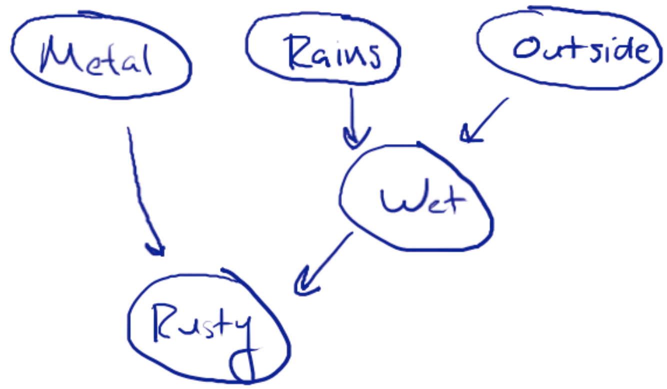
Review: probability

- Covariance, correlation
  - relationship to independence
- Law of iterated expectations
- Bayes Rule
- Examples: emacsitis, weighted dice
- Model learning

Review: graphical models

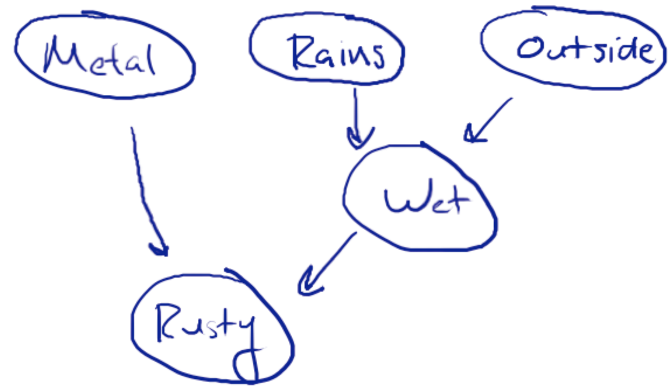


- Bayes net = DAG + CPT
- Factored representation of distribution
  - fewer parameters
- Inference: showed Metal & Outside independent for rusty-robot network

Independence

independent root

- Showed M ⊥ O



- Any other independences?

$M \perp Ra \quad Ra \perp O \quad M \perp W$

- Didn't use CPTs !!
  - independences depend only on DAG
- May also be "accidental" independences

depend on CPTs

$P(W \mid Ra, O)$

Ra	O	
T	T	0.3 → 0.30001
T	F	0.3
F	T	0.3
F	F	0.3

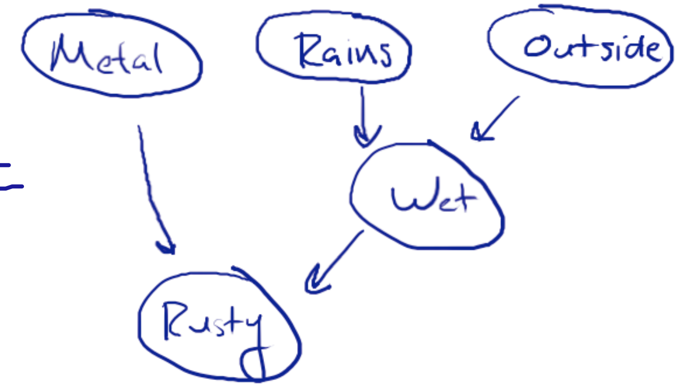
indep

depend

Conditional independence

- How about O, Ru? O $\not\perp$ Ru

$W=F$



- Suppose we know we're not wet
- P(M, Ra, O, W, Ru) =

  P(M) P(Ra) P(O) P(W|Ra,O) P(Ru|M,W)

- Condition on W=F, find marginal of O, Ru

$$P(O,Ru) = \sum_{M\in T,F} \sum_{Ra\in T,F} P(M)\,P(Ra)\,P(O)\,P(W=F\mid Ra,O)\,P(Ru\mid M,W=F) \,/\, P(W=F)$$

$$= \left[\sum_{Ra} P(Ra)\,P(O)\,P(W=F\mid Ra,O)\right]\left[\sum_{M} P(M)\,P(Ru\mid M,W=F) \,/\, P(W=F)\right]$$

$O \perp Ru \mid W=F$ conditional independence

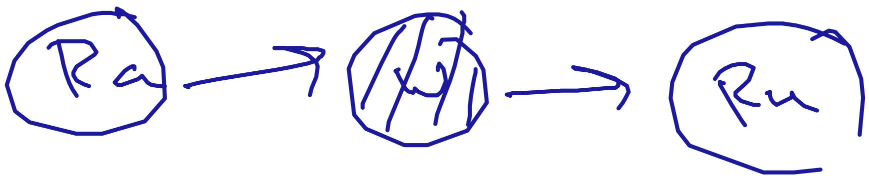
Conditional independence

- This is generally true
  - conditioning on evidence can make or break independences
  - many (conditional) independences can be derived from graph structure alone
  - “accidental” ones are considered less interesting → except "context specific"

