Directed Graphical Probabilistic Models: inference

William W. Cohen

Machine Learning 10-601

Feb 2008

REVIEW OF DIRECTED GRAPHICAL MODELS AND D-SEPARATION

Summary of Monday(1): Bayes nets

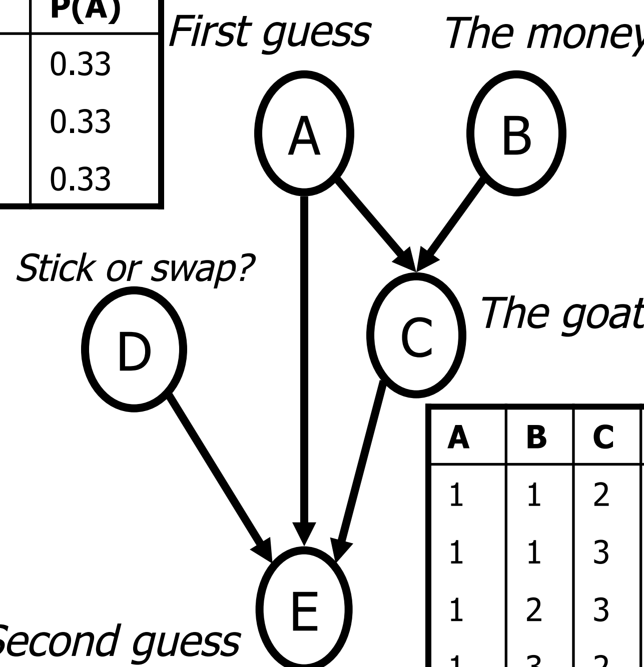
- Many problems can be solved using the <u>joint probability</u> P(X_1, …,X_n).
- Bayes nets describe a way to compactly write the joint.
- For a Bayes net:

$$P(X_1,\ldots,X_n) = \prod_i P(X_i \mid X_1,\ldots,X_{i-1})$$
$$= \prod_i P(X_i \mid parents(X_i))$$

- Conditional independence:

$$X \perp Y \mid E \;\equiv\; I\langle X,E,Y\rangle$$
$$\equiv P(X \mid E,Y) = P(X \mid E)$$

A	P(A)
1	0.33
2	0.33
3	0.33



B	P(B)
1	0.33
2	0.33
3	0.33

A	B	C	P(C|A,B)
1	1	2	0.5
1	1	3	0.5
1	2	3	1.0
1	3	2	1.0
…	…	…	…

A	C	D	P(E|A,C,D)
…	…	…	…

$$P(A,B,C,D,E) = P(A)P(B \mid \cancel{A})P(C \mid A,B)P(D \mid \cancel{A,B,C})P(E \mid A,\cancel{B},C,D)$$
$$= P(A)P(B)P(C \mid A,B)P(D)P(E \mid A,C,D)$$

Conditional Independence

"It's really a lot like independence"

Independence: $P(X,Y) = P(X)P(Y)$

Conditional independence: $\underline{P(X \mid Y, E)} = \underline{P(X \mid E)}$

Claim: if I<X,E,Y> then $P(X,Y \mid E) = P(X \mid E)P(Y \mid E)$

$P(X,Y \mid E) = \underline{P(X \mid Y, E)}P(Y \mid E)$ (Fancy version of c.r.) $P(X,Y) = P(X \mid Y)P(Y)$

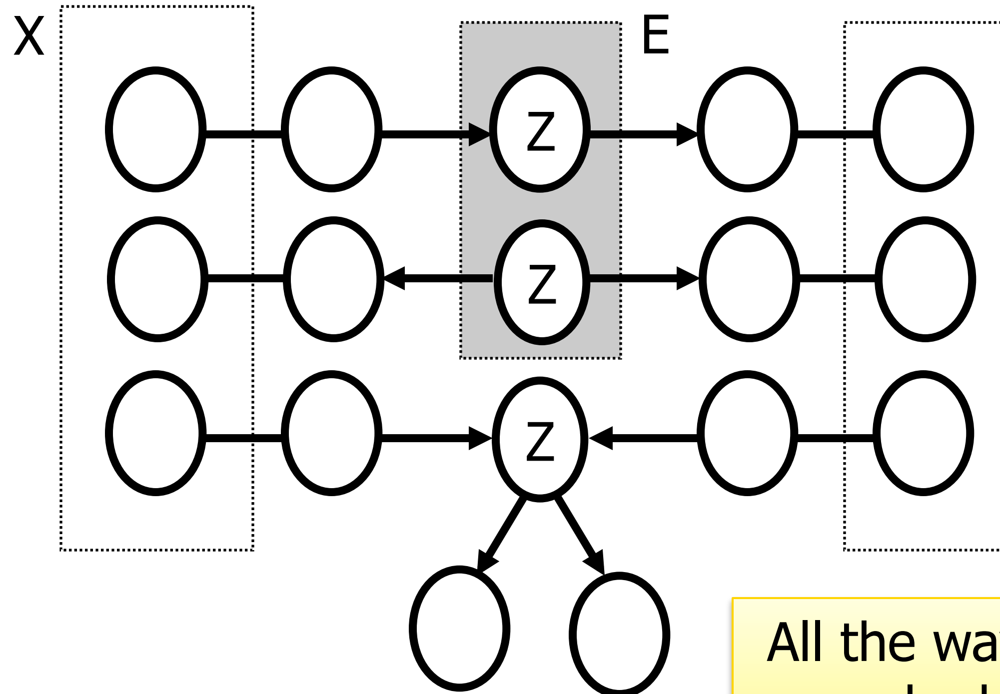
$P(X,Y \mid E) = \underline{P(X \mid E)}P(Y \mid E)$ (Def' n of cond. Indep.)

Summary of Monday(2): *d*-separation

There are three ways paths from X to Y given evidence E can be *blocked.*

X is *d-separated* from Y given E iff all paths from X to Y given E are blocked

If X is d-separated from Y given E, then *I<X,E,Y>*



All the ways paths can be blocked

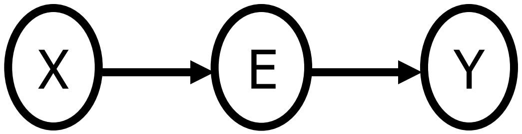
$$I\langle X,E,Y\rangle \equiv X \perp Y \mid E$$

$$\equiv P(X \mid E,Y) = P(X \mid E)$$

$$\equiv \forall x,y,e : P(X=x \mid E=e, Y=y) = P(X=x \mid E=e)$$

d-separation continued…

Question: is $Y \perp X \mid \{\}$?



$$P(Y \mid X) \stackrel{?}{=} P(Y)$$

It depends…on the CPTs

This is why d-separation implies conditional independence but not the converse…

X	P(X)
0	0.5
1	0.5

X	E	P(E|X)
0	0	0.01
0	1	0.99
1	0	0.99
1	1	0.01

$E \cong \neg X$

E	Y	P(Y|E)
0	0	0.5
0	1	0.5
1	0	0.5
1	1	0.5

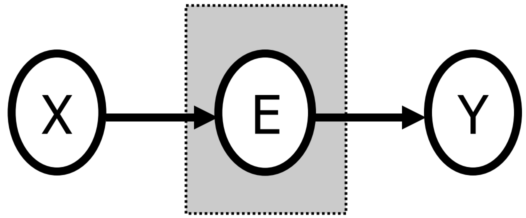
E	Y	P(Y|E)
0	0	0.01
0	1	0.99
1	0	0.99
1	1	0.01

$Y \cong \neg E$

$$P(X,E,Y) = P(X)P(E \mid X)P(Y \mid E,X)$$
$$= P(X)P(E \mid X)P(Y \mid E)$$

d-separation continued...

Question: is $Y \perp X \mid E$?



Yes!

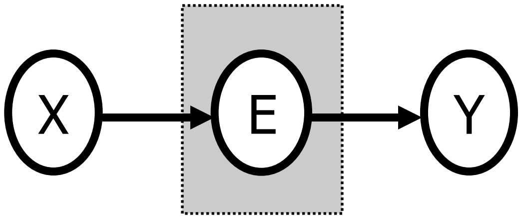
$$P(Y \mid E, X) \stackrel{?}{=} P(Y \mid E)$$

$$P(X, E, Y) = P(X)P(E \mid X)P(Y \mid E, X)$$
$$= P(X)P(E \mid X)P(Y \mid E)$$

$$P(X, E, Y) = P(X)P(E \mid X)P(Y \mid E, X)$$
$$= P(X)P(E \mid X)P(Y \mid E)$$

d-separation continued...

Question: is $X \perp Y \mid E$?



Yes!

$$P(X \mid E, Y) \stackrel{?}{=} P(X \mid E)$$

$$P(X \mid Y) = \frac{P(Y \mid X) P(X)}{P(Y)}$$ Bayes rule

$$P(X \mid Y, E) = \frac{P(Y \mid X, E) P(X \mid E)}{P(Y \mid E)}$$ Fancier version of B.R.

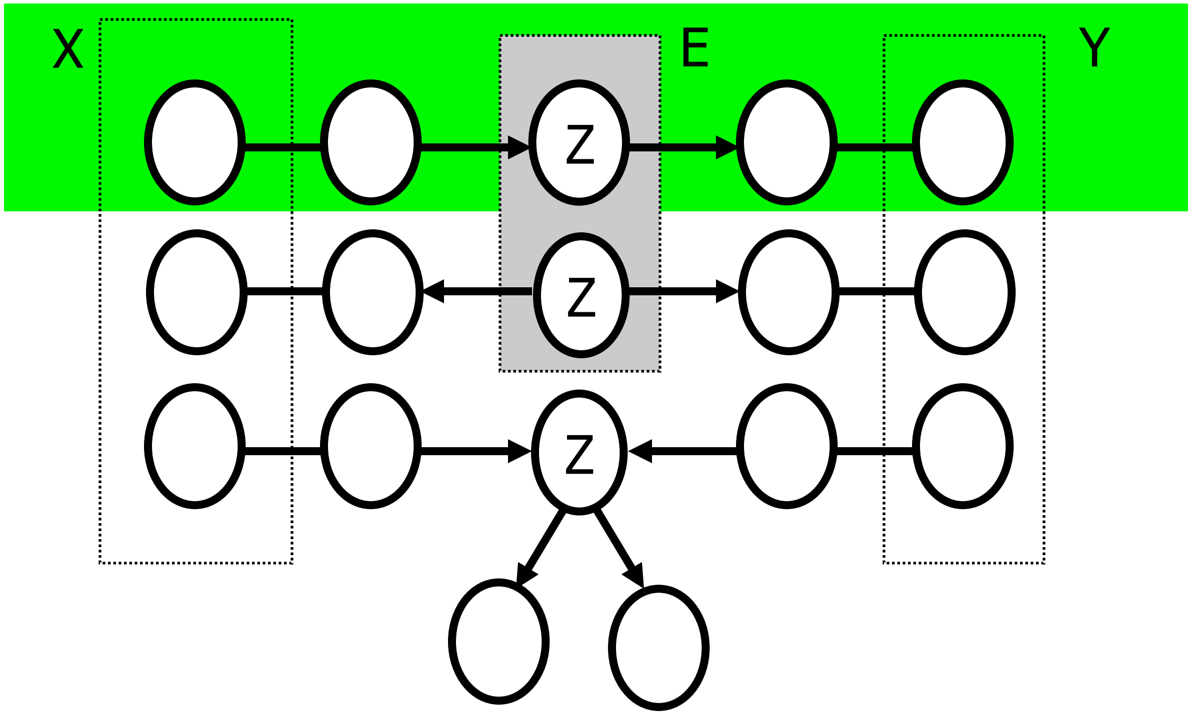
From previous slide...

$$P(X, E, Y) = P(X) P(E \mid X) P(Y \mid E, X)$$
$$= P(X) P(E \mid X) P(Y \mid E)$$

d-separation

$$X \perp Y \mid E$$

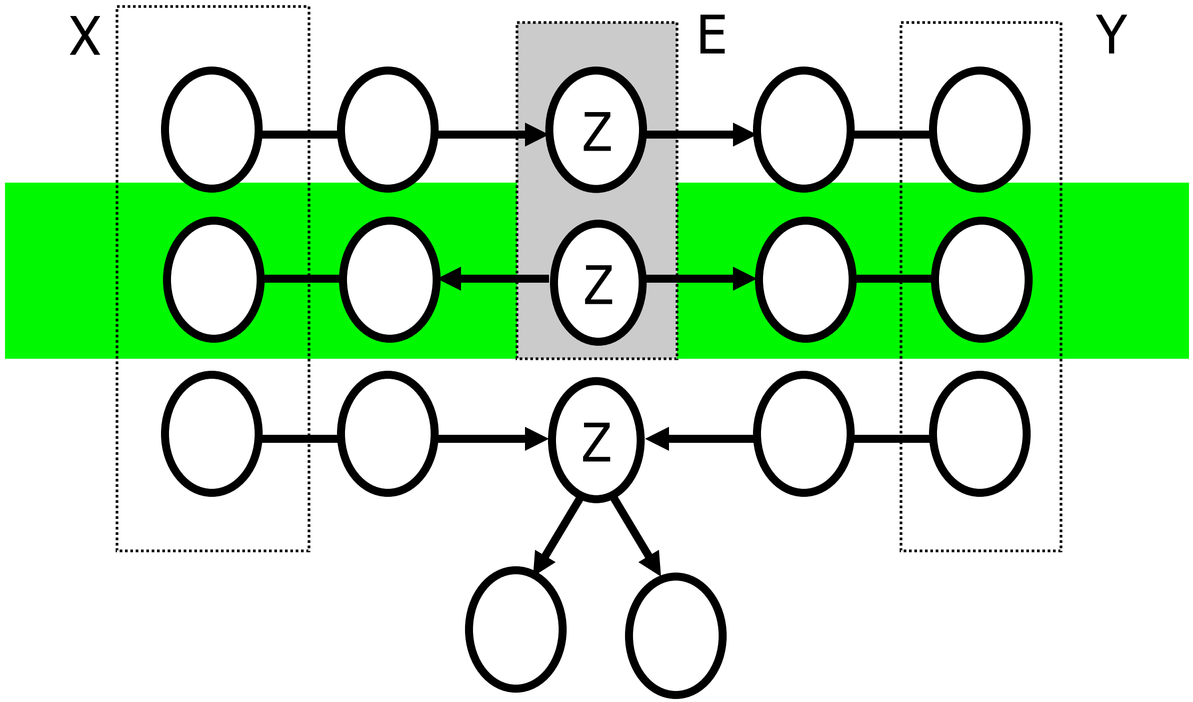
$$X \not\perp Y \mid \{\}$$



d-separation

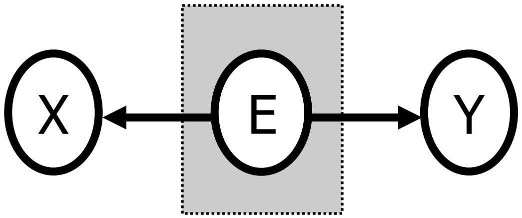
$X \perp Y \mid E$?

$X \perp Y \mid \{\}$?



d-separation continued…

Question: is $Y \perp X \mid E$?



