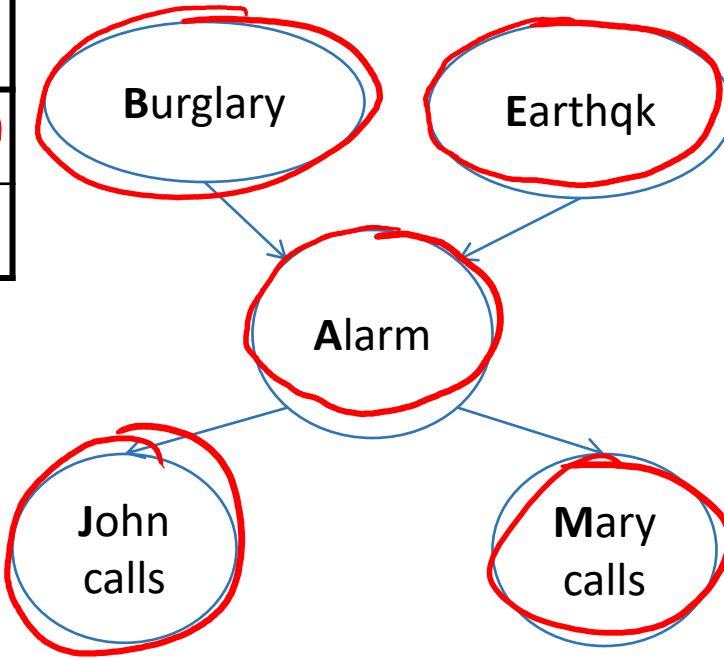
$$P(B, E, A, J, M) = P(B) * P(E) * P(A|B, E) * P(J|A) * P(M|A)$$



Example: Alarm Network

$$P(+b, -e, -a, -m, -j) =$$

B	P(B)
+b	0.001
-b	0.999

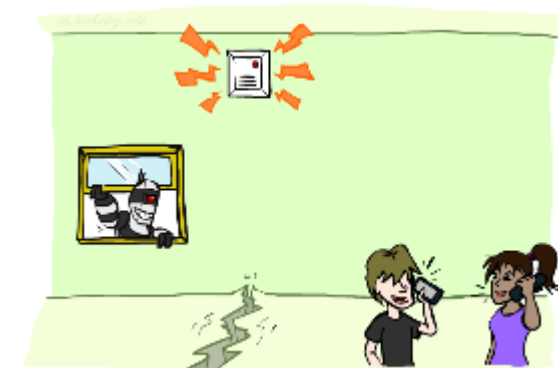


E	P(E)
+e	0.002
-e	0.998

A	J	P(J A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
<u>+b</u>	<u>-e</u>	<u>-a</u>	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999



Inference by Enumeration in Bayes Net

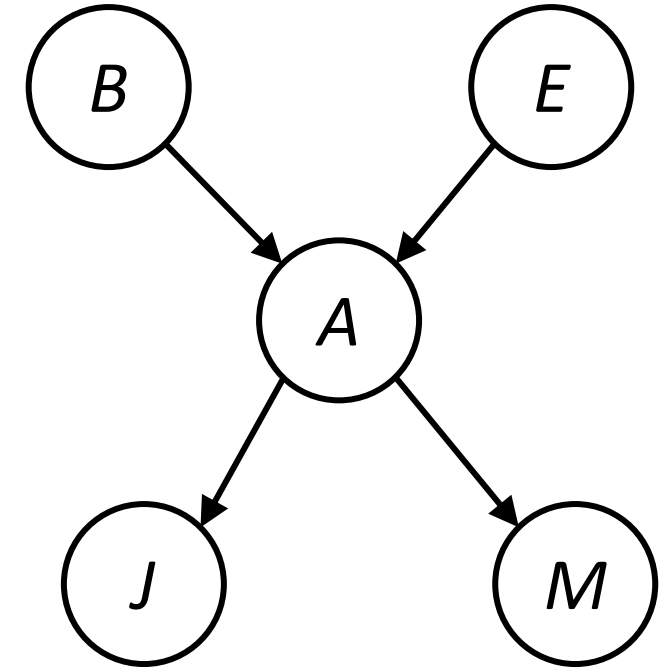
Reminder of inference by enumeration:

- Any probability of interest can be computed by summing entries from the joint distribution
- Entries from the joint distribution can be obtained from a BN by multiplying the corresponding conditional probabilities

$$\begin{aligned} P(B \mid j, m) &= \alpha P(B, j, m) \\ &= \alpha \sum_{e,a} P(B, e, a, j, m) \\ &= \alpha \sum_{e,a} P(B) P(e) P(a \mid B, e) P(j \mid a) P(m \mid a) \end{aligned}$$

So inference in Bayes nets means computing sums of products of numbers: sounds easy!!

Problem: sums of **exponentially many** products!



Can we do better?

Consider

- $x_1y_1z_1 + x_1y_1z_2 + x_1y_2z_1 + x_1y_2z_2 + x_2y_1z_1 + x_2y_1z_2 + x_2y_2z_1 + x_2y_2z_2$
- 16 multiplies, 7 adds
- Lots of repeated subexpressions!

Rewrite as

- $(x_1 + x_2)(y_1 + y_2)(z_1 + z_2)$
- 2 multiplies, 3 adds

$$\sum_e \sum_a P(B) P(e) P(a | B, e) P(j | a) P(m | a)$$

$$\begin{aligned} &= P(B) P(+e) P(+a | B, +e) P(j | +a) P(m | +a) \\ &\quad + P(B) P(-e) P(+a | B, -e) P(j | +a) P(m | +a) \\ &\quad + P(B) P(+e) P(-a | B, +e) P(j | -a) P(m | -a) \\ &\quad + P(B) P(-e) P(-a | B, -e) P(j | -a) P(m | -a) \end{aligned}$$

- Lots of repeated subexpressions!

Variable elimination: The basic ideas

Move summations inwards as far as possible

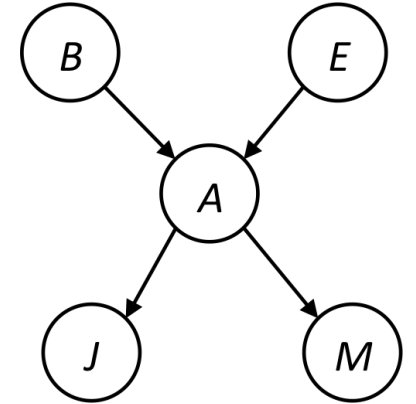
$$\begin{aligned} P(B \mid j, m) &= \alpha \sum_e \sum_a P(B, e, a, j, m) \\ &= \alpha \sum_e \sum_a P(j|a) P(e) P(m|a) P(a|B, e) P(B) \\ &= \alpha \sum_e \sum_a P(j|a) P(e) P(m|a) P(a|B, e) P(B) \end{aligned}$$

Variable elimination: The basic ideas

Move summations inwards as far as possible

$$\begin{aligned} P(B \mid j, m) &= \alpha \sum_e \sum_a P(B, e, a, j, m) \\ &= \alpha \sum_e \sum_a P(j|a) P(e) P(m|a) P(a|B, e) P(B) \\ &= \alpha P(B) \sum_e P(e) \sum_a P(j|a) P(m|a) P(a|B, e) \end{aligned}$$

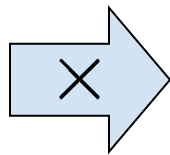
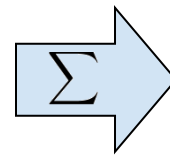
Example



Query $P(B \mid j, m)$

$P(B)$	$P(E)$	$P(A \mid B, E)$	$P(j \mid A)$	$P(m \mid A)$
--------	--------	------------------	---------------	---------------

Choose A

 $P(A \mid B, E)$ $P(j \mid A)$ $P(m \mid A)$  $P(A, j, m \mid B, E)$  $P(j, m \mid B, E)$