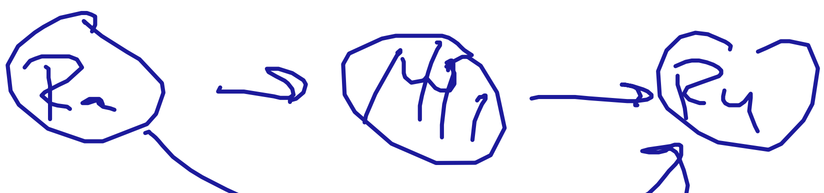
Graphical tests for independence

- We derived (conditional) independence by looking for factorizations
- It turns out there is a purely graphical test
  - this was one of the key contributions of Bayes nets
- Before we get there, a few more examples

Blocking



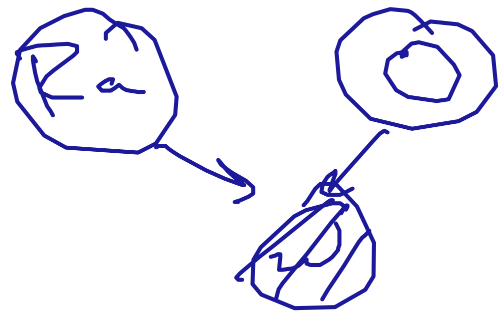
$Ra \perp Ru \mid W$



$Ra \not\perp Ru \mid W$

- Shaded = observed (by convention)

Explaining away

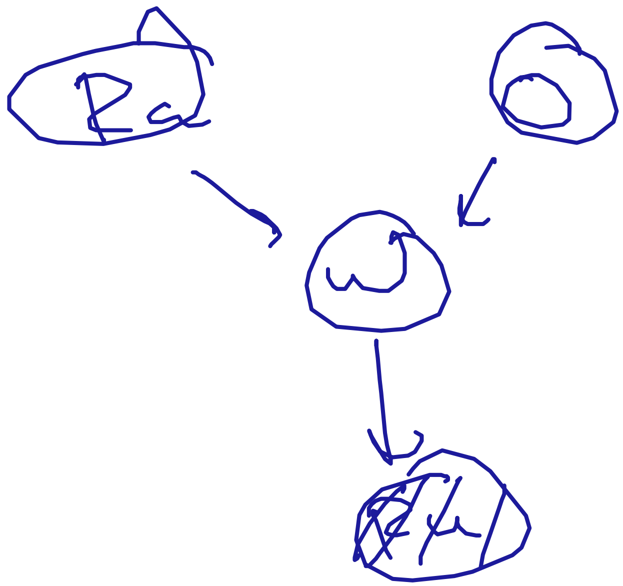


$Ra \perp O$

$Ra \not\perp O \mid W$

- Intuitively: know W=F , $Ra \xRightarrow{\text{probably}}$ not O

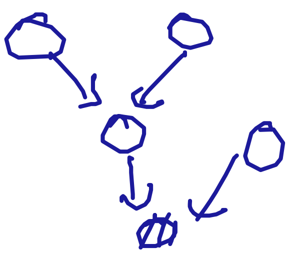
Son of explaining away



$Ra \perp O$

$R \not\perp O \mid Ru$

$R \not\perp W \mid Ru$



works same way

d-separation

- General graphical test: "d-separation"
  - d = dependence
- $X \perp Y \mid Z$ when there are no ***active paths*** between X and Y
- Active paths from X to Y (W ***outside*** conditioning set):

$X \rightarrow W \rightarrow Y$ Z in conditioning set

$X \leftarrow W \leftarrow Y$

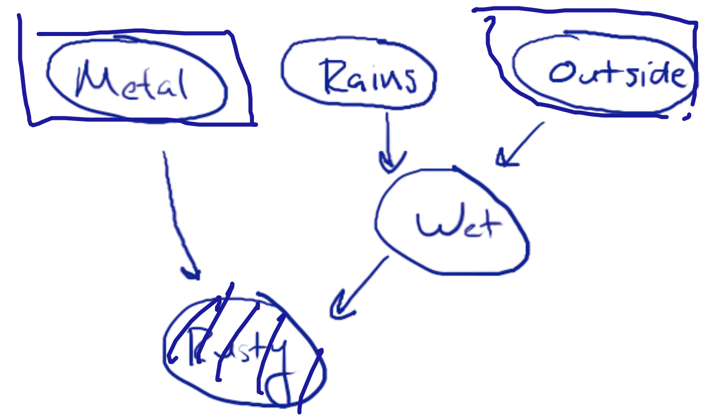
$X \leftarrow W \rightarrow Y$ "common cause"