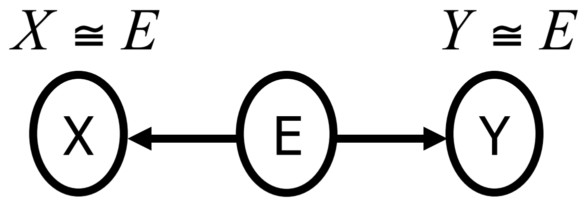
$$P(Y \mid E, X) \overset{?}{=} P(Y \mid E)$$

Yes!

$$P(E, X, Y) = P(E)P(X \mid E)P(Y \mid E, X)$$
$$= P(E)P(X \mid E)P(Y \mid E)$$

$$P(E, X, Y) = P(E)P(X \mid E)P(Y \mid E, X)$$
$$= P(E)P(X \mid E)P(Y \mid E)$$

d-separation continued…



E	P(E)
0	0.5
1	0.5

Question: is $Y \perp X \mid \{\}$?

$$P(Y \mid X) \stackrel{?}{=} P(Y)$$

No

$$P(Y = 1) \cong 0.5$$

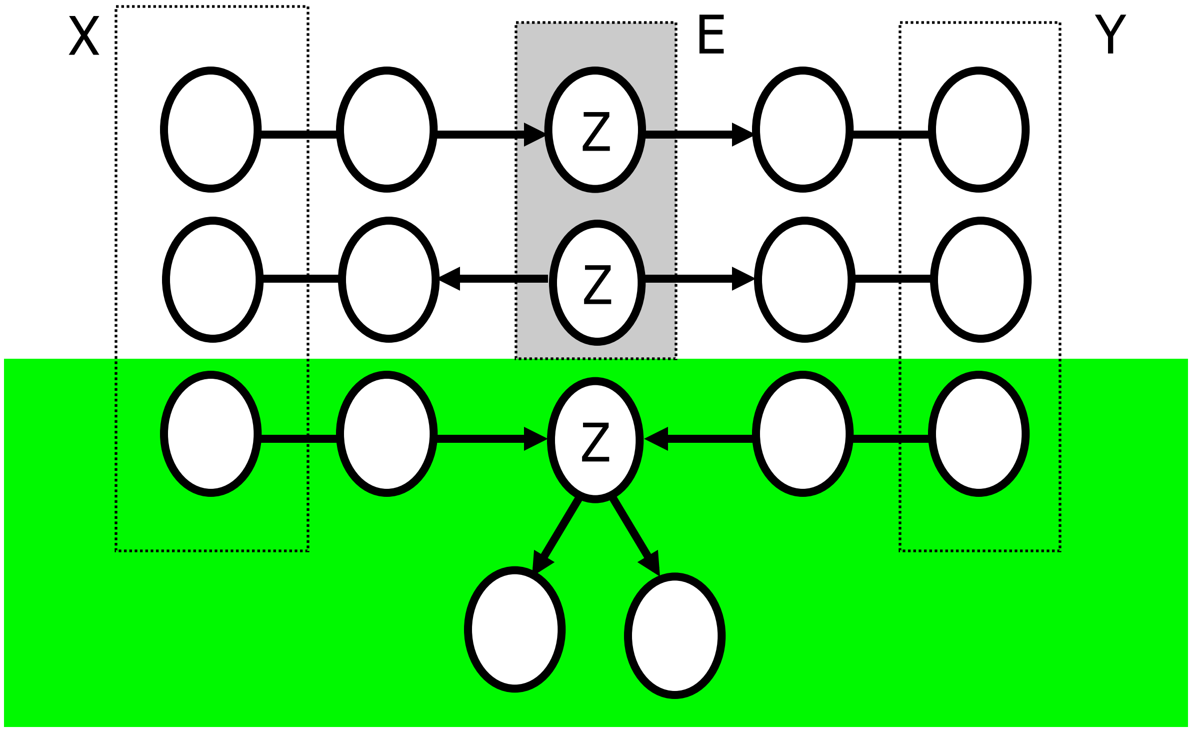
$$P(Y = 1 \mid X = 1) \cong 1$$

$$P(E, X, Y) = P(E)P(X \mid E)P(Y \mid E, X)$$
$$= P(E)P(X \mid E)P(Y \mid E)$$

d-separation

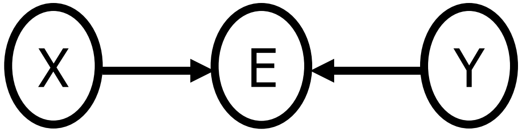
$X \perp Y \mid E$?

$X \perp Y \mid \{\}$?



d-separation continued...

Question: is $Y \perp X \mid \{\} \ ?$



Yes!

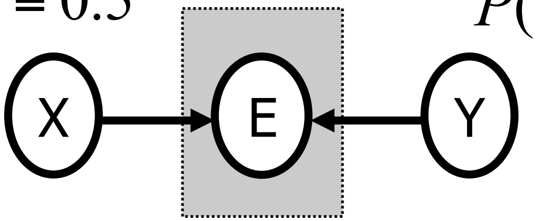
$$P(Y \mid X) \overset{?}{=} P(Y)$$

$$\begin{aligned} P(X,Y,E) &= P(X)P(Y \mid X)P(E \mid X,Y) \\ &= P(X)P(Y)P(E \mid X,Y) \end{aligned}$$

$$\begin{aligned} P(X,Y,E) &= P(X)P(Y \mid X)P(E \mid X,Y) \\ &= P(X)P(Y)P(E \mid X,Y) \end{aligned}$$

d-separation continued...

$P(X) = 0.5$ $P(Y) = 0.5$



Question: is $Y \perp X \mid E$?

$$P(Y \mid E, X) \stackrel{?}{=} P(Y \mid E)$$

$$P(Y=1 \mid E=1, X=0) = \frac{P(Y=1, E=1, X=0)}{P(E=1, X=0)} = \frac{0.24}{0.25}$$

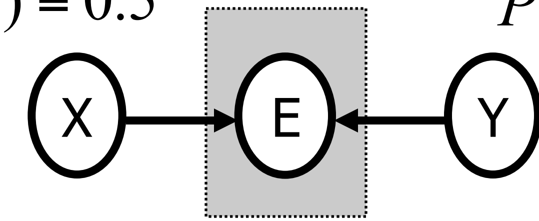
$E \cong X \vee Y$

X	Y	E	P(E|X,Y)	P(E,X,Y)
0	0	0	0.96	0.24
0	0	1	0.04	0.01
0	1	0	0.04	0.01
0	1	1	0.96	0.24
1	0	0	0.04	0.01
1	0	1	0.96	0.24
1	1	0	0.04	0.01
1	1	1	0.96	0.24

$$P(X, Y, E) = P(X)P(Y \mid X)P(E \mid X, Y)$$
$$= P(X)P(Y)P(E \mid X, Y)$$

d-separation continued…

$P(X) = 0.5$ $\quad$ $P(Y) = 0.5$



Question: is $Y \perp X \mid E$?

No!

$$P(Y \mid E, X) \stackrel{?}{=} P(Y \mid E)$$

$$P(Y = 1 \mid E = 1, X = 0) = \frac{P(Y = 1, E = 1, X = 0)}{P(E = 1, X = 0)} \cong 1$$

$$P(Y = 1 \mid E = 1) = \frac{P(Y = 1, E = 1)}{P(E = 1)} \cong 2/3$$

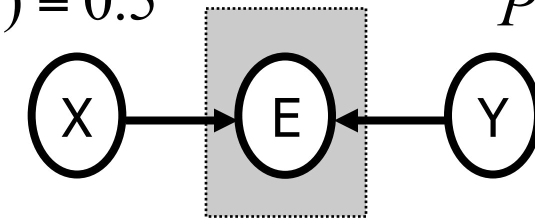
$E \cong X \vee Y$

X	Y	E	P(E|X,Y)	P(E,X,Y)
0	0	0	0.96	0.24
0	0	1	0.04	0.01
0	1	0	0.96	0.01
0	1	1	0.96	0.24
1	0	0	0.04	0.01
1	0	1	0.96	0.24
1	1	0	0.04	0.01
1	1	1	0.96	0.24

$$P(X, Y, E) = P(X)P(Y \mid X)P(E \mid X, Y)$$
$$= P(X)P(Y)P(E \mid X, Y)$$

d-separation continued…

$P(X) = 0.5$ $P(Y) = 0.5$



Question: is $Y \perp X \mid E$?

No!

$$P(Y \mid E, X) \stackrel{?}{=} P(Y \mid E)$$

$$P(Y=1 \mid E=1, X=1) = \frac{P(Y=1, E=1, X=1)}{P(E=1, X=1)} = 0.5$$

$$P(Y=1 \mid E=1) = \frac{P(Y=1, E=1)}{P(E=1)} \cong \frac{2}{3}$$

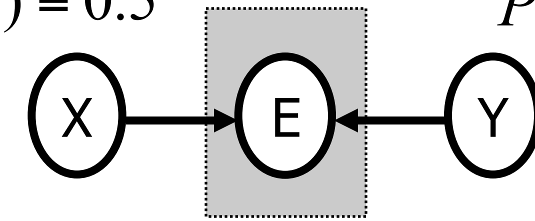
$E \cong X \vee Y$

X	Y	E	P(E|X,Y)	P(E,X,Y)
0	0	0	0.96	0.24
0	0	1	0.04	0.01
0	1	0	0.96	0.01
0	1	1	0.96	0.24
1	0	0	0.04	0.01
1	0	1	0.96	0.24
1	1	0	0.04	0.01
1	1	1	0.96	0.24

$$P(X,Y,E) = P(X)P(Y \mid X)P(E \mid X,Y)$$
$$= P(X)P(Y)P(E \mid X,Y)$$

"Explaining away"

$P(X) = 0.5$ $P(Y) = 0.5$



$E \cong X \vee Y$

$X \perp Y \mid E$ NO

$X \perp Y \mid \{\}$ YES

This is "explaining away":

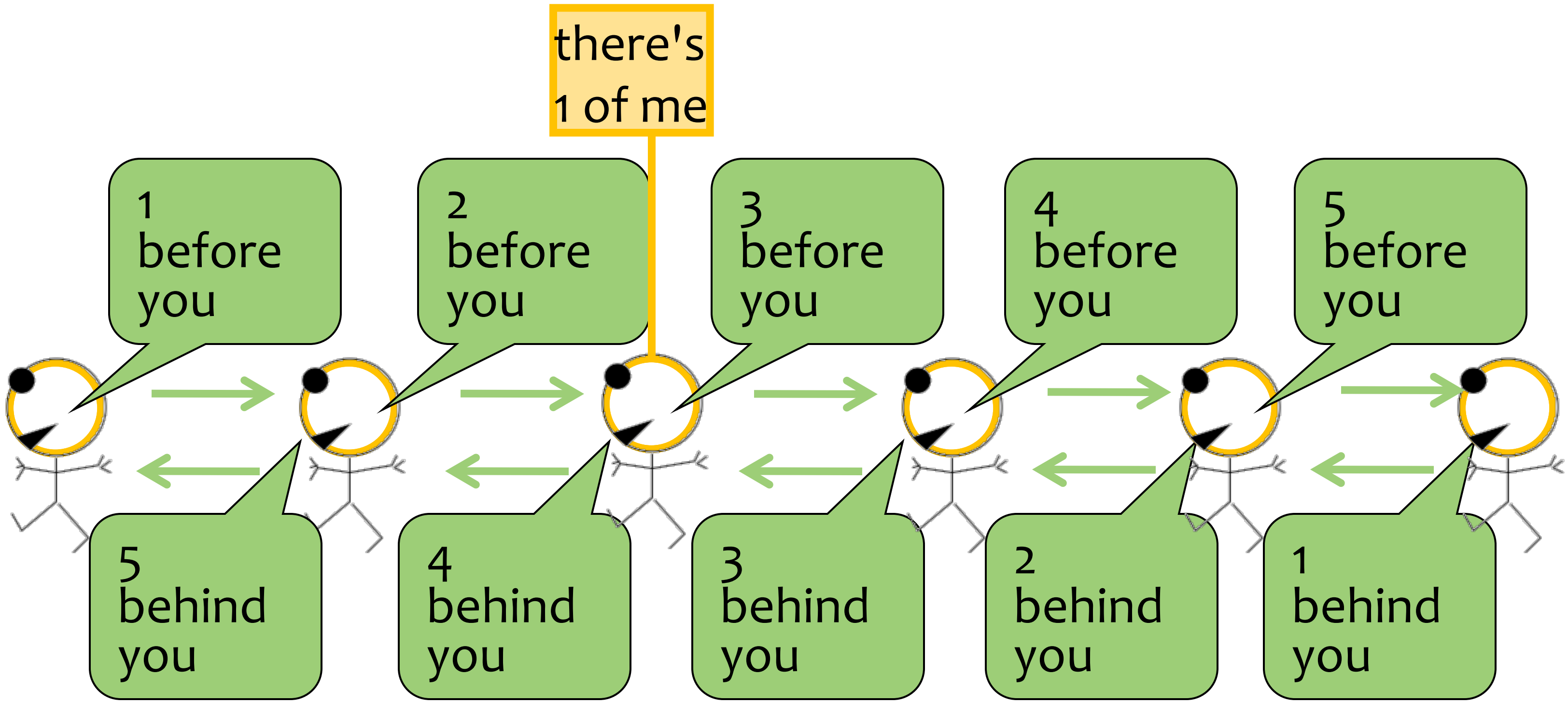
- E is common symptom of two causes, X and Y
- After observing E=1, *both* X and Y become *more* probable
- After observing E=1 and X=1, Y becomes *less* probable (compared to just E=1)
    - since X alone is enough to "explain" E=1