$P(B)$	$P(E)$	$P(j, m \mid B, E)$
--------	--------	---------------------

Example

$P(B)$	$P(E)$	$P(j,m B,E)$
--------	--------	--------------

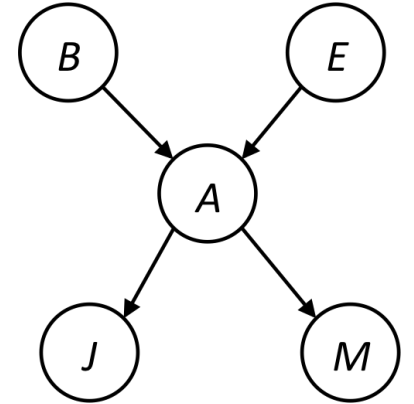
Choose E

$$\begin{array}{c} P(E) \\ P(j,m|B,E) \end{array} \xrightarrow{\times} P(E,j,m|B) \xrightarrow{\Sigma} P(j,m|B)$$

$P(B)$	$P(j,m B)$
--------	------------

Finish with B

$$\begin{array}{c} P(B) \\ P(j,m|B) \end{array} \xrightarrow{\times} P(j,m,B) \xrightarrow{\text{Normalize}} P(B | j,m)$$



Order matters

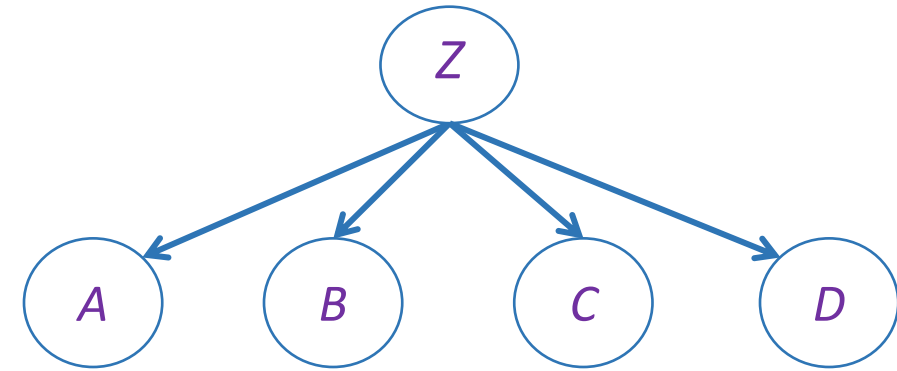
- Elimination Order: C, B, A, Z

- $P(D) = \alpha \sum_{z,a,b,c} P(D|z) P(z) P(a|z) P(b|z) P(c|z)$
- $= \alpha \sum_z P(D|z) P(z) \sum_a P(a|z) \sum_b P(b|z) \sum_c P(c|z)$
- Largest factor has 2 variables (D,Z)

- Elimination Order: Z, C, B, A

- $P(D) = \alpha \sum_{a,b,c,z} P(a|z) P(b|z) P(c|z) P(D|z) P(z)$
- $= \alpha \sum_a \sum_b \sum_c \sum_z P(a|z) P(b|z) P(c|z) P(D|z) P(z)$
- Largest factor has 4 variables (A,B,C,D) (or 5 if you count pre-summation over Z)

- In general, with n leaves, factor of size 2^n



VE: Computational and Space Complexity

The computational and space complexity of variable elimination is determined by the largest factor (and it's space that kills you)

The elimination ordering can greatly affect the size of the largest factor.