$X \rightarrow Z \leftarrow Y$

$X \rightarrow W \leftarrow Y$

$\hookrightarrow \cdots \rightarrow Z$

Longer paths



- Node is active if:

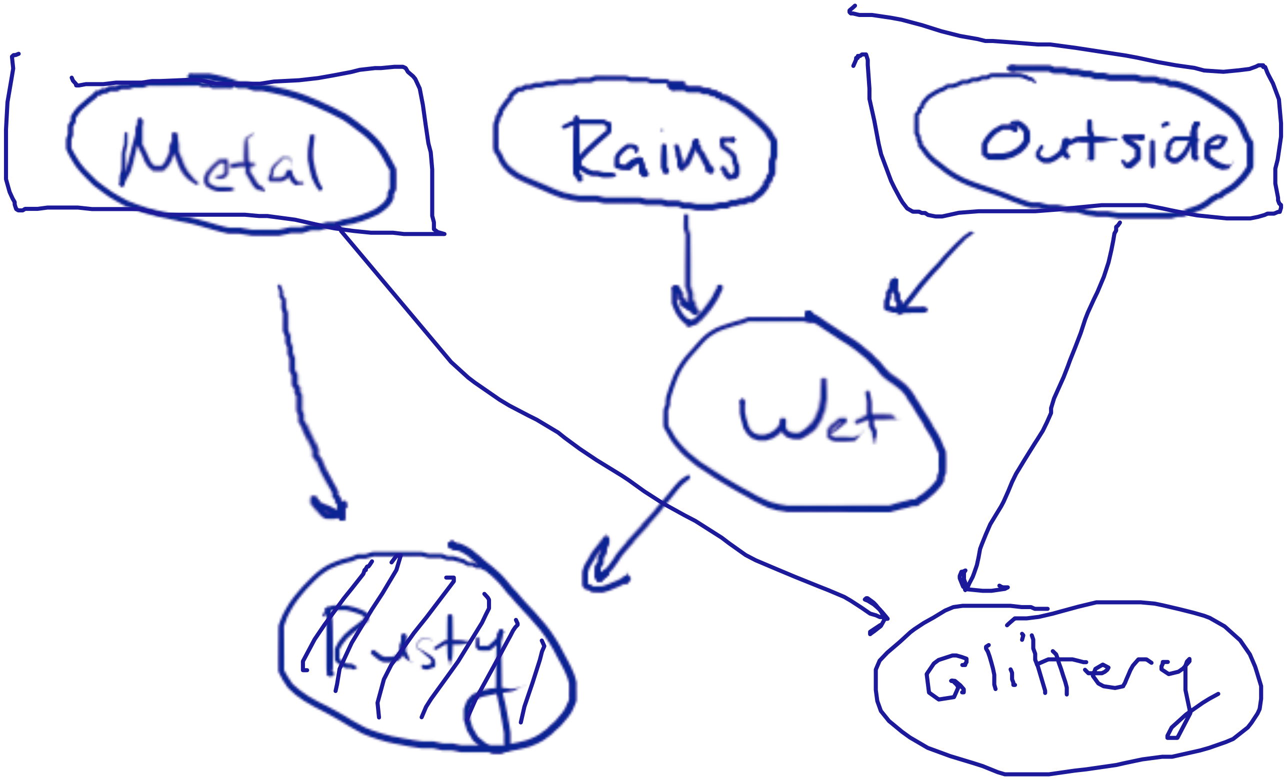
  unshaded and $\rightarrow\,\rightarrow$ $\leftarrow\,\leftarrow$ $\leftarrow\,\rightarrow$