INFERENCE IN DGM

from: Russell and Norvig

Great Ideas in ML: Message Passing

Count the soldiers

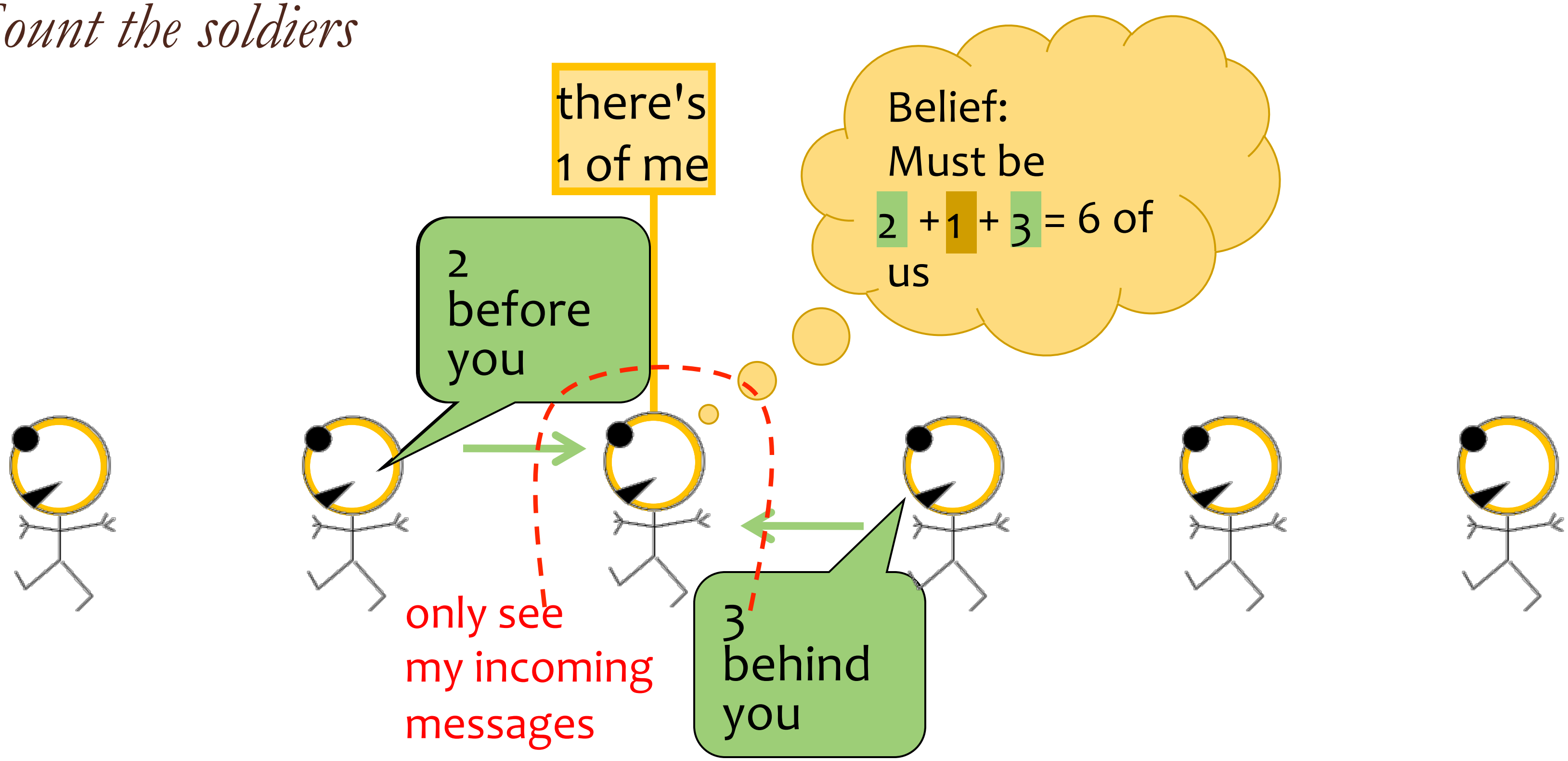


adapted from MacKay (2003) textbook

Thanks Matt Gormley

Great Ideas in ML: Message Passing

Count the soldiers

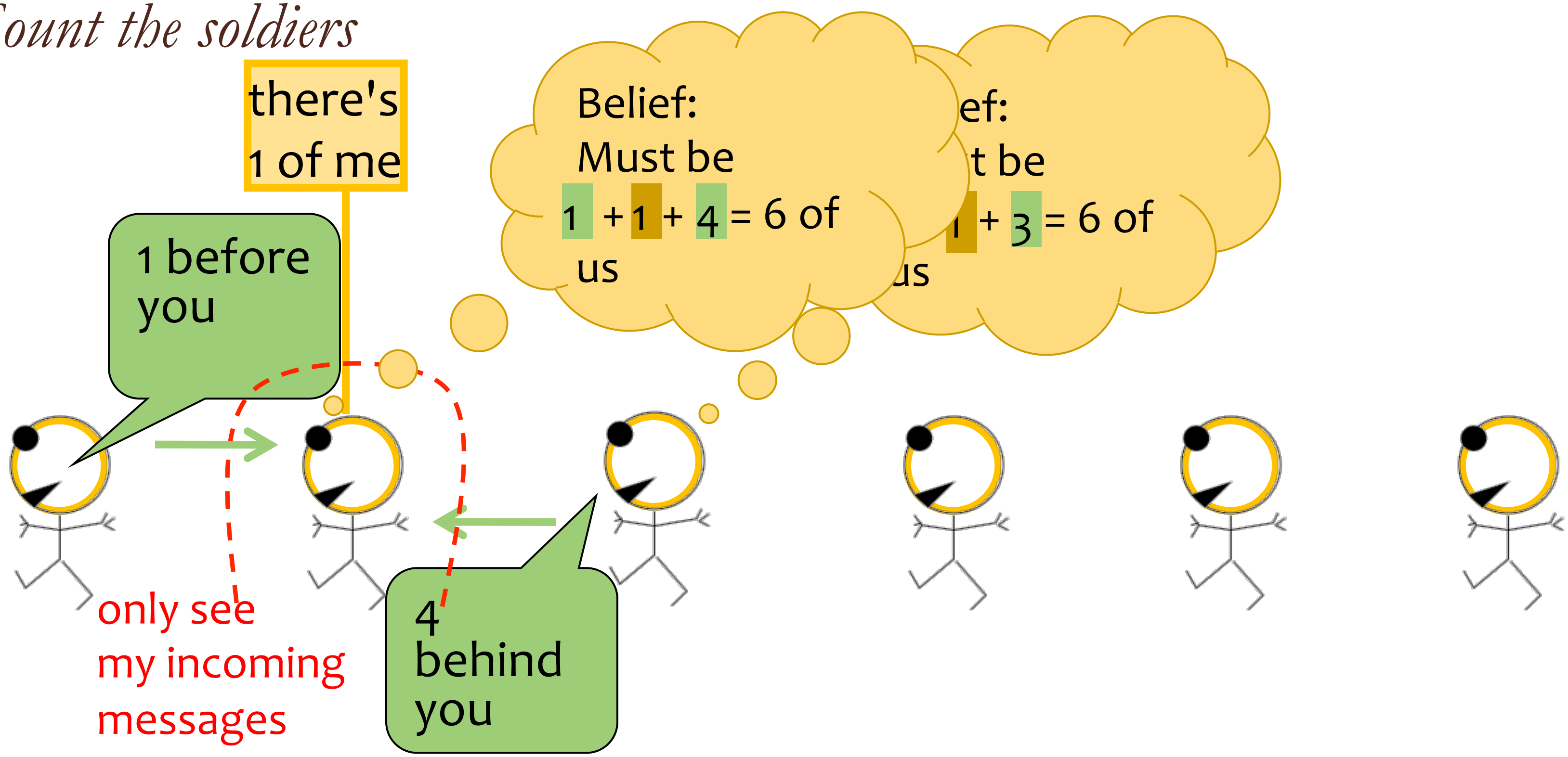


adapted from MacKay (2003) textbook

Thanks Matt Gormley

Great Ideas in ML: Message Passing

Count the soldiers

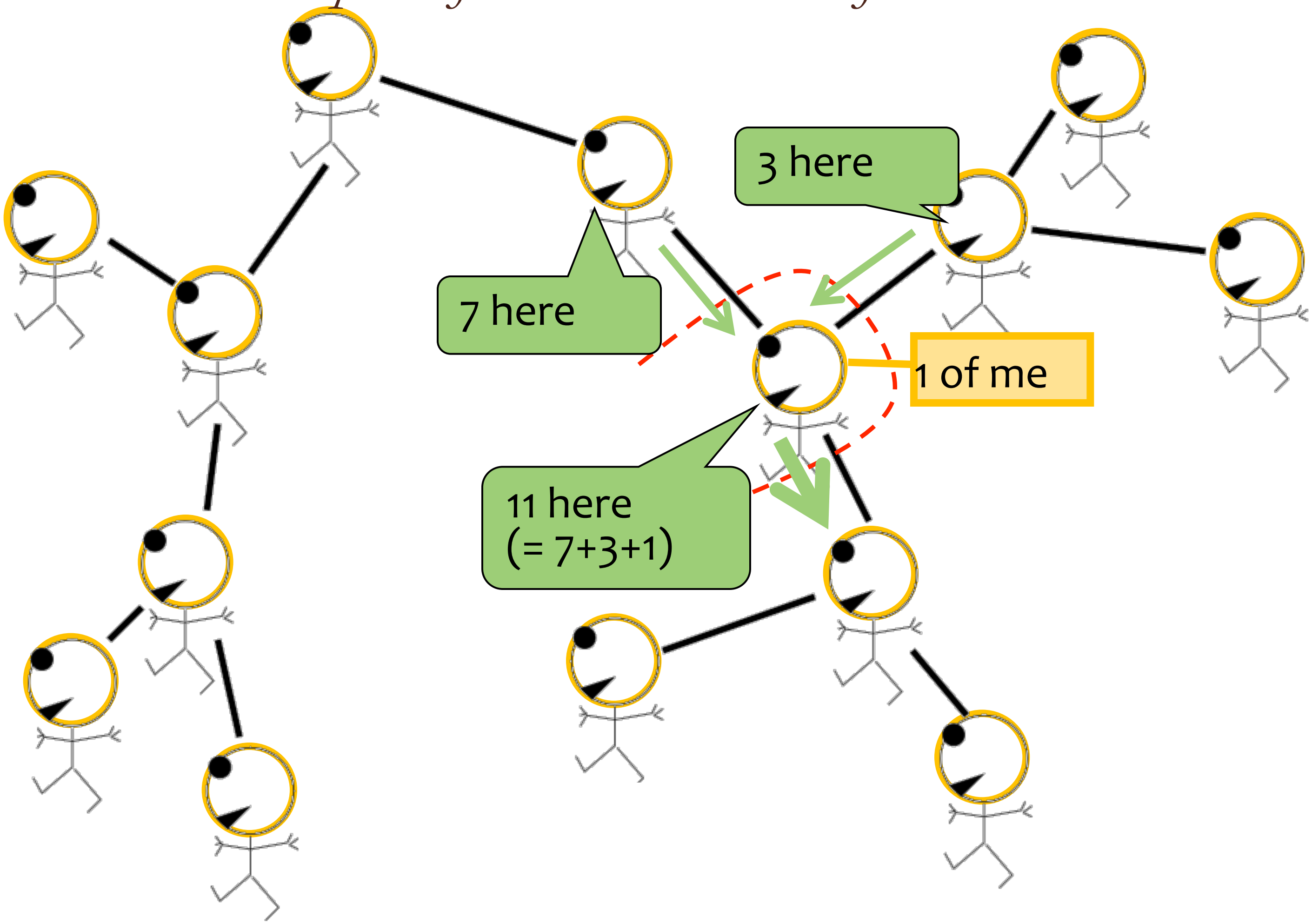


adapted from MacKay (2003) textbook

Thanks Matt Gormley

Great Ideas in ML: Message Passing

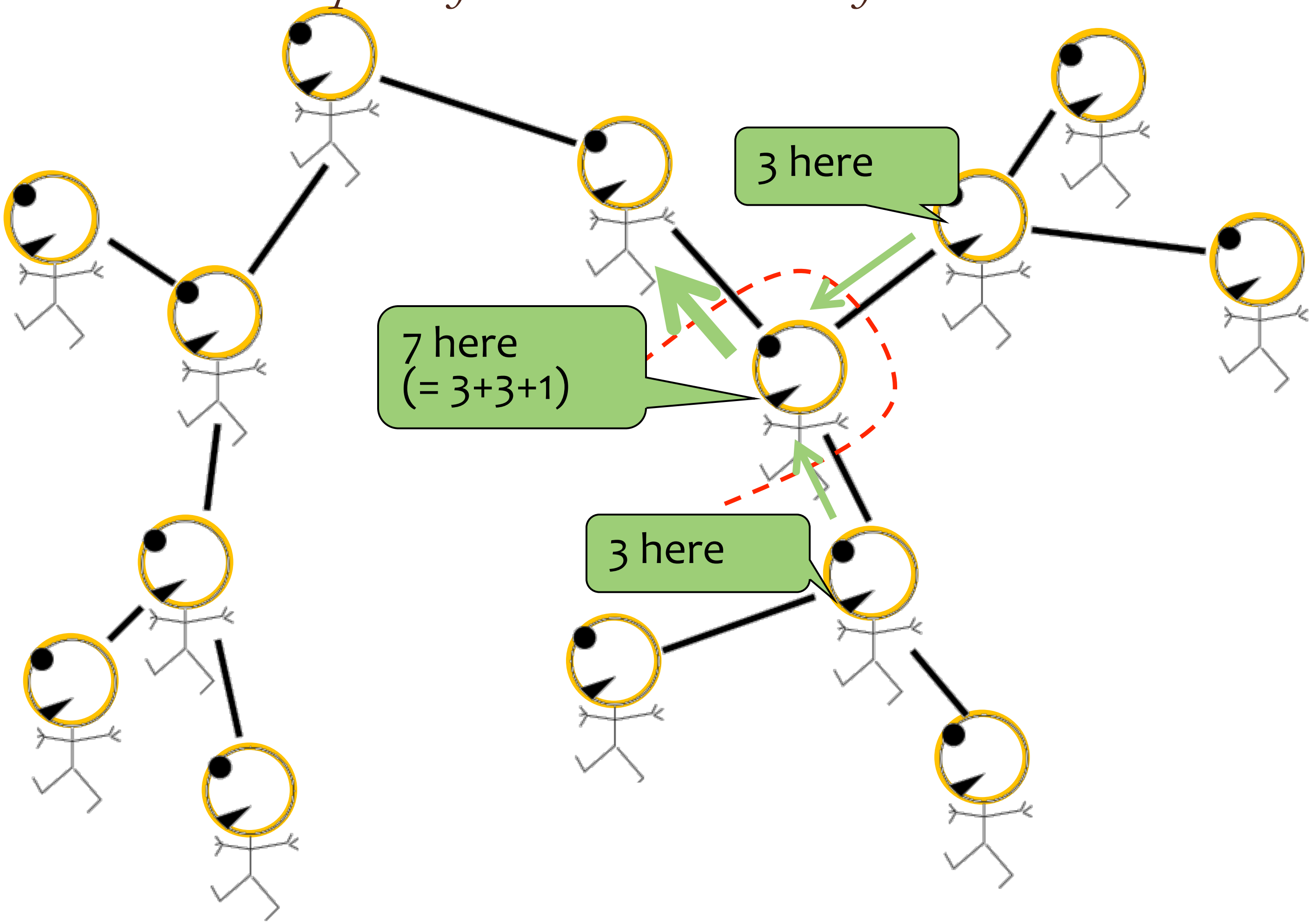
Each soldier receives reports from all branches of tree