- E.g., previous slide's example 2^n vs. 2

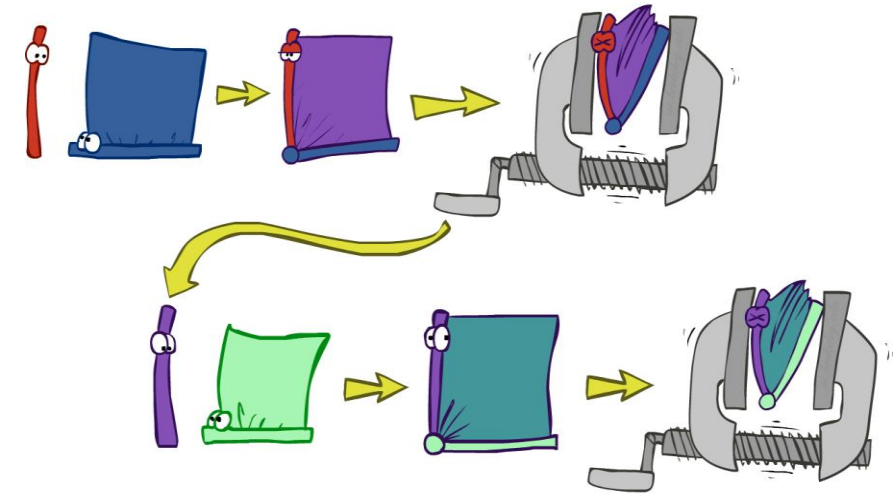
Does there always exist an ordering that only results in small factors?

- No!

Variable elimination: The basic ideas

Move summations inwards as far as possible

- $$P(B \mid j, m) = \alpha \sum_e \sum_a P(B) P(e) P(a \mid B, e) P(j \mid a) P(m \mid a)$$
$$= \alpha P(B) \sum_e P(e) \sum_a P(a \mid B, e) P(j \mid a) P(m \mid a)$$



Do the calculation from the inside out

- I.e., sum over a first, then sum over e
- Problem: $P(a \mid B, e)$ isn't a single number, it's a bunch of different numbers depending on the values of B and e
- Solution: use arrays of numbers (of various dimensions) with appropriate operations on them; these are called **factors**

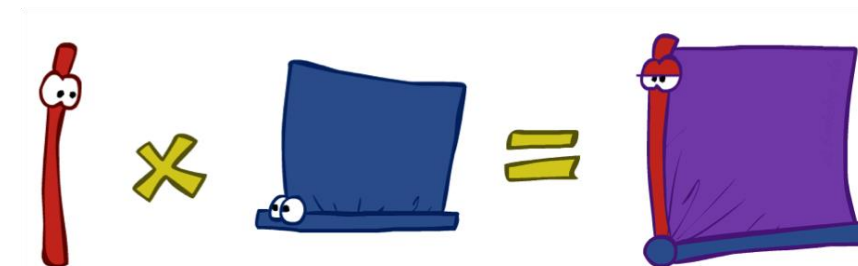
Operation 1: Pointwise product

First basic operation: **pointwise product** of factors
(similar to a **database join**, **not** matrix multiply!)

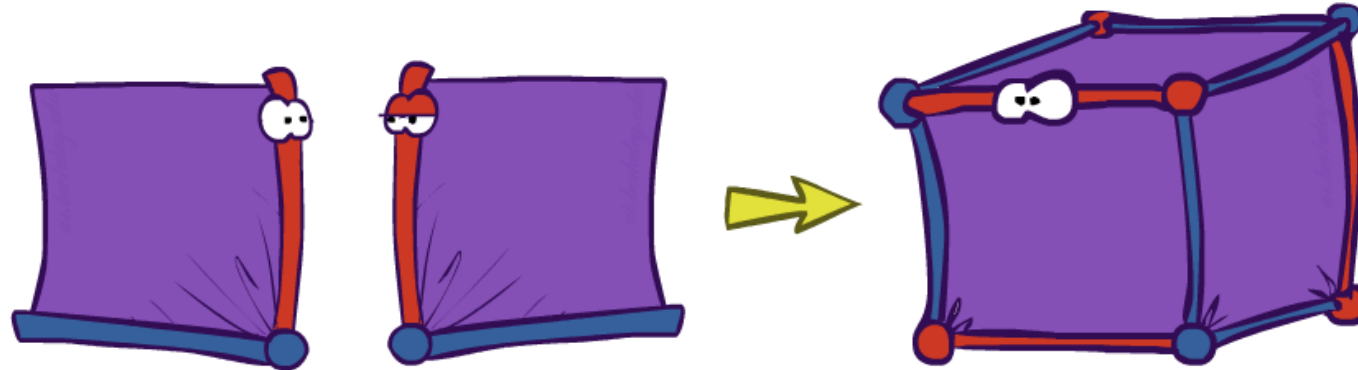
- New factor has **union** of variables of the two original factors
- Each entry is the product of the corresponding entries from the original factors