  shaded or descendant shaded and $\rightarrow\,\leftarrow$

  and inactive o/w

- Path is active if all intermediate nodes are active

Another example

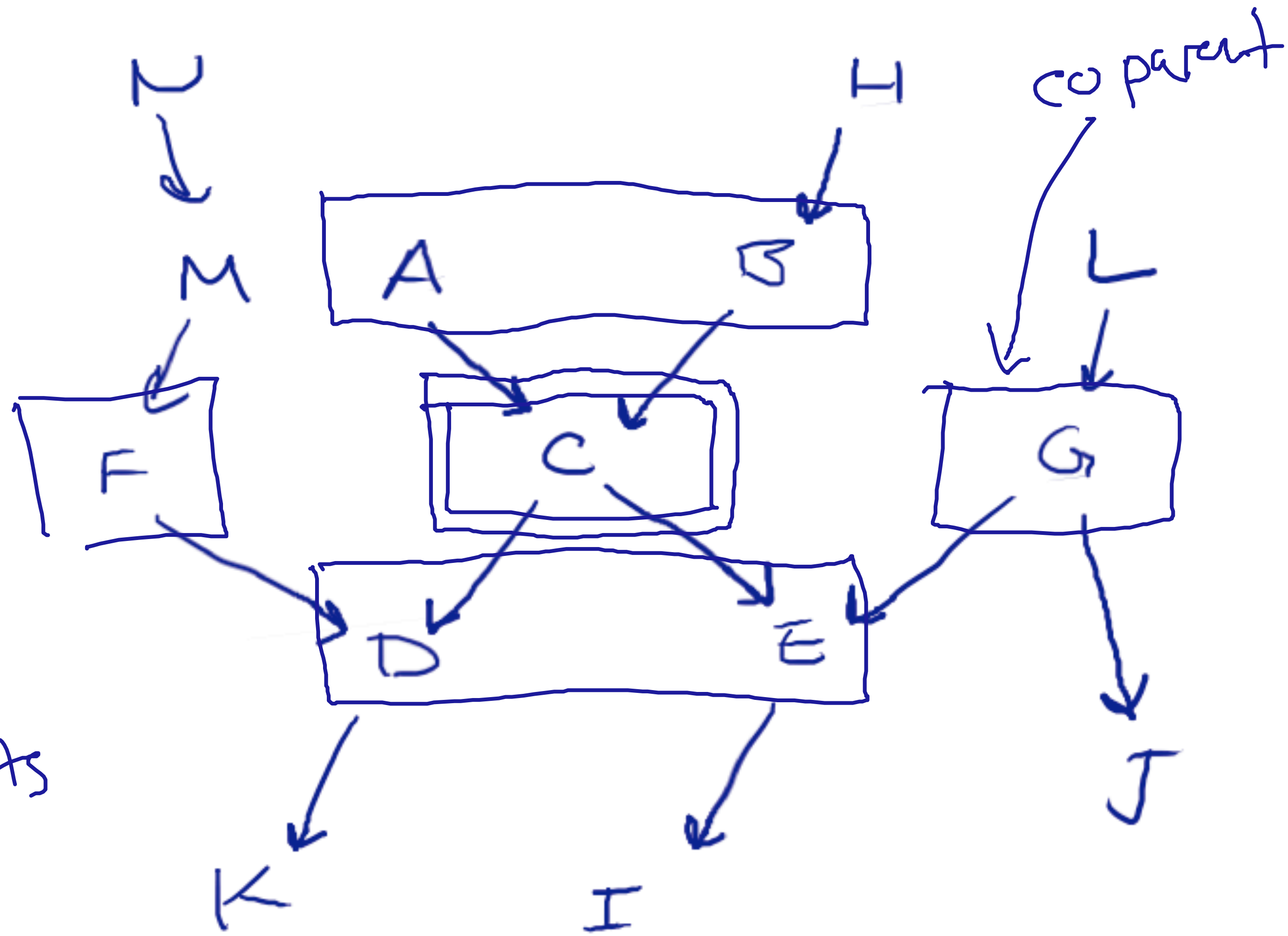


$M \not\perp O \mid Ru$

Markov blanket

Markov blanket of C = minimal set of observations to render C independent of rest of graph

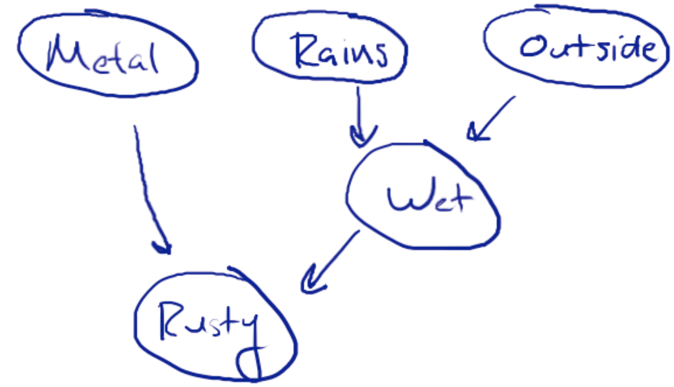
= parents, children and co-parents of C



Parameter

Learning Bayes nets

by Counting



P(M) = 3/5

P(Ra) = 2/5

P(O) = 4/5

P(W | Ra, O) =

TT 1/2

TF 0/0 ← prob #1 divide by 0

FT 1/2

FF 0/1 = 0 ← prob #2 extreme prob.