adapted from MacKay (2003) textbook

Great Ideas in ML: Message Passing

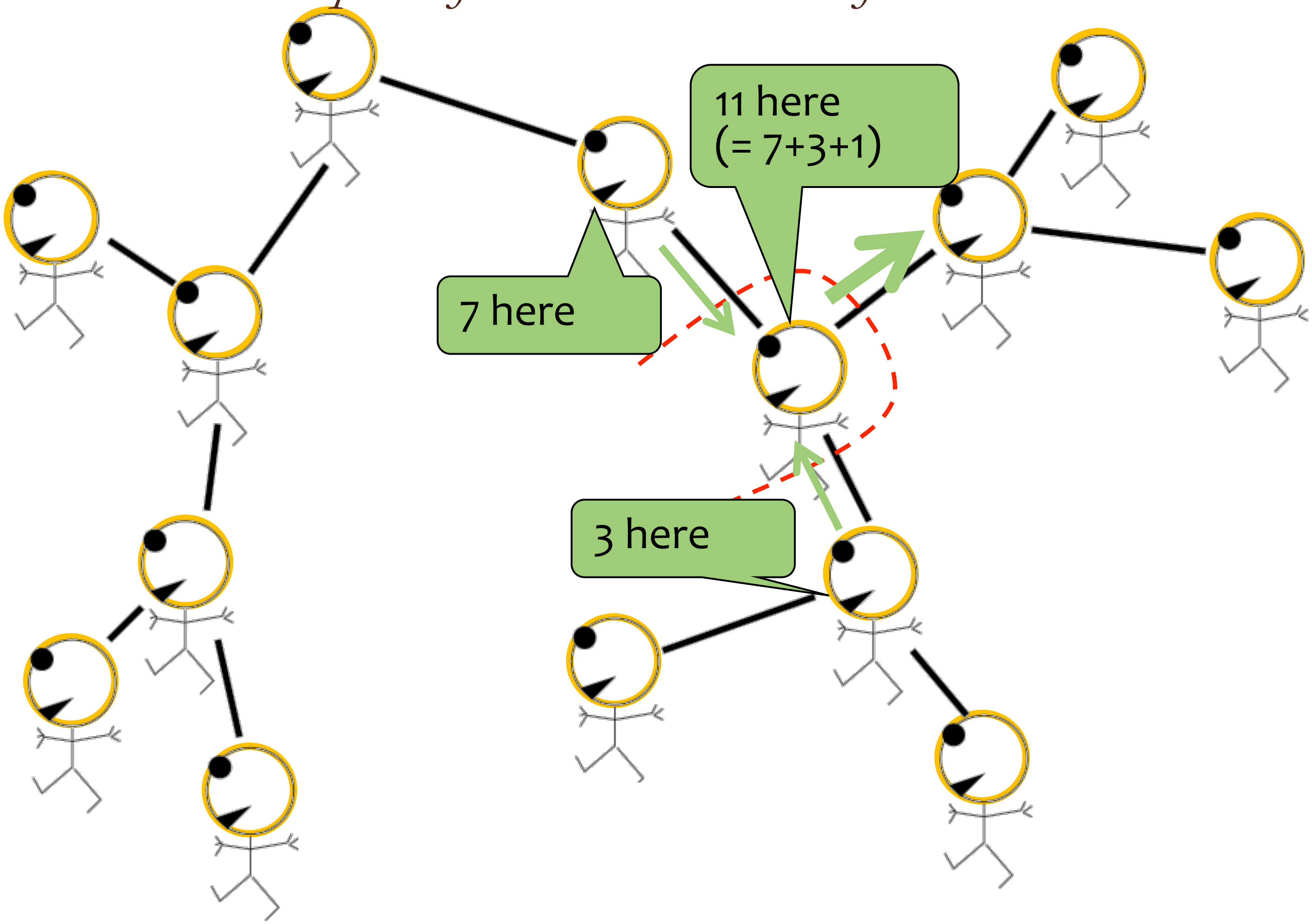
Each soldier receives reports from all branches of tree



adapted from MacKay (2003) textbook

Great Ideas in ML: Message Passing

Each soldier receives reports from all branches of tree

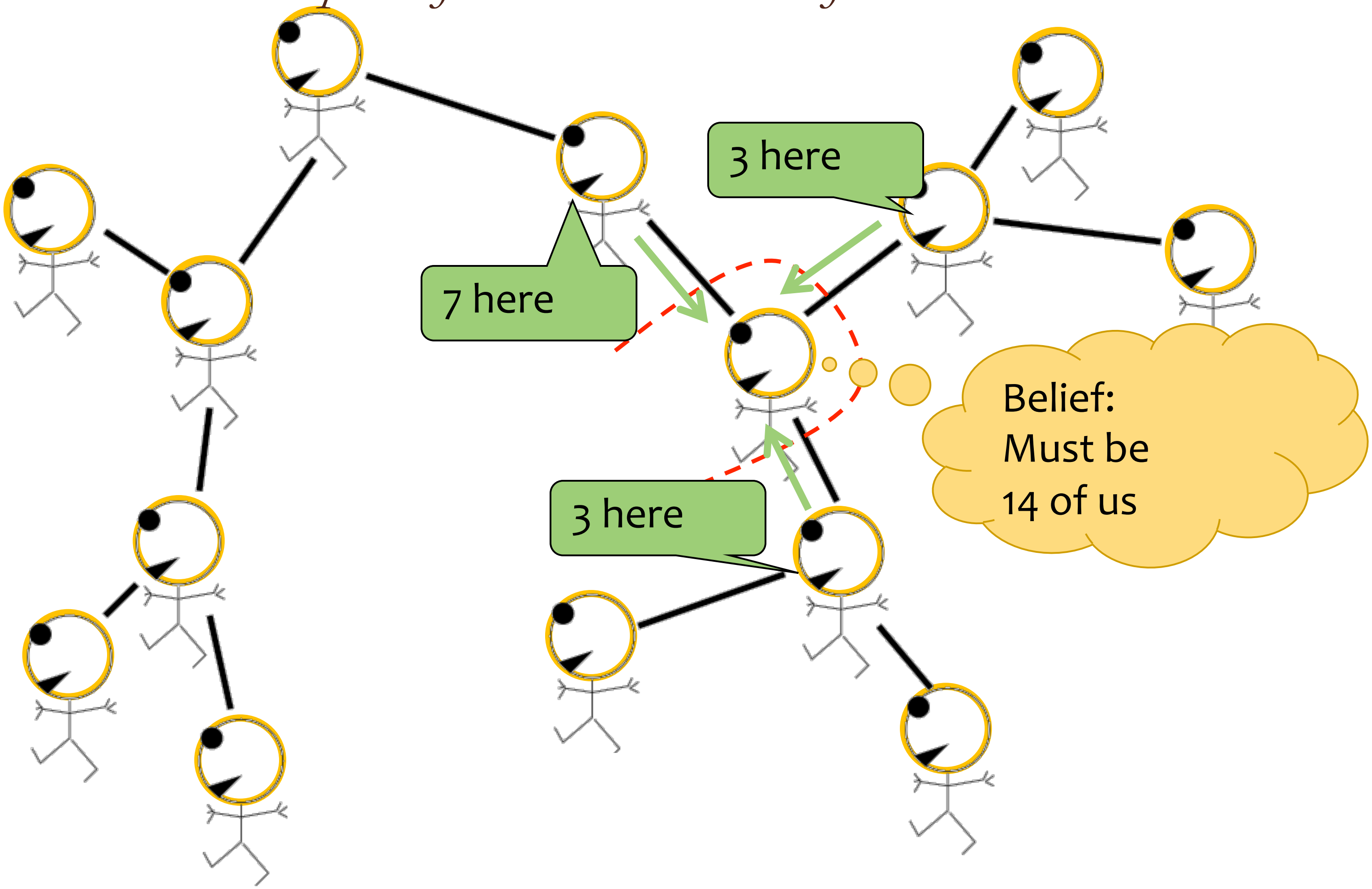


adapted from MacKay (2003) textbook

Thanks Matt Gormley

Great Ideas in ML: Message Passing

Each soldier receives reports from all branches of tree

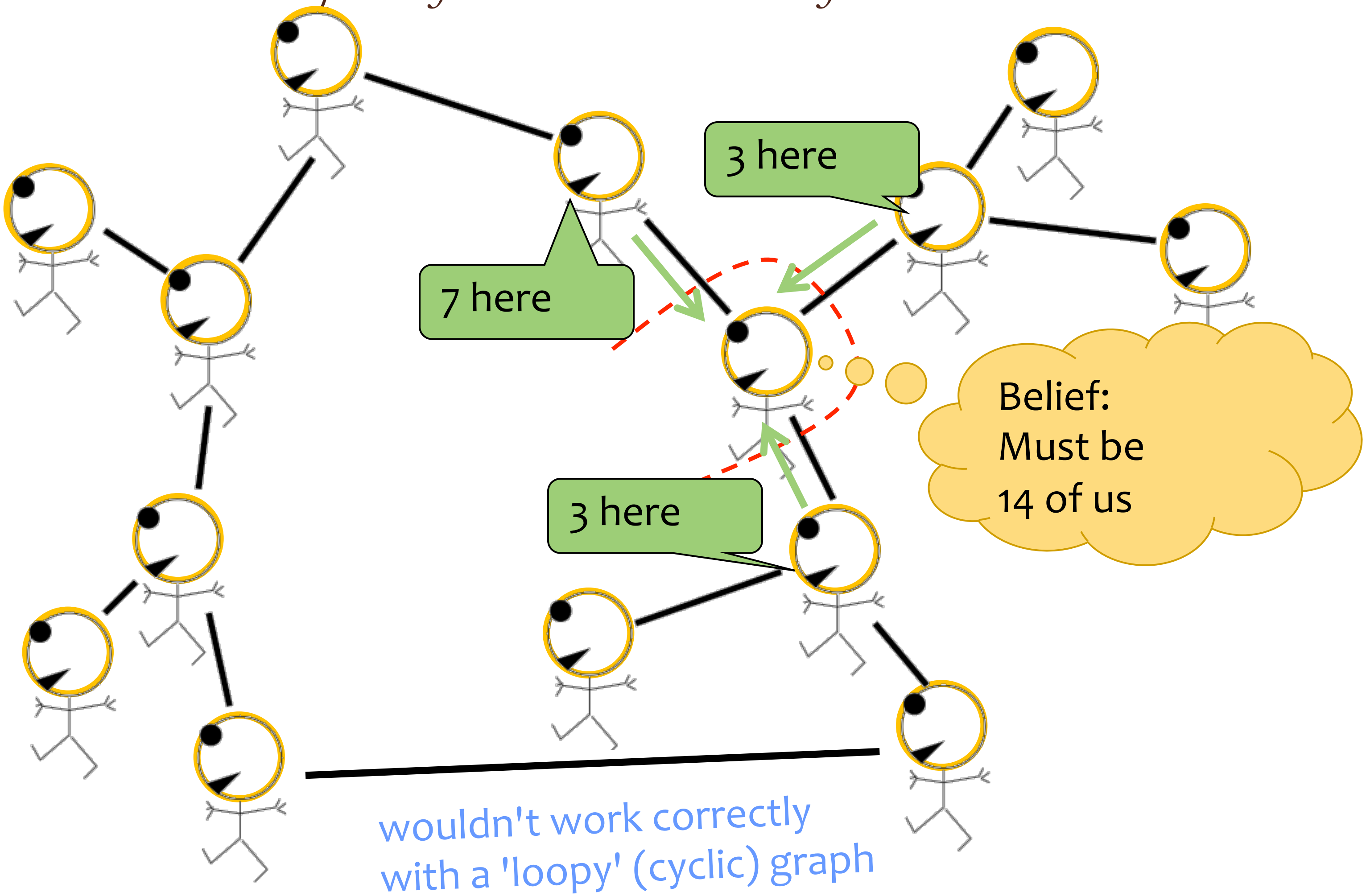


adapted from MacKay (2003) textbook

Thanks Matt Gormley

Great Ideas in ML: Message Passing

Each soldier receives reports from all branches of tree



adapted from MacKay (2003) textbook

Thanks Matt Gormley

Message Passing and Inference

- Message passing:
  – Handles the “simple” (and tractable) case of polytree* exactly
  – Handles intractable cases approximately
  – Often an important part of exact algorithms for more complex cases

Message Passing and Counting

- Message passing is *almost* counting
- Instead of passing counts, we'll be passing probability distributions (beliefs) of various types

Inference in Bayes Nets

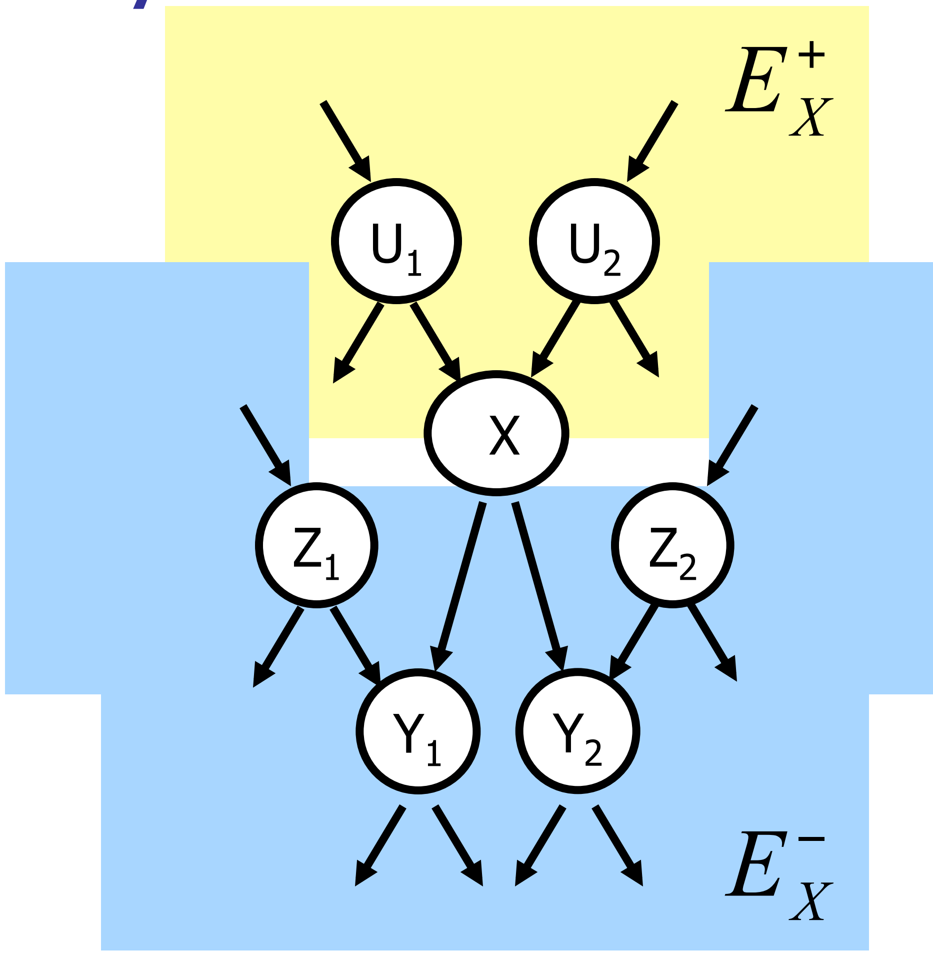
- The belief propagation algorithm

Inference in Bayes nets

- General problem: given evidence $E_1,\ldots,E_k$ compute $P(X|E_1,..,E_k)$ for any X
- Big assumption: graph is "polytree"
  - <=1 undirected path between any nodes X,Y
- Notation:

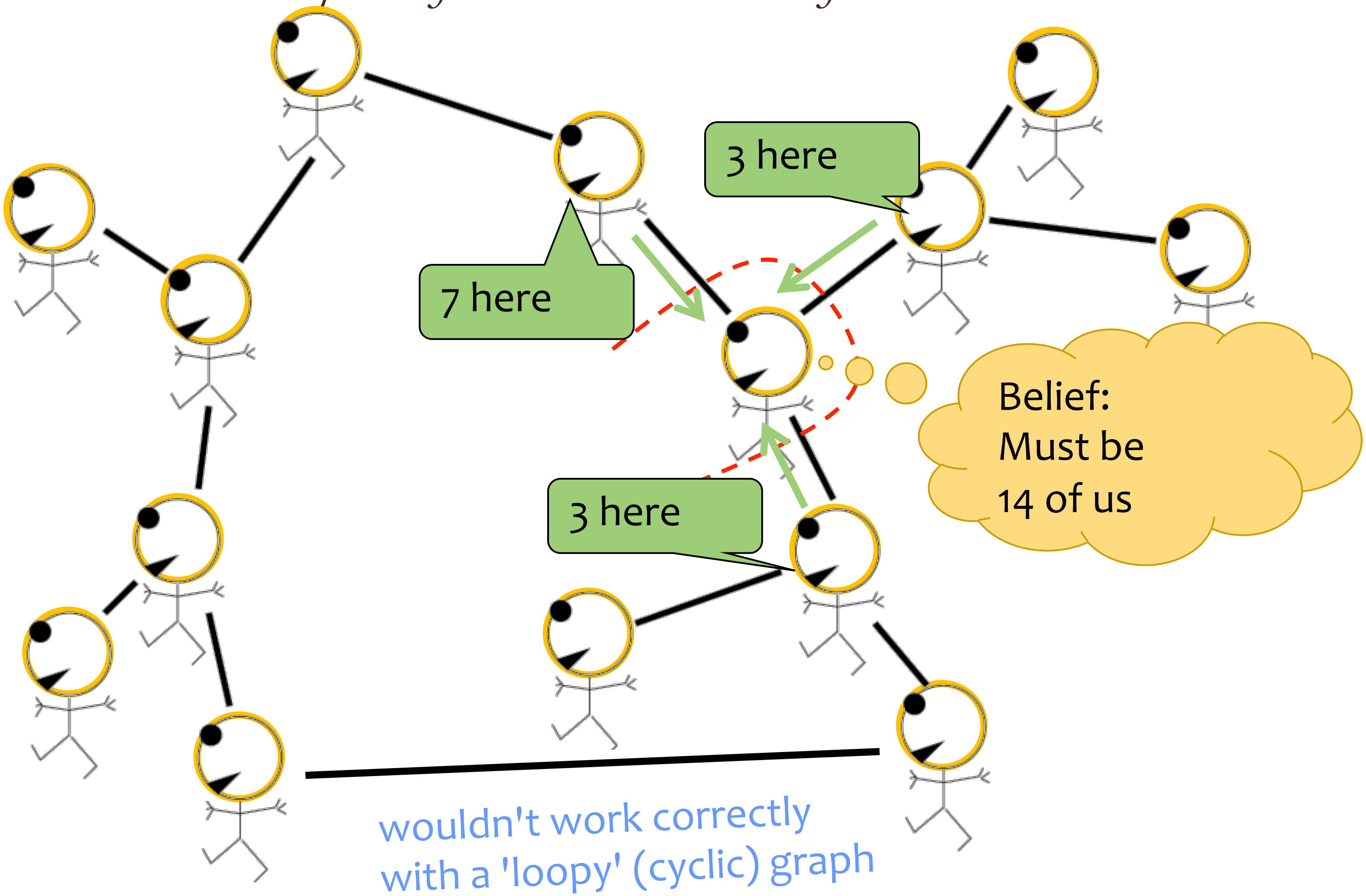
$$E_X^+ = \text{"causal support" for X}$$

$$E_X^- = \text{"evidential support" for X}$$



Great Ideas in ML: Message Passing

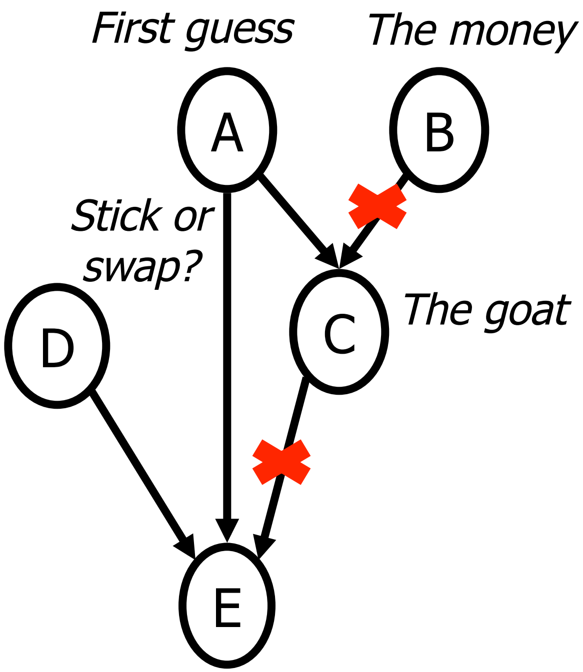
Each soldier receives reports from all branches of tree

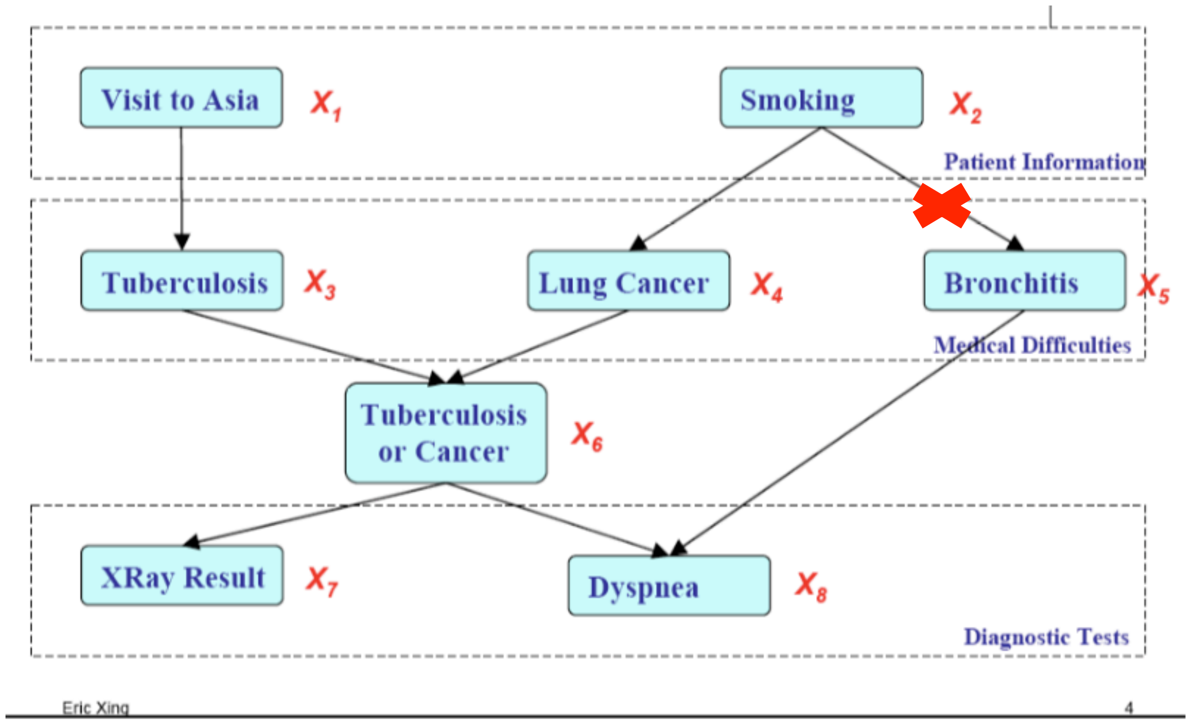


adapted from MacKay (2003) textbook

Thanks Matt Gormley

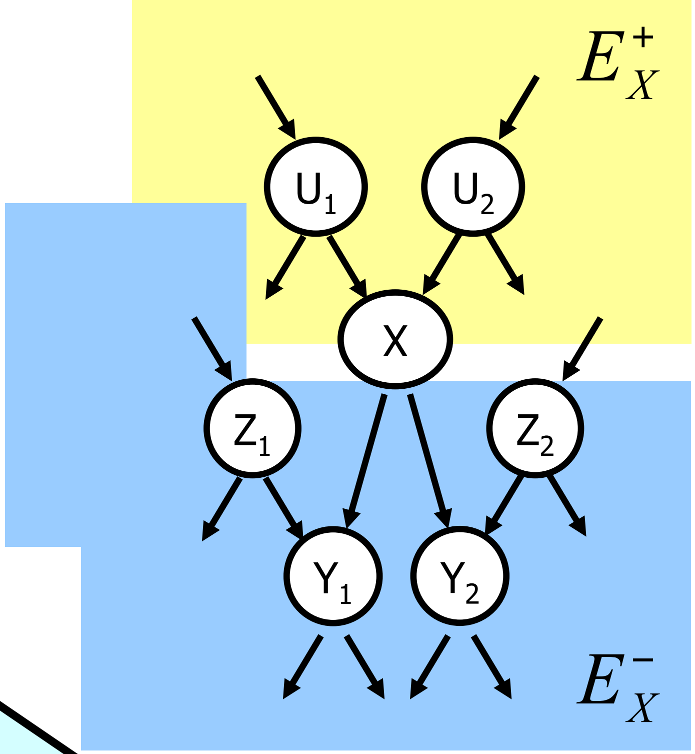
Hot* or not?





*Hot = polytree
(Must be one undirected path between every pair of nodes)

Inference in Bayes nets: P(X|E)



$$P(X \mid E) = P(X \mid E_X^+, E_X^-)$$

$$= \frac{P(E_X^- \mid X, \cancel{E_X^+})P(X \mid E_X^+)}{P(E_X^- \mid E_X^+)}$$

$$P(X \mid E) \propto P(E_X^- \mid X)\underline{P(X \mid E_X^+)}$$

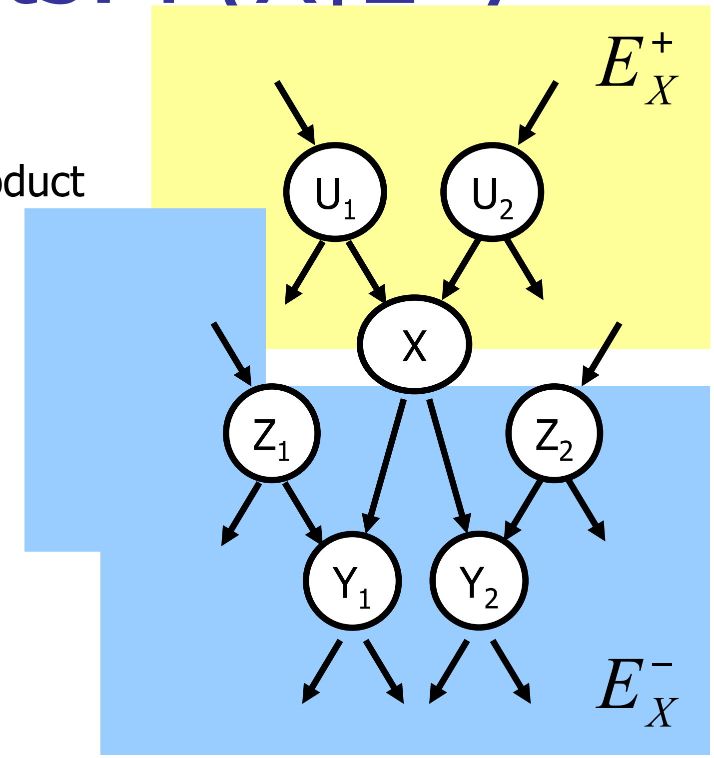
$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

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

lets start with $\underline{P(X \mid E_X^+)}$....

E+: causal support
E-: evidential support

Inference in Bayes nets: $P(X|E^+)$



$$P(X \mid E_X^+) =_{\text{d-sep.}}$$

d-sep – write as product

$$\sum_{u_1,u_2} P(X \mid u_1, u_2, \cancel{E_X^+}) \underbrace{P(u_1, u_2 \mid E_X^+)}$$

$$\sum_{u_1,u_2} P(X \mid U_1 = u_1, U_2 = u_2, \ldots.)$$

Inference in Bayes nets: P(X|E+)

$$P(X \mid E_X^+) =$$

$$\sum_{u_1,u_2} P(X \mid u_1, u_2, \cancel{E_X^+}) P(u_1, u_2 \mid E_X^+)$$

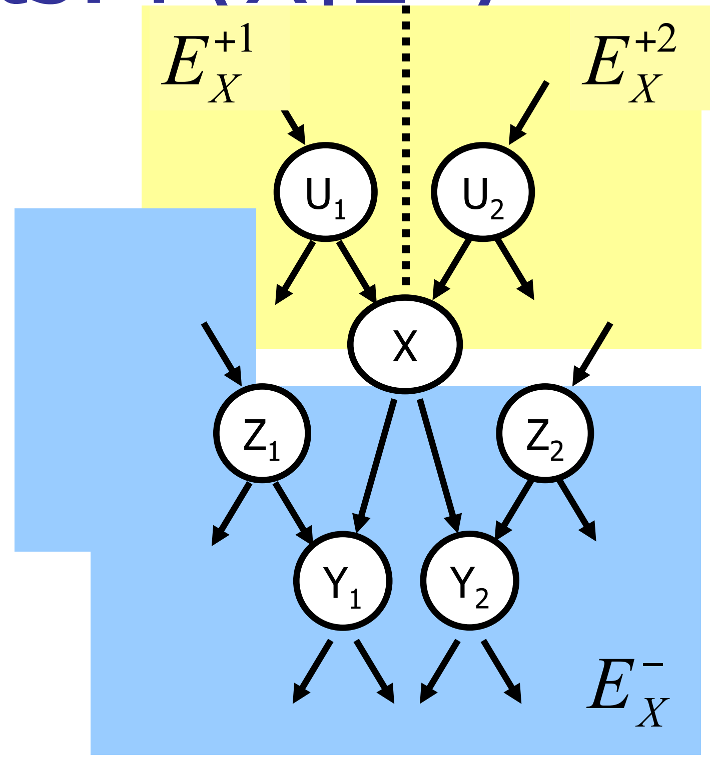
$$= \underbrace{\sum_{u_1,u_2} P(X \mid u_1, u_2,)}_{\text{CPT table lookup}} \underbrace{\prod_j P(u_j \mid E_X^{+j})}_{\text{Recursive call to } P(\cdot \mid E^+)}$$

CPT table lookup

Recursive call to P(·|E+)

$$= \sum_{\mathbf{u}} P(X \mid \mathbf{u}) \prod_j P(u_j \mid E_{Uj \setminus X})$$

Evidence for Uj that doesn't go thru X



So far: simple way of propagating requests for "belief due to causal evidence" **up** the tree

I.e. info on Pr(X|E+) flows down

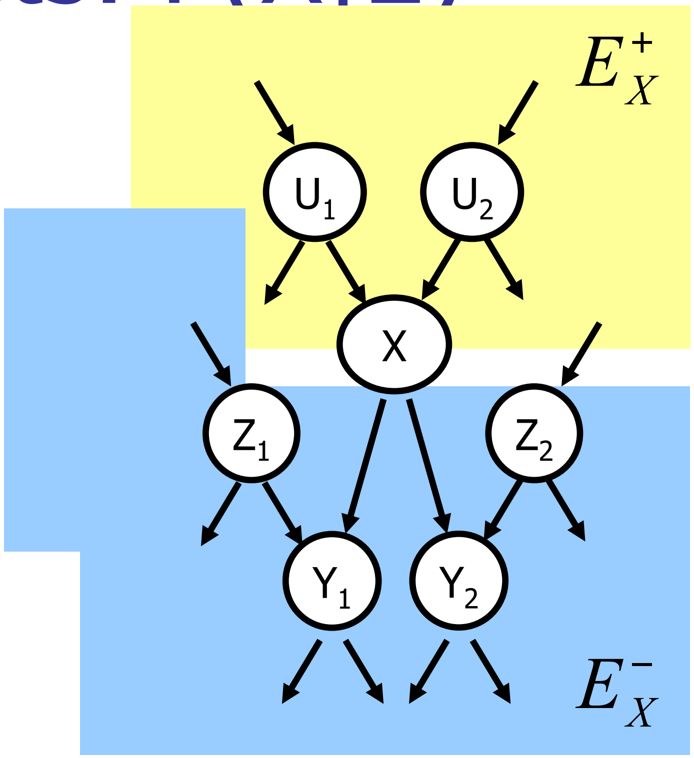
Inference in Bayes nets: P(X|E)

$$P(X \mid E) = P(X \mid E_X^+, E_X^-)$$

$$= \frac{P(E_X^- \mid X, E_X^+) P(X \mid E_X^+)}{P(E_X^- \mid E_X^+)}$$

$$P(X \mid E) \propto \underline{P(E_X^- \mid X)} P(X \mid E_X^+)$$

now: $\underline{P(E_X^- \mid X)}$....