Example: $P(A) \times P(J|A) = P(A,J)$

$P(A)$		$P(J A)$			$P(A,J)$		
true	0.1	A \ J	true	false	A \ J	true	false
false	0.9	true	0.9	0.1	true	0.09	0.01
		false	0.05	0.95	false	0.045	0.855



Example: Making larger factors



Example: $P(J/A) \times P(M/A) = P(J,M/A)$

$P(J/A)$

$A \setminus J$	true	false
true	0.99	0.01
false	0.145	0.855

$\times$

$P(M/A)$

$A \setminus M$	true	false
true	0.97	0.03
false	0.019	0.891

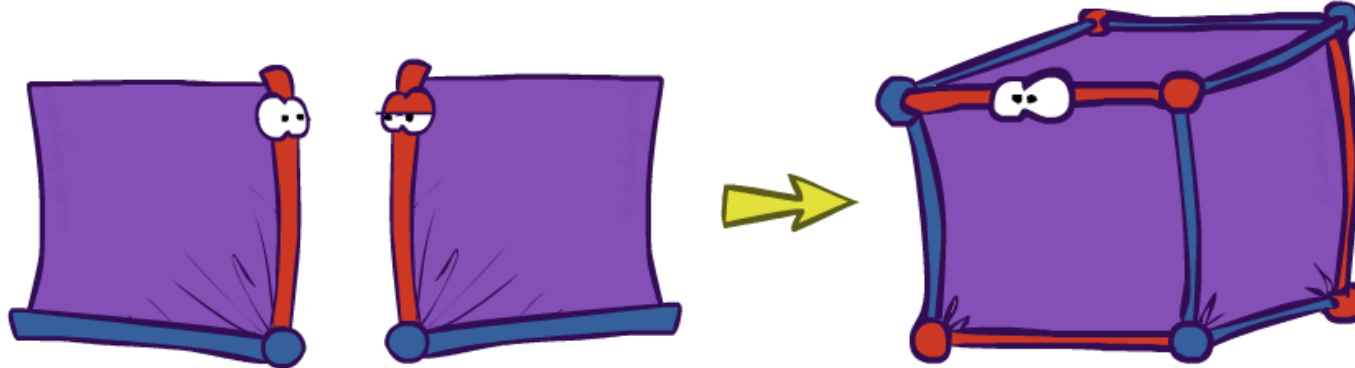
$=$

$P(J,M/A)$

	$J \setminus M$		
	true	false	
$J \setminus M$	true	false	
true			18
false		.0003	

A=false
A=true

Example: Making larger factors



Example: $f_1(U,V) \times f_2(V,W) \times f_3(W,X) = f_4(U,V,W,X)$

Sizes: $[10,10] \times [10,10] \times [10,10] = [10,10,10,10]$

I.e., 300 numbers blows up to 10,000 numbers!

Factor blowup can make VE very expensive

Operation 2: Summing out a variable

Second basic operation: **summing out** (or eliminating) a variable from a factor