P(Ru | M,W) =

TT 1/2

TF

FT

FF

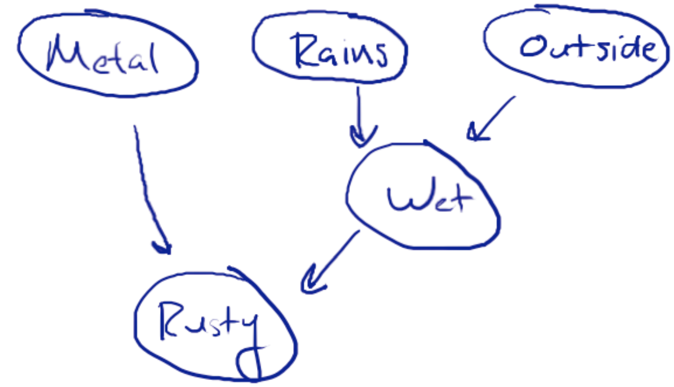
	M	Ra	O	W	Ru
obs 1	T	F	T	T	F
ob 2	T	T	T	T	T
3	F	T	T	F	F
4	T	F	F	F	T
5	F	F	T	F	T

no missing data

Parameter learning by

Laplace smoothing

asyptoticclly unbiased



$P(M) = 4/7 = \frac{3+1}{3+1+2+1}$

$P(Ra) = 3/7$

$P(O) = 5/7$

$P(W \mid Ra, O) =$

TT $= 2/4 = 1/2$

TF $= 1/2$

FT

FF $= 2/3$

$P(Ru \mid M, W) =$

M	Ra	O	W	Ru
T	F	T	T	F
T	T	T	T	T
F	T	T	F	F
T	F	F	F	T
F	F	T	F	T

Advantages of Laplace

- No division by zero
- No extreme probabilities
    - No near-extreme probabilities unless lots of evidence

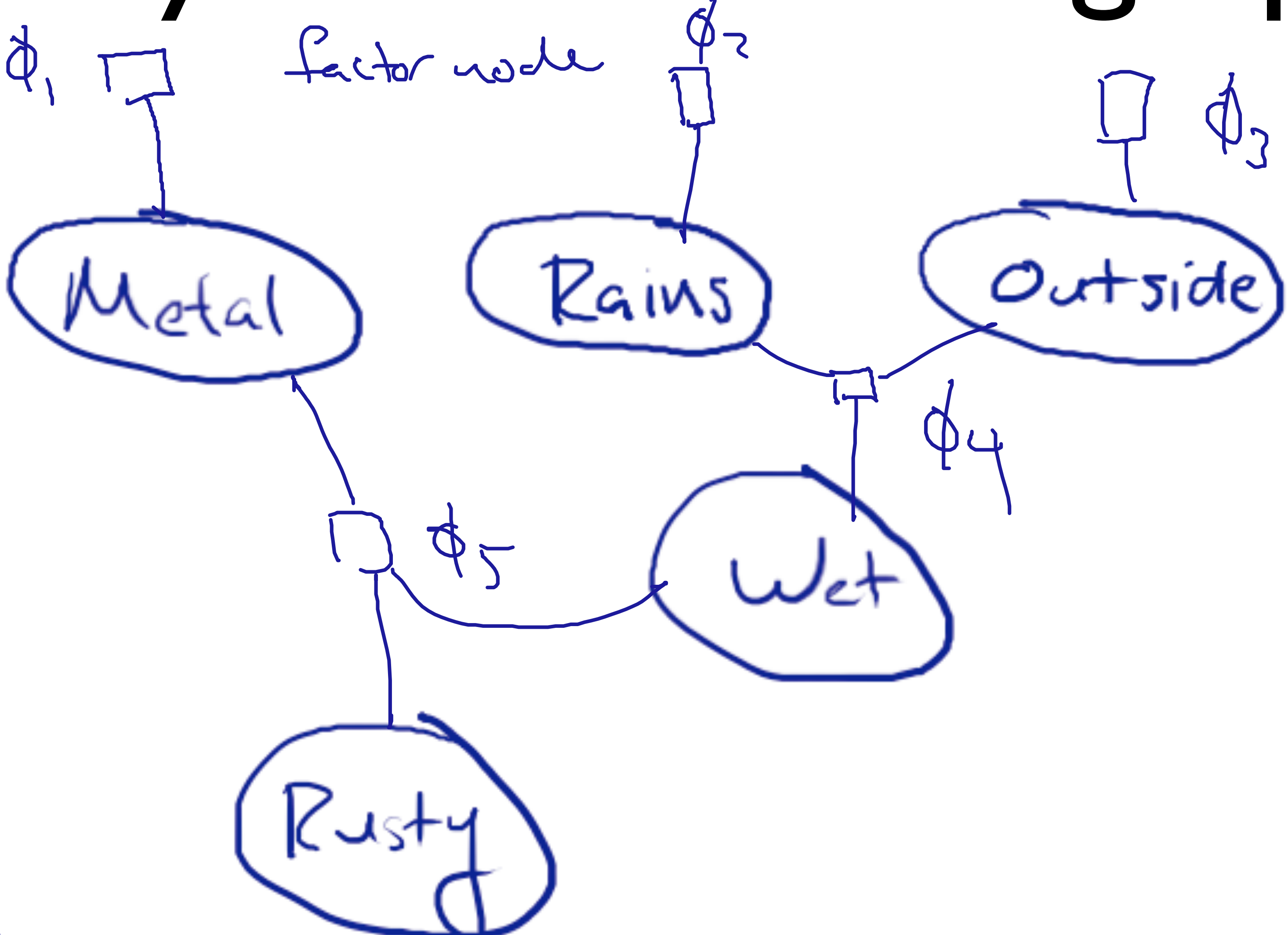
Limitations of counting and Laplace smoothing