Inference in Bayes nets: $P(E^-|X)$ **simplified**

$$P(X \mid E) \propto P(E_X^- \mid X)P(X \mid E_X^+)$$

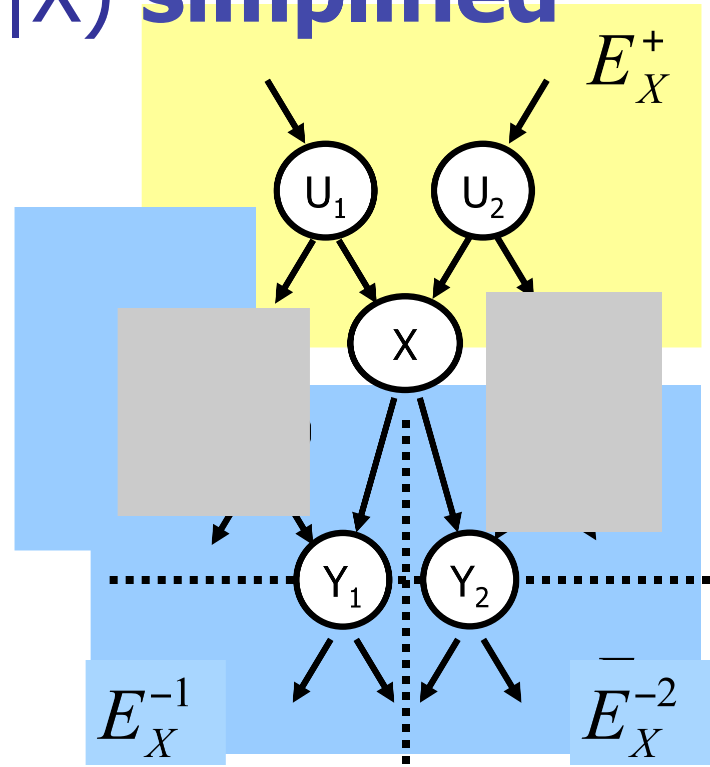
$$P(E_X^- \mid X) = \prod_j P(E_{Yj\backslash X}^- \mid X)$$

d-sep + polytree

$$= \prod_j \sum_{y_{jk}} P(y_{jk} \mid X)P(E_{Yj\backslash X}^- \mid X, y_{jk})$$

$$= \prod_j \sum_{y_{jk}} P(y_{jk} \mid X)\underbrace{P(E_{Yj}^{-j} \mid y_{jk})}$$

Recursive call to $P(E^-|\cdot)$



So far: simple way of propagating requests for "belief due to evidential support" **down** the tree

I.e. info on Pr(E-|X) flows up

$E_{Y_j}^{-j}$: evidential support for Y_j

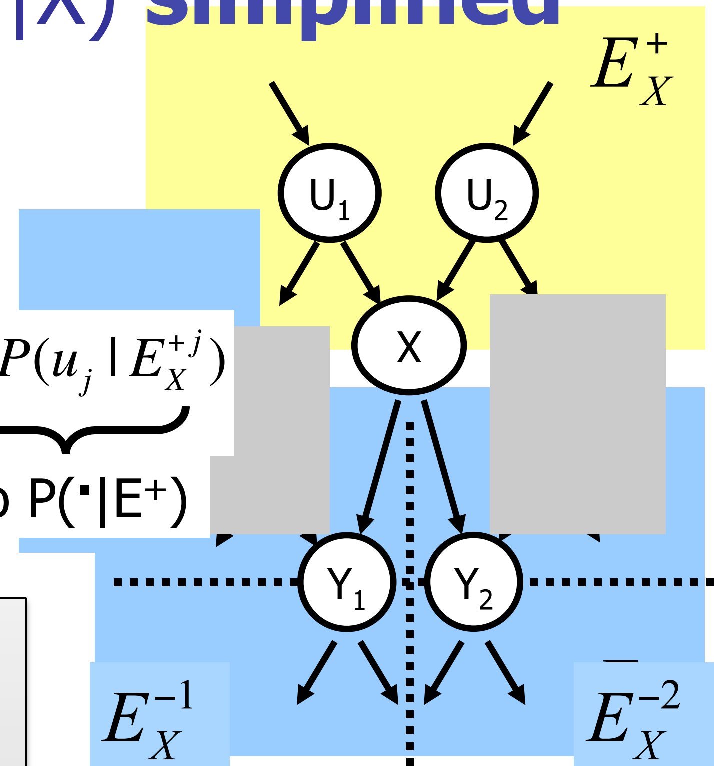
$E_{Y_j\backslash X}$: evidence for Y_j excluding support through X

Inference in Bayes nets: P(E⁻|X) **simplified**

$$P(X \mid E) \propto P(E_X^- \mid X)\, P(X \mid E_X^+)$$

$$= \prod_j \sum_{y_{jk}} P(y_{jk} \mid X) \underbrace{P(E_{Yj}^{-j} \mid y_{jk})} * \sum_{u_1,u_2} P(X \mid u_1, u_2) \prod_j \underbrace{P(u_j \mid E_X^{+j})}$$

Recursive call to $P(E^- \mid \cdot)$ Recursive call to $P(\cdot \mid E^+)$



Usual implementation is message passing:

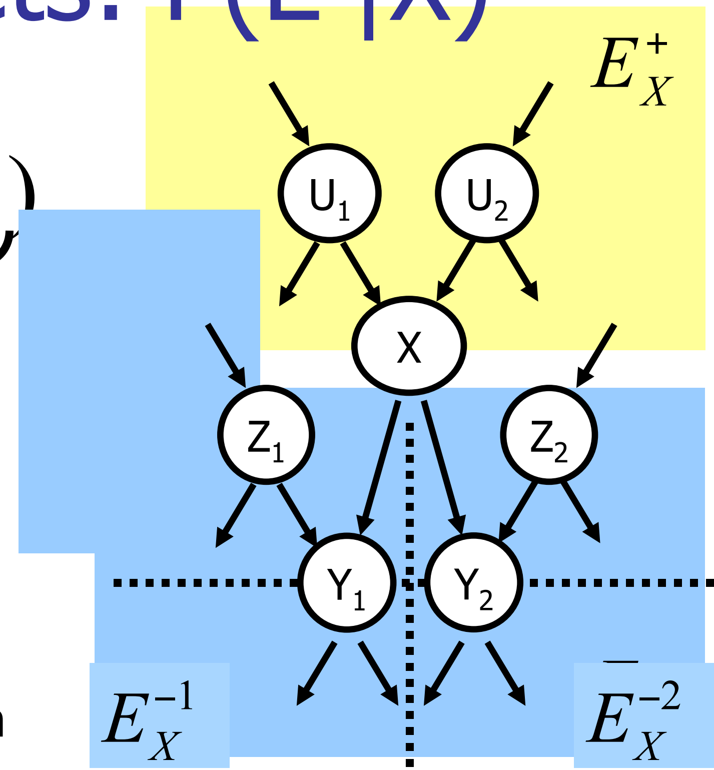
- Send values for P(E-|X) up the tree (vertex to parent)
    - Wait for all children's msgs before sending
- Send values for P(X|E+) down the tree (parent to child)
    - Wait for all parent's msgs before sending
- Compute P(X|E) when after all msgs to X are recieved

$E_{Y_j}^{-j}$: evidential support for Y_j

$E_{Y_j \setminus X}$: evidence for Y_j excluding support through X

Inference in Bayes nets: $P(E^-|X)$

$$P(X \mid E) \propto P(E_X^- \mid X)\underbrace{P(X \mid E_X^+)}_{\text{recursion}}$$



$$P(E_X^- \mid X) = \prod_j P(E_{Yj\backslash X} \mid X)$$

$$= \prod_j \sum_{y_{jk}, z_{jk}} P(y_{jk}, z_{jk} \mid X) P(E_{Yj\backslash X} \mid X, y_{jk}, z_{jk})$$

$$= \prod_j \sum_{y_{jk}, z_{jk}} P(y_{jk}, z_{jk} \mid X) \cdot$$

our decomposition

$$\cdot\ P(E_{Y_j}^{-j} \mid X, y_{jk}, z_{jk}) \cdot P(E_{Y_j\backslash X}^{+j} \mid X, y_{jk}, z_{jk})$$

d-sep

$$= \prod_j \sum_{y_{jk}} P(E_{Y_j}^{-j} \mid y_{jk}) \sum_{z_{jk}} P(E_{Y_j\backslash X}^{+j} \mid z_{jk}) P(y_{jk}, z_{jk} \mid X)$$

$E_{Y_j}^{-j}$: evidential support for Y_j

$E_{Y_j\backslash X}$: evidence for Y_j excluding support through X

Inference in Bayes nets: P(E⁻|X)

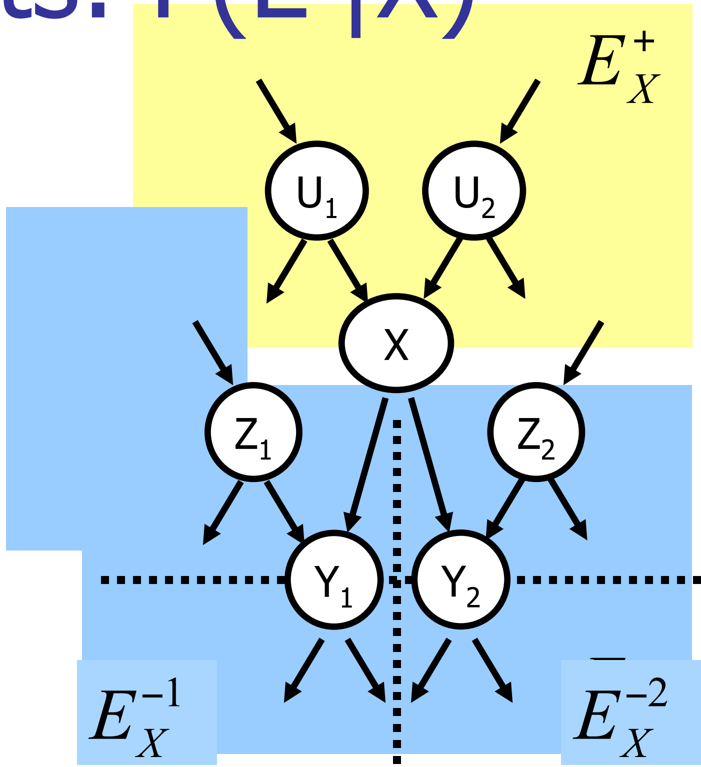
$$P(E_X^- \mid X) = \prod_j P(E_X^{-j} \mid X)$$

$$= \prod_j \sum_{y_{jk}} P(E_{Y_j}^{-j} \mid y_{jk}) \sum_{z_{jk}} P(E_{Y_j \setminus X}^{+} \mid z_{jk}) P(y_{jk}, z_{jk} \mid X)$$

$$P(E_{Y_j \setminus X}^{+} \mid z_{jk}) = \frac{P(z_{jk} \mid E_{Y_j \setminus X}^{+}) P(E_{Y_j \setminus X}^{+})}{P(z_{jk})}$$

$$P(y_{jk}, z_{jk} \mid X) = P(y_{jk} \mid X, z_{jk}) P(z_{jk} \mid X)$$

$$= \prod_j \sum_{y_{jk}} P(E_{Y_j}^{-j} \mid y_{jk}) \sum_{z_{jk}} \beta_j P(z_{jk} \mid E_{Y_j \setminus X}^{+}) P(y_{jk} \mid X, z_{jk})$$



Inference in Bayes nets: $P(E^-|X)$

$$P(E_X^- \mid X) = \prod_j P(E_X^{-j} \mid X)$$

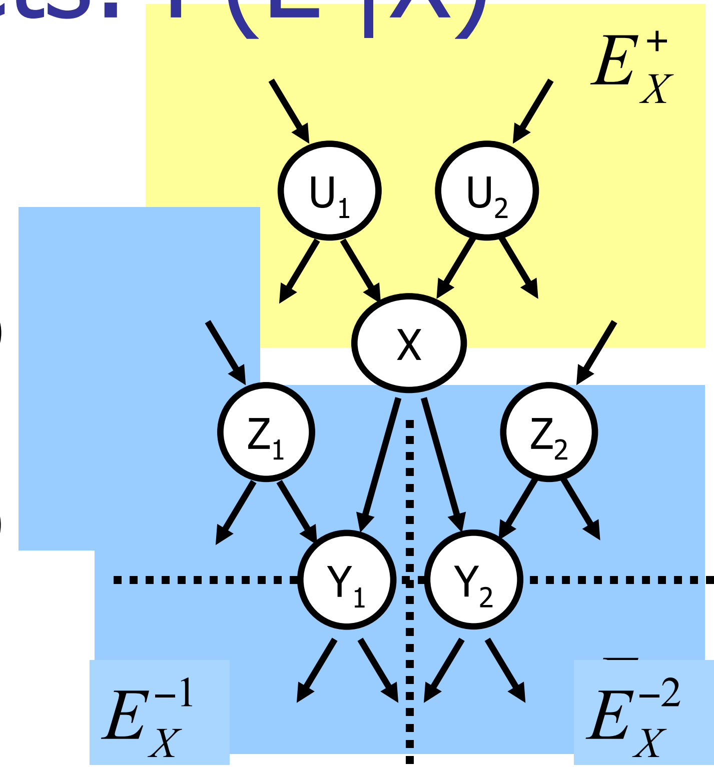
$$= \prod_j \sum_{y_{jk}} P(E_{Y_j}^{-j} \mid y_{jk}) \sum_{z_{jk}} \beta_j P(z_{jk} \mid E_{Y_j \backslash X}^{+}) P(y_{jk} \mid X, z_{jk})$$

$$= \beta \prod_j \sum_{y_{jk}} \underbrace{P(E_{Y_j}^{-j} \mid y_{jk})}_{\text{Recursive call to } P(E^-|\cdot)} \sum_{z_{jk}} \underbrace{P(y_{jk} \mid X, z_{jk})}_{\text{CPT}} \underbrace{P(z_{jk} \mid E_{Z_j \backslash Y_k})}_{\text{Recursive call to } P(\cdot|E)}$$