- Shrinks a factor to a smaller one

Example: $\sum_j P(A, J) = P(A, j) + P(A, \neg j) = P(A)$

$P(A, J)$

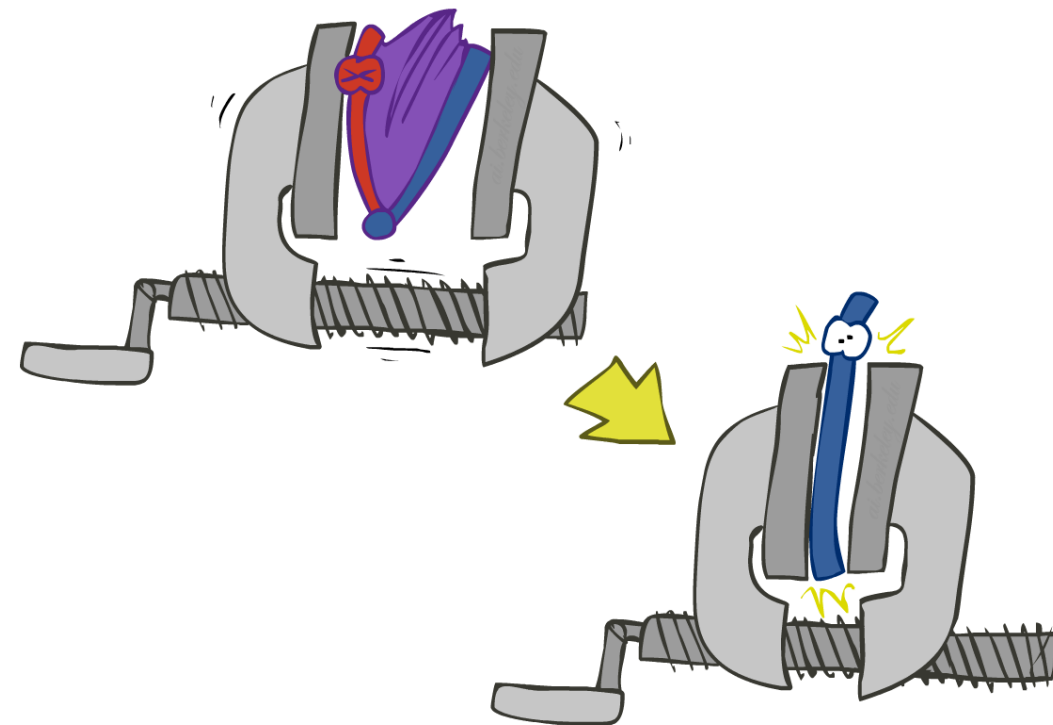
$A \setminus J$	true	false
true	0.09	0.01
false	0.045	0.855