- Work ***only*** when all variables are observed in all examples
- If there are ***hidden*** or ***latent*** variables, more complicated algorithm—we'll cover a related method later in course
  - or just use a toolbox!

Factor graphs

- Another common type of graphical model
- Uses ***undirected, bipartite*** graph instead of DAG

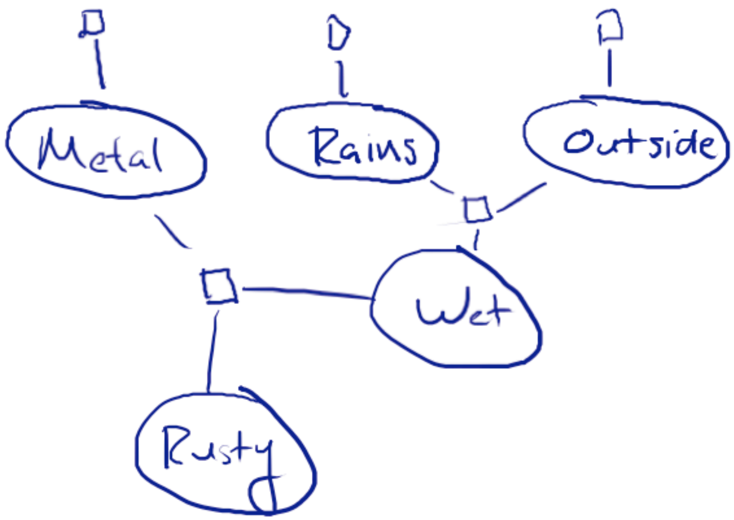
Rusty robot: factor graph

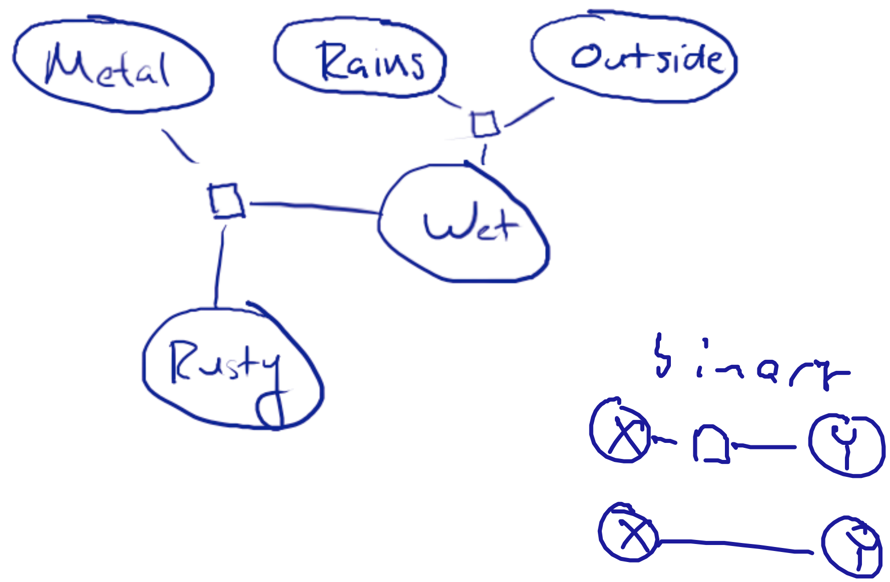


$\phi_1(M)$

$$\underbrace{P(M)}\ \underbrace{P(Ra)}_{\phi_2(Ra)}\ \underbrace{P(O)}_{\phi_3}\ \underbrace{P(W|Ra,O)}_{\phi_4(R,O,W)}\ \underbrace{P(Ru|M,W)}_{\phi_5}$$

Convention





- Don’t need to show unary factors
- Why? They don’t affect algorithms below.