Recursive call to $P(E^-|\cdot)$

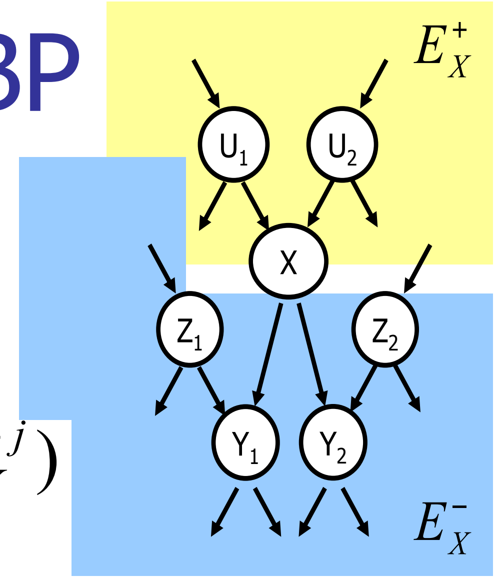
CPT

Recursive call to $P(\cdot|E)$



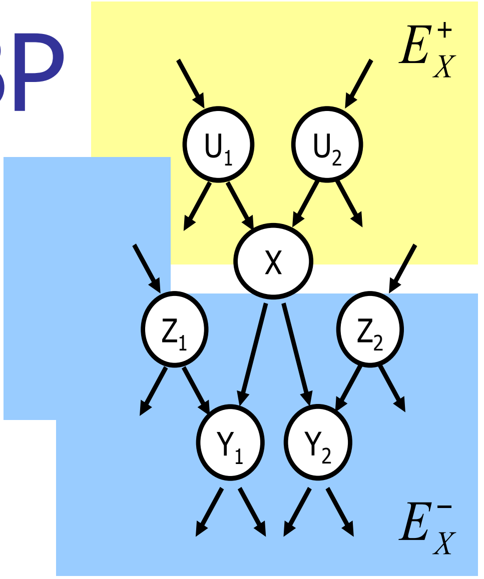
where $\beta = \prod \beta_j$

Recap: Message Passing for BP



- We reduced P(X|E) to product of two recursively calculated parts:
  - $P(X=x|E^+) = \sum_{u_1,u_2} P(X \mid u_1, u_2)\prod_j P(u_j \mid E_X^{+j})$
    - i.e., CPT for X and product of "forward" messages from parents
  - $P(E^-|X=x) = \beta \prod_j \sum_{y_{jk}} P(E_{Y_j}^{-j} \mid y_{jk}) \sum_{z_{jk}} P(y_{jk} \mid X, z_{jk}) P(z_{jk} \mid E_{Z_j \setminus Y_k})$
    - i.e., combination of "backward" messages from parents, CPTs, and $P(Z|E_{Z\setminus Yk})$, a simpler instance of P(X|E)
  - This can also be implemented by message-passing (belief propagation)
    - Messages are distributions – i.e., vectors

Recap: Message Passing for BP



- Top-level algorithm
  - Pick one vertex as the “root”
  - Any node with only one edge is a “leaf”
  - Pass messages from the leaves to the root
  - Pass messages from root to the leaves
  - Now every X has received P(X|E+) and P(E-|X) and can compute P(X|E)

function BELIEF-NET-ASK(X) **returns** a probability distribution over the values of X
  inputs: X, a random variable

  SUPPORT-EXCEPT(X, *null*)

---

function SUPPORT-EXCEPT(X, V) **returns** $\mathbf{P}(X|E_{X\setminus V})$

  if EVIDENCE?(X) **then return** observed point distribution for X
  else
    calculate $\mathbf{P}(E^-_{X\setminus V}|X)$= EVIDENCE-EXCEPT($X$, V)
    U — PARENTS [X]
    if U is empty
      then return $\alpha\ \mathbf{P}(E^-_{X\setminus V}|X)\mathbf{P}(X)$
    else
      for each U_i **in** U
        calculate and store $\mathbf{P}(U_i|E_{U_{ii}\setminus X})$ = SUPPORT-EXCEPT(U_i, X)
      return $\alpha\ \mathbf{P}(E^-_{X\setminus V}|X)\sum_{\mathbf{u}} \mathbf{P}(X|\mathbf{u}) \prod_i \mathbf{P}(U_i|E_{u_i\setminus X})$

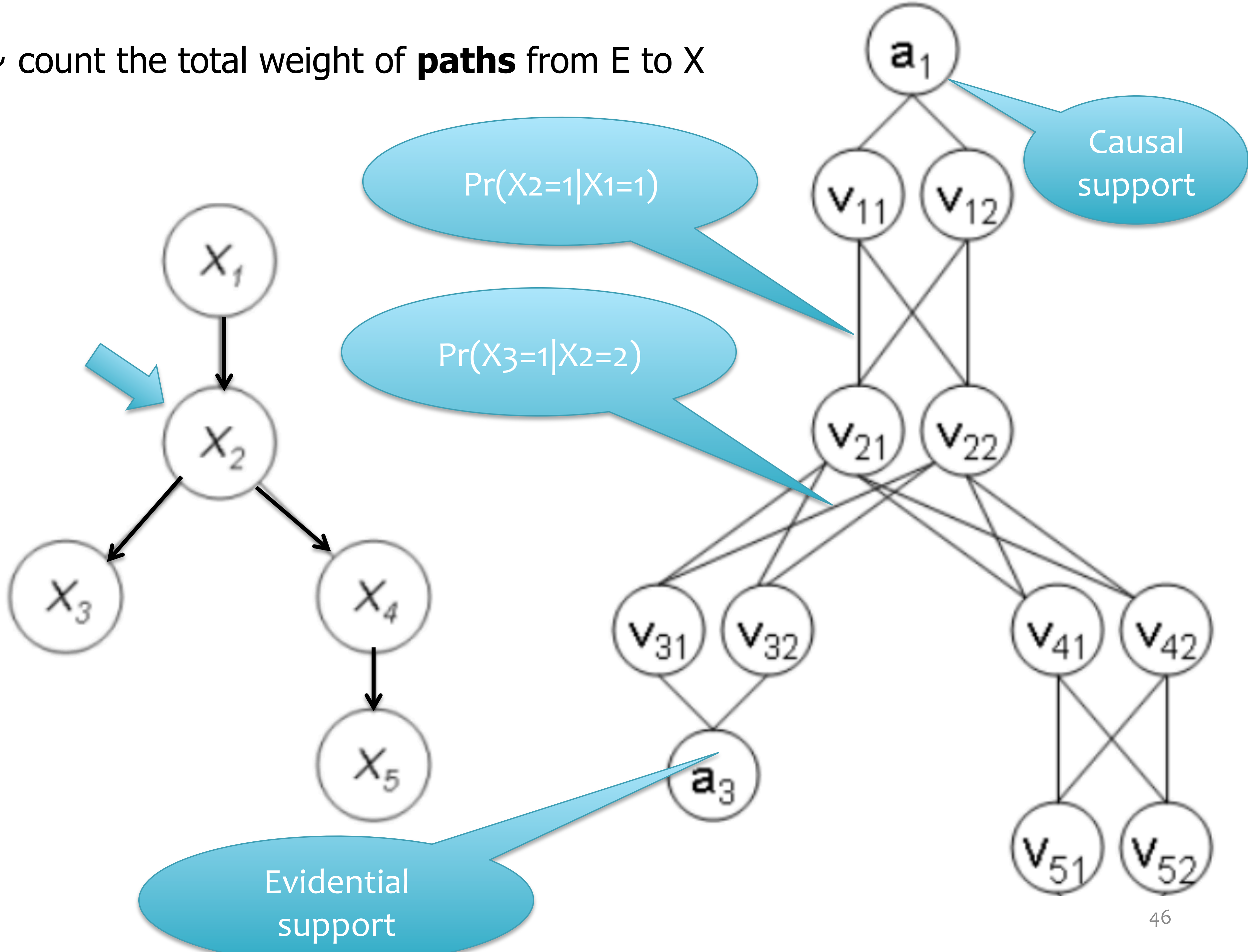
---

function EVIDENCE-EXCEPT(X, V) **returns** $\mathbf{P}(E^-_{X\setminus V}|X)$

  Y — CHILDREN[X] - V
  if Y is empty
    then return a uniform distribution
  else
    for each Y_i **in Y do**
      calculate $\mathbf{P}(E^-_{Y_i}|y_i)$ = EVIDENCE-EXCEPT(Y_i, null)
      Z_i — PARENTS [Y_i] - X
      for each Z_{ij} **in** $\mathbf{Z}_i$
        calculate $\mathbf{P}(Z_{ij}|E_{Z_{ii}\setminus Y_i})$ = SUPPORT-EXCEPT(Z_{ij}, Y_i)
    return $\beta \prod_i \sum_{y_i} P(E^-_{Y_i}|y_i) \sum_{\mathbf{z}_i} \mathbf{P}(y_i|X,\mathbf{z}_i) \prod_j P(z_{ij}|E_{Z_{ij}\setminus Y_i})$

From Russell and Norvig

~ count the total weight of **paths** from E to X

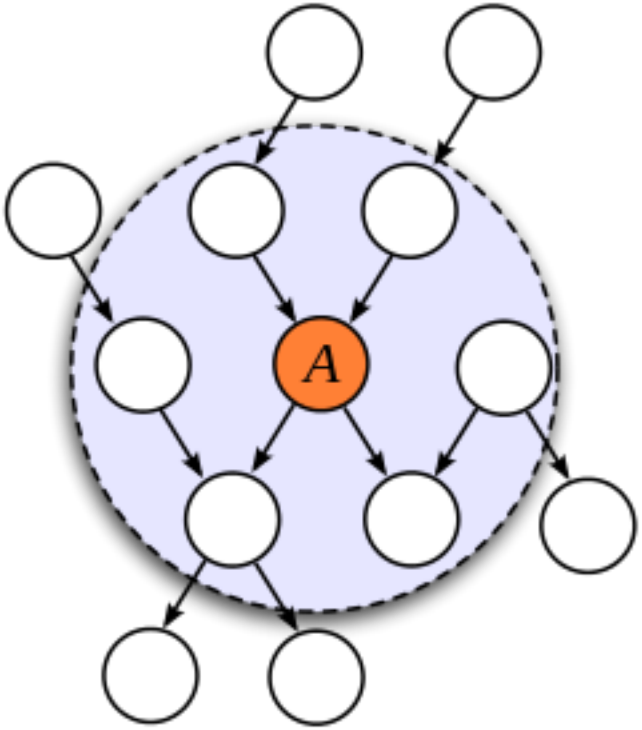


More on message passing/BP

- BP for other graphical models
  - Markov networks
  - Factor graphs
- BP for non-polytrees:
  - Small tree-width graphs
  - Loopy BP

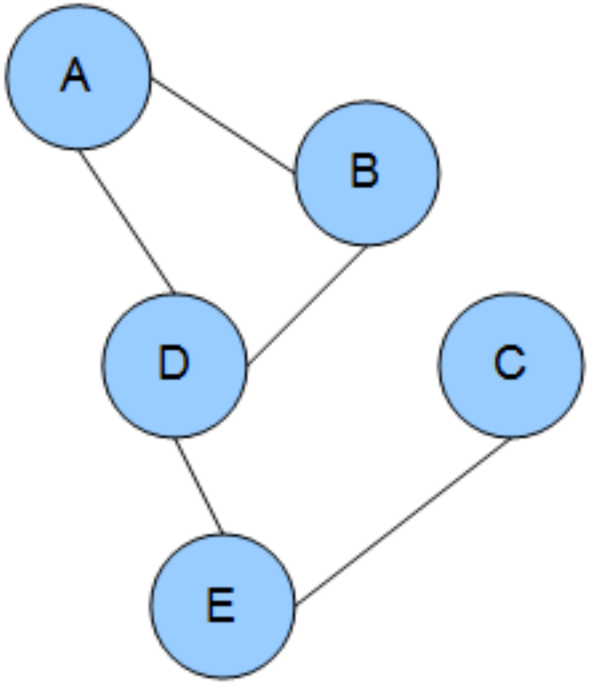
Markov blanket

- The Markov blanket for a random variable A is the set of variables $B_1, \ldots, B_k$ that A is **not** conditionally independent of
  - I.e., not d-separated from
- For DGM this includes parents, children, and "co-parents" (other parents of children)
  - why? explaining away



Markov network

- A Markov network is a set of random variables in an **undirected** graph where
  - each variable is conditionally independent of **all other variables** given its **neighbors**
  - E.g:
    - I<B,{A,D},C>
    - I<B,{A,D},E>

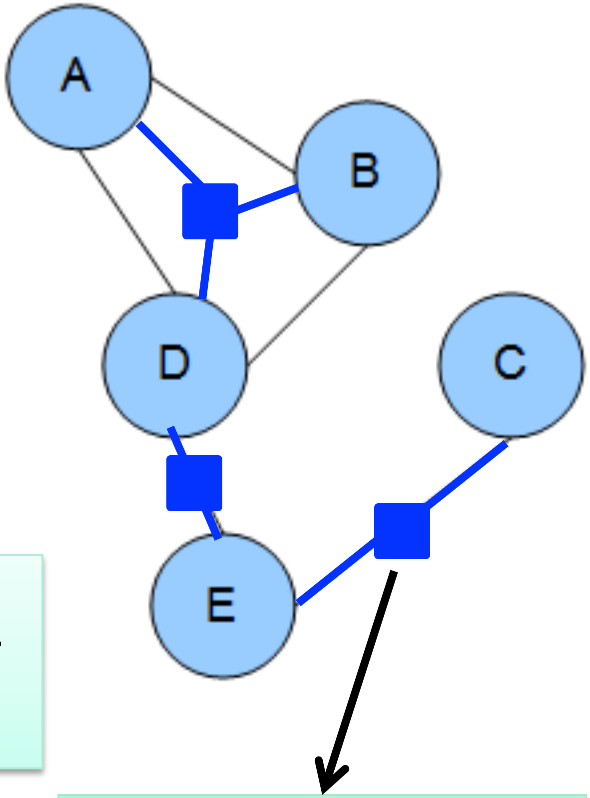


Markov networks

- A Markov network is a set of random variables in an **undirected** graph where
  - each variable is conditionally independent of **all other variables** given its **neighbors**
  - E.g:
    - I<B,{A,D},C>
    - I<B,{A,D},E>
- Instead of CPTs there are **clique potentials**

So the Markov blanket, d-sep stuff is **much** simpler

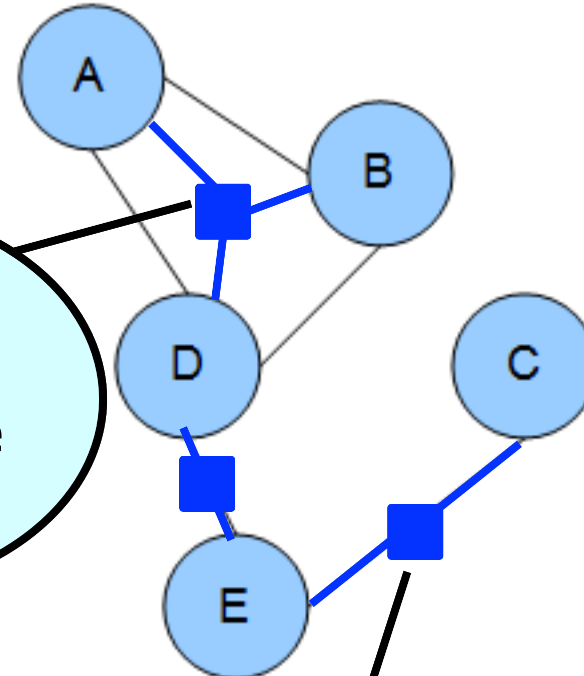
But there's no "generative" story



C	E	φ
0	0	1
0	1	10
1	0	10
1	1	1

Markov networks

A	B	D	φ
0	0	0	1
0	0	1	3.5
0	1	0	[illegible]
…	…	[illegible]	[illegible]
1	1	[illegible]	[illegible]