Sum out J



$P(A)$

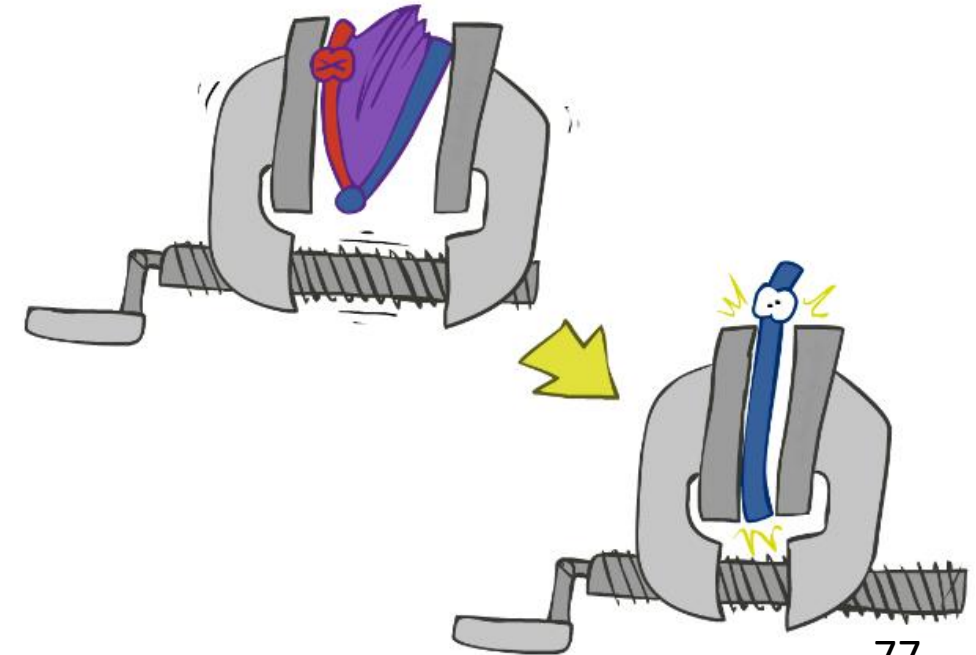
	true	false
true	0.1	
false	0.9	



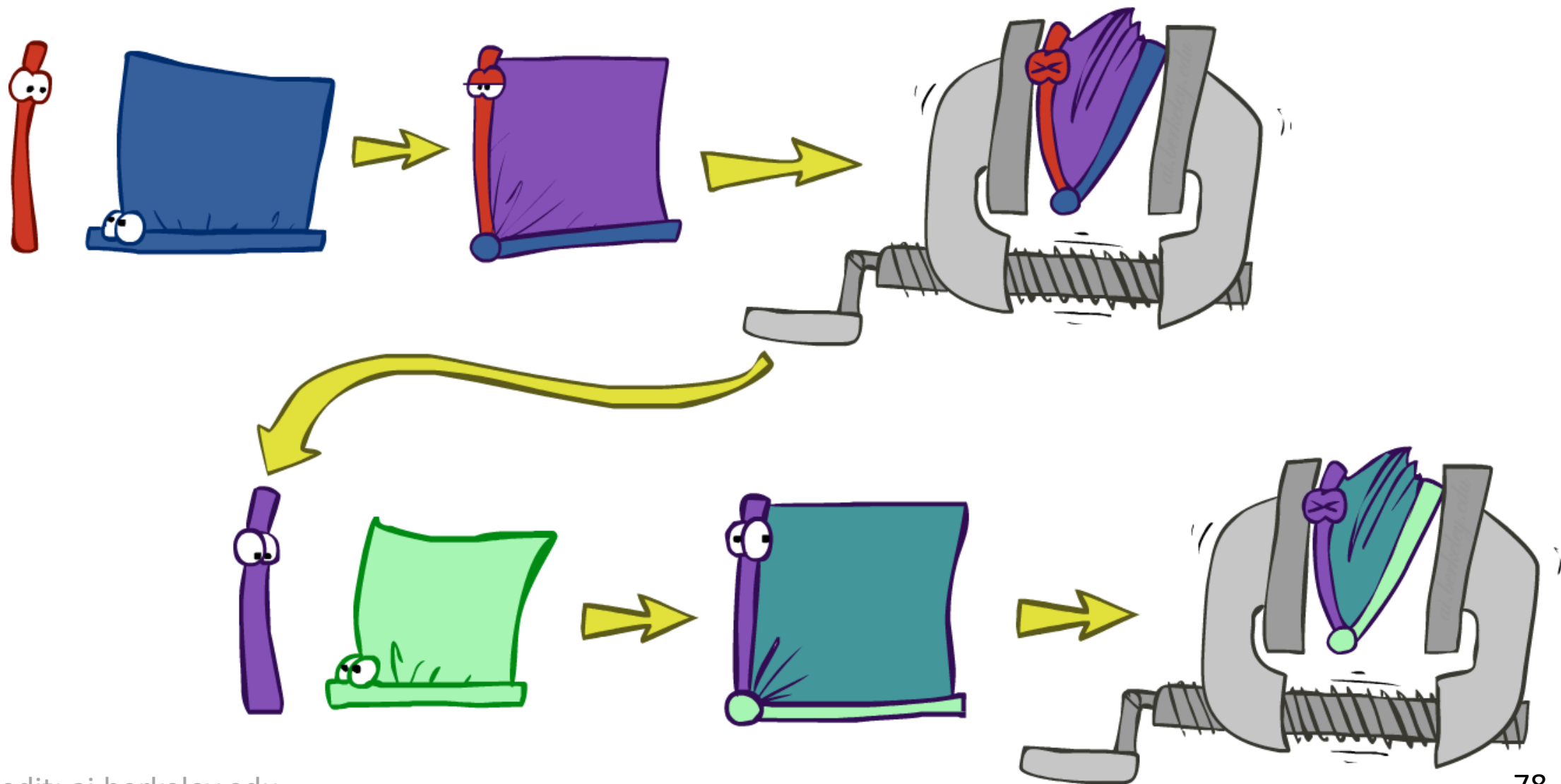
Summing out from a product of factors

Project the factors each way first, then sum the products

- Example: $\sum_a P(a|B,e) P(j|a) P(m|a)$
= $P(a|B,e) P(j|a) P(m|a) + P(\neg a|B,e) P(j|\neg a) P(m|\neg a)$
= $P(a,j,m|B,e) + P(\neg a,j,m|B,e)$
= $P(j,m|B,e)$