Non-CPT factors

- Just saw: easy to convert Bayes net → factor graph
- In general, factors need not be CPTs: any nonnegative #s allowed

$\phi_1(x, y) =$

T T	5
T F	1
F T	1
F F	5

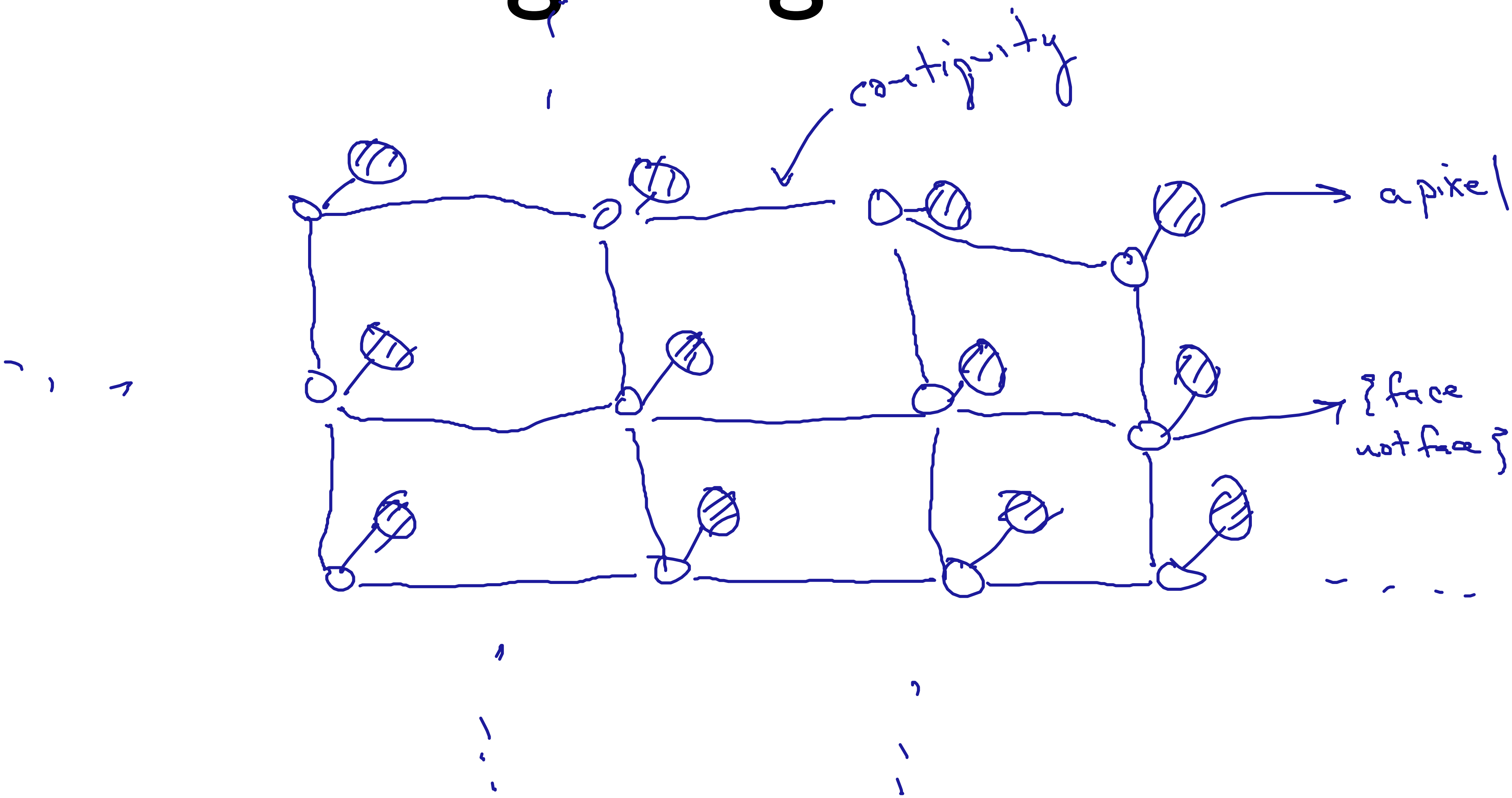
- In general, P(A, B, …) = $\dfrac{\tilde{P}(A, B, \ldots)}{Z}$

$\tilde{P}(A, B, \ldots) = \prod_{i \in \text{factor nodes}} \phi_i(\text{nbr}(i))$

nbr(i) = neighbor set

- Z = $\sum_{A \in \text{rng}(A)} \sum_{B \in \text{rng}(B)} \ldots \sum \tilde{P}(A, B, \ldots)$