C	E	φ
0	0	1
0	1	10
1	0	10
1	1	1

$$\Pr(X_1,\dots,X_n = x_1,\dots,x_n)$$

$$= \frac{1}{Z}\prod_{C} \varphi(X_{C,1},\dots,X_{C,k} = x_{C,1},\dots,x_{C,k})$$

these values mapped to these variable indices

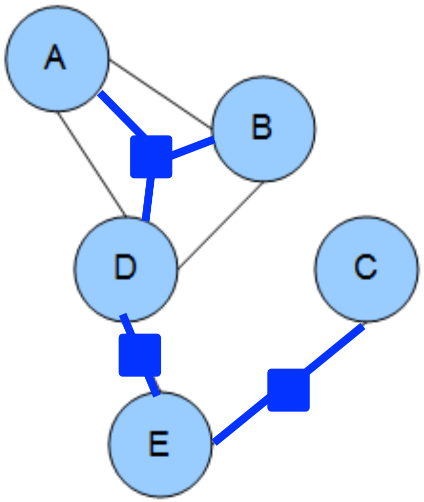
clique C

vars in clique C

- Instead of CPTs there are **clique potentials** which define a joint

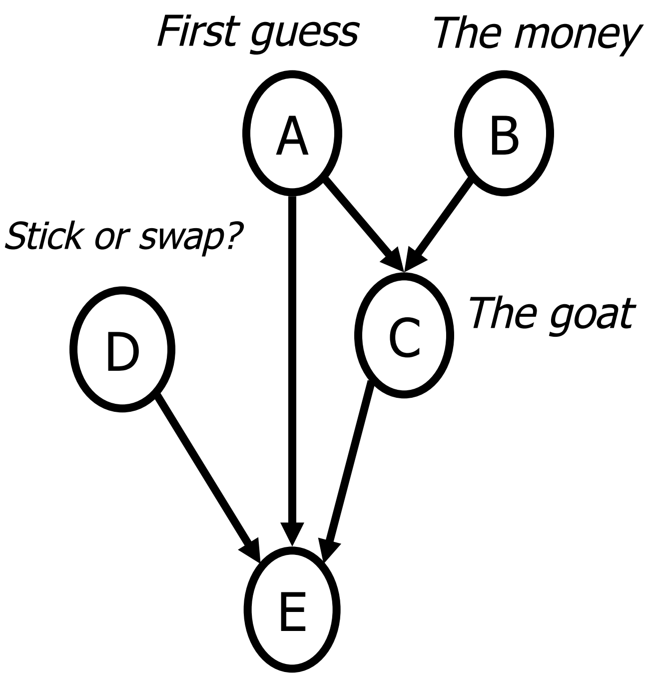
More on message passing/BP

- BP can be extended to other graphical models
  - Markov networks, if they are polytrees
  - "Factor graphs"---which are a generalization of both Markov networks and DGMs
  - Arguably cleaner than DGM BP
- BP for non-polytrees:
  - Small tree-width graphs
  - Loopy BP

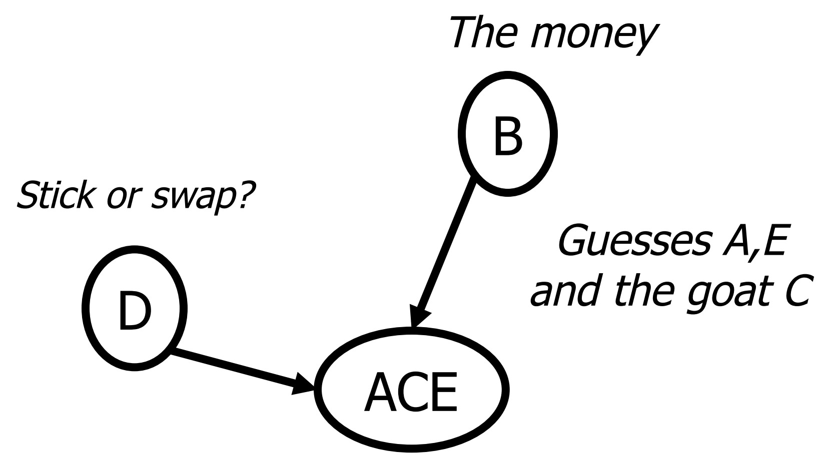


Small tree-width graphs

Not a polytree



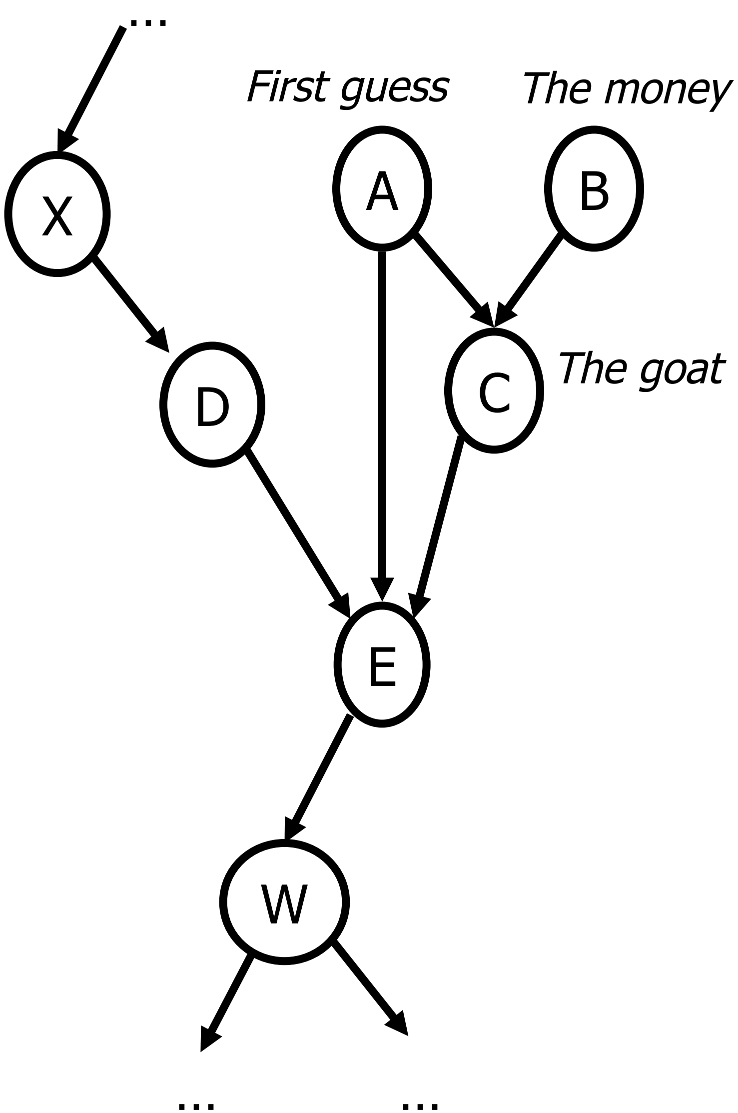
A polytree



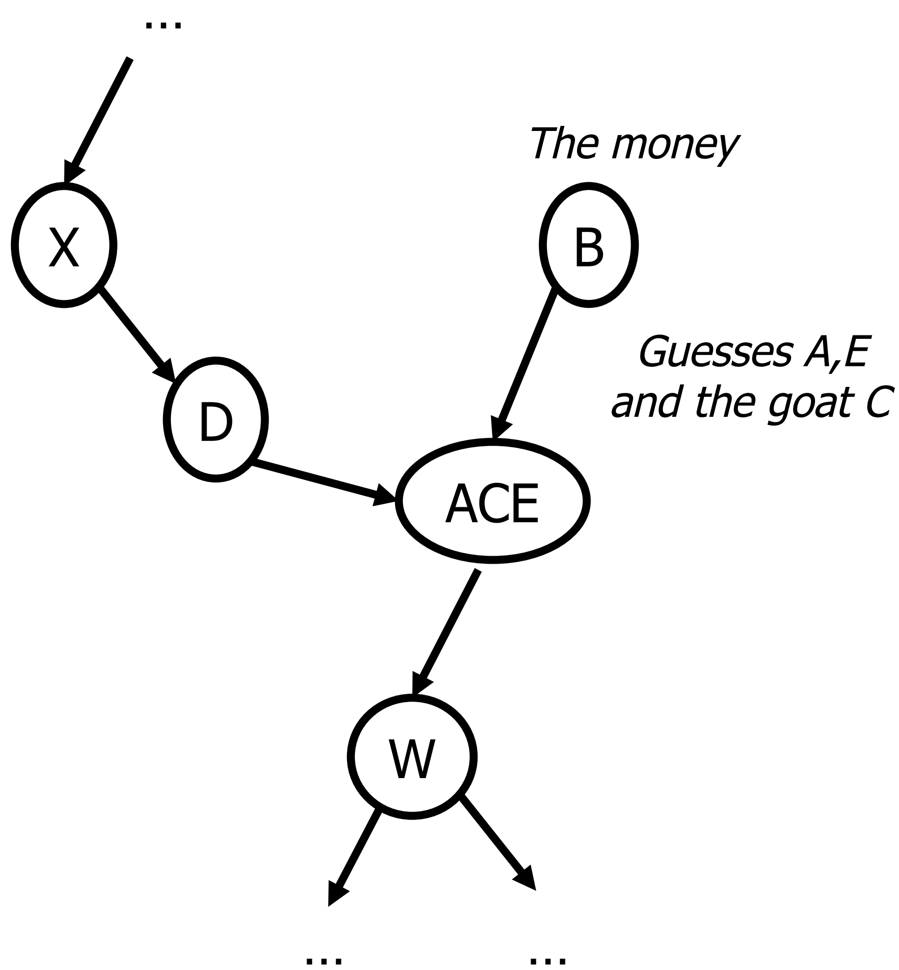
B	D	ACE	P(ACE|B,D)
0	0	1,1,1	…
0	0	1,1,2	…
…	…	…	…
0	0	3,3,3	…
0	1	1,1,1	…
…	…	…	…

Small tree-width graphs

Not a polytree



A polytree

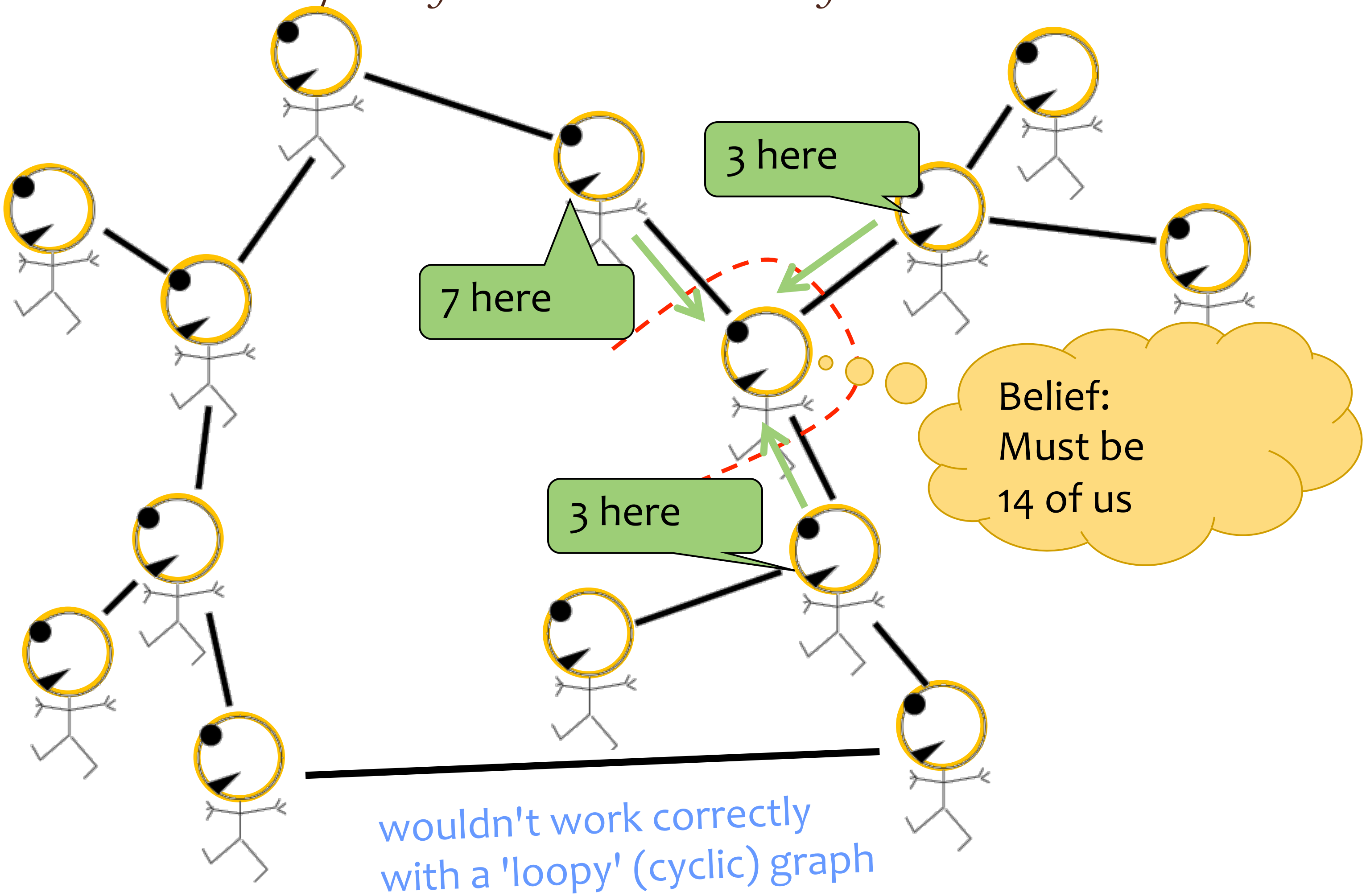


More on message passing/BP

- BP can be extended to other graphical models
  - Markov networks, if they are polytrees
  - “Factor graphs”
- BP for non-polytrees:
  - Small tree-width graphs:
    - convert to polytrees and run normal BP
  - Loopy BP

Great Ideas in ML: Message Passing

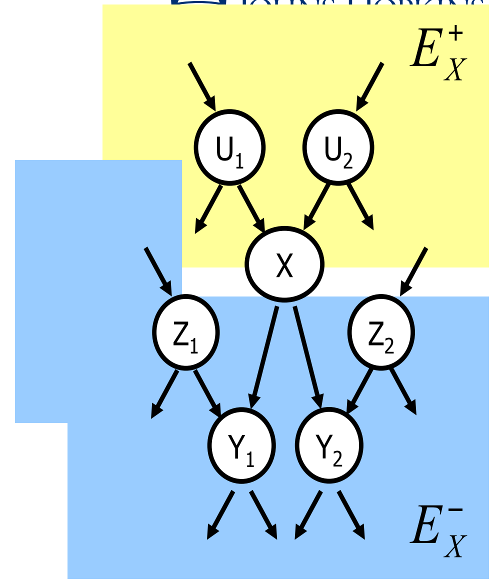
Each soldier receives reports from all branches of tree



adapted from MacKay (2003) textbook

Thanks Matt Gormley

Recap: Loopy BP



- Top-level algorithm
  - Initialize every X's messages P(X|E+) and P(E-|X) to vectors of all 1's.
  - Repeat:
    - Have every X send its children/parents messages
      - Which will be incorrect at first
    - For every X, update P(X|E+) and P(E-|X) based on last messages
      - Which will be incorrect at first
  - In a tree this will eventually converge to the right values
  - In a graph if *might* converge
    - Non-trivial to predict when and if but…
    - it's often a good approximation