Variable Elimination



Variable Elimination

- Query: $P(Q|E_1=e_1, \dots, E_k=e_k)$

Start with initial factors:

- Local CPTs (but instantiated by evidence)

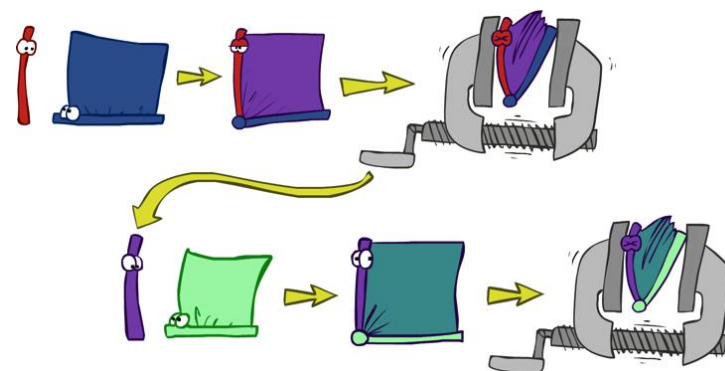


x	P(x)
-3	0.05
-1	0.25
0	0.07
1	0.2
5	0.01

2 0.15

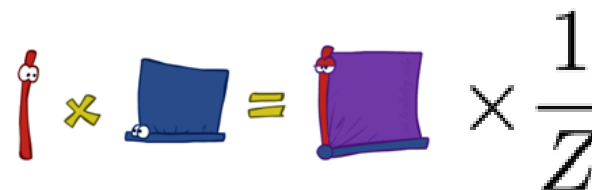
While there are still hidden variables (not Q or evidence):

- Pick a hidden variable H
- Join all factors mentioning H
- Eliminate (sum out) H



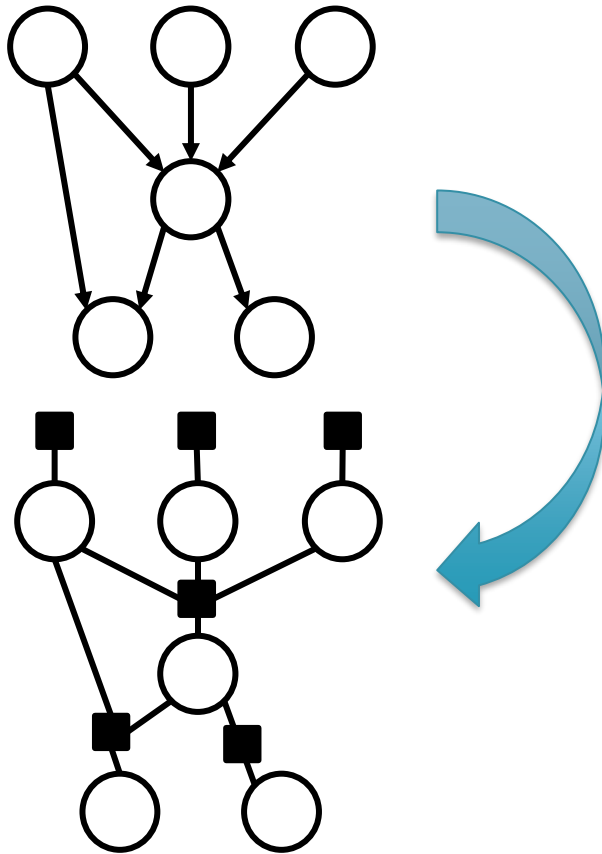
Join all remaining factors and normalize

Slide credit: ai.berkeley.edu


$$\text{stick figure} \times \text{blue pillow} = \text{purple pillow} \times \frac{1}{Z}$$

Converting to Factor Graphs

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



Each clique in an **undirected GM** becomes a factor