Ex: image segmentation



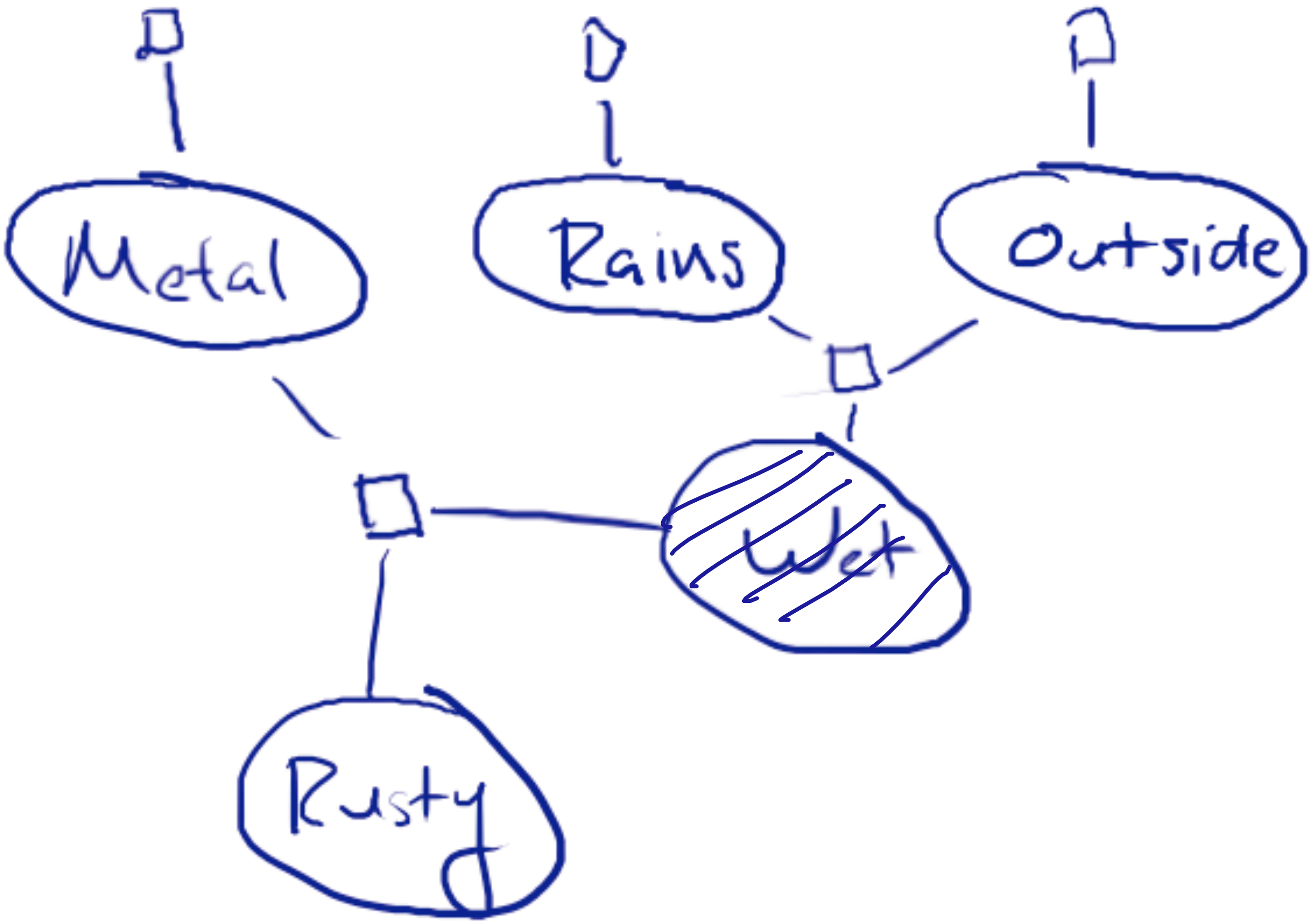
Factor graph → Bayes net

- Conversion possible, but much more involved
  - Each representation can handle ***any*** distribution
- Without adding nodes: # P - complete
  → we think exp time
- Adding nodes:
  poly-time
  somewhat complicated

Independence

- Just like Bayes nets, there are graphical tests for independence and conditional independence
- Simpler, though:
  - Cover up all observed nodes
  - Look for a path

Independence example



$$M \not\perp O$$

$$M \perp O \mid W$$

Modeling independence

- Take a Bayes net, list the (conditional) independences
- Convert to a factor graph, list the (conditional) independences
- Are they the same list? No!
- What happened?

accidental indep. are different

Inference

- We gave an example of inference in a Bayes net, but not a general algorithm
- Reason: general algorithm uses factor-graph representation
- Steps: instantiate evidence, eliminate nuisance nodes, answer query