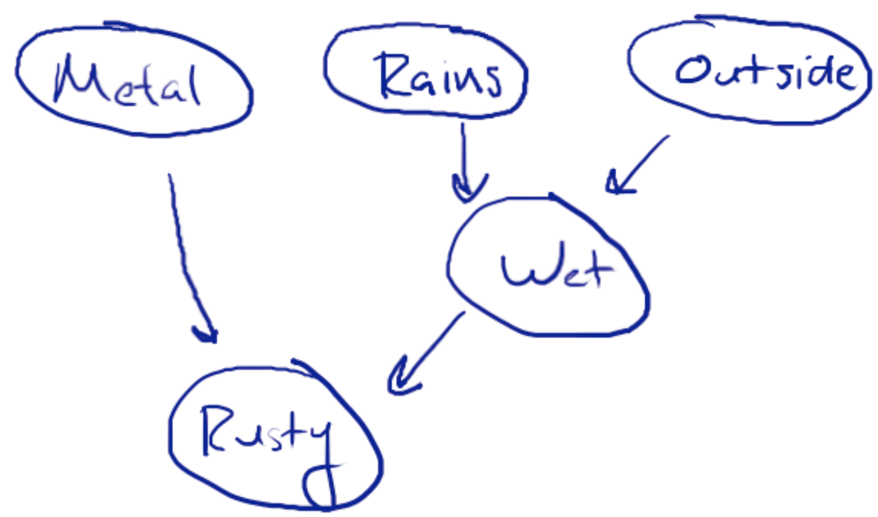
Review: probability

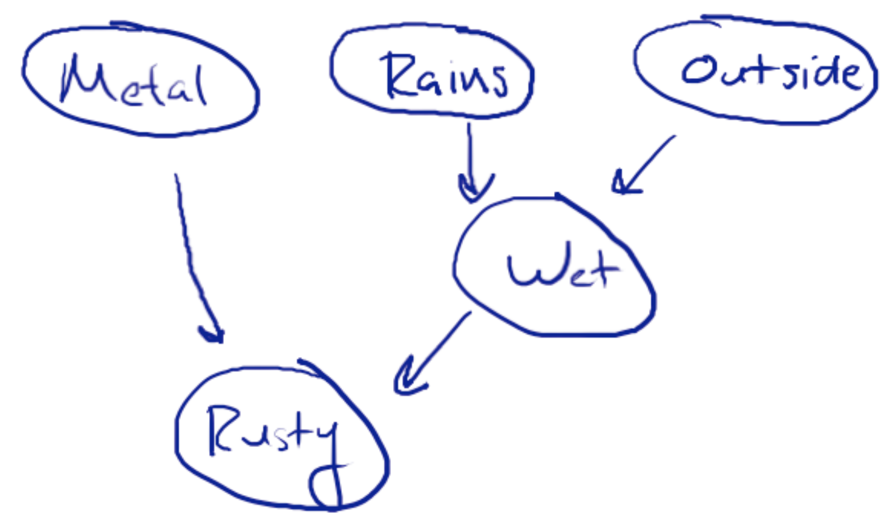
- Covariance, correlation
  - relationship to independence
- Law of iterated expectations
- Bayes Rule
- Examples: emacsitis, weighted dice
- Model learning

Review: graphical models



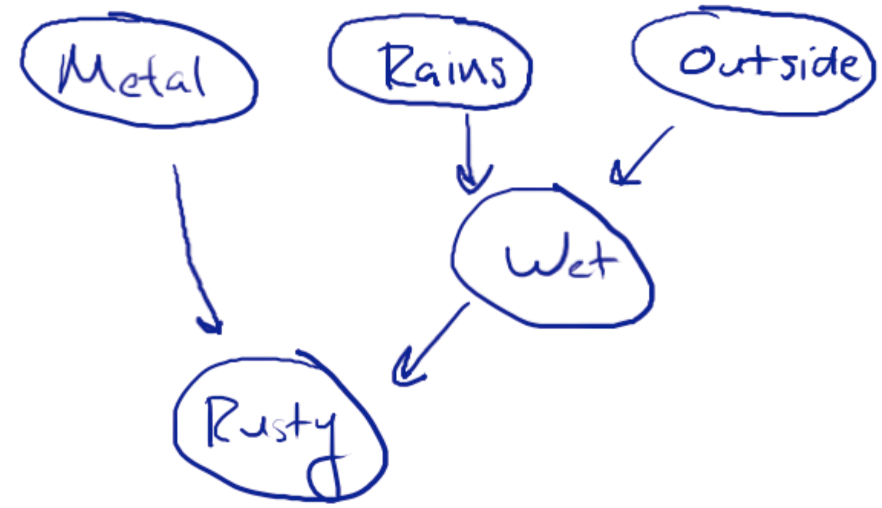
- Bayes net = DAG + CPT
- Factored representation of distribution
  - fewer parameters
- Inference: showed Metal & Outside independent for rusty-robot network

Independence



- Showed M $\perp$ O
- Any other independences?
- Didn't use
    - independences depend only on
- May also be "accidental" independences

Conditional independence



- How about O, Ru?   O   Ru
- 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

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

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

- Shaded = observed (by convention)

Explaining away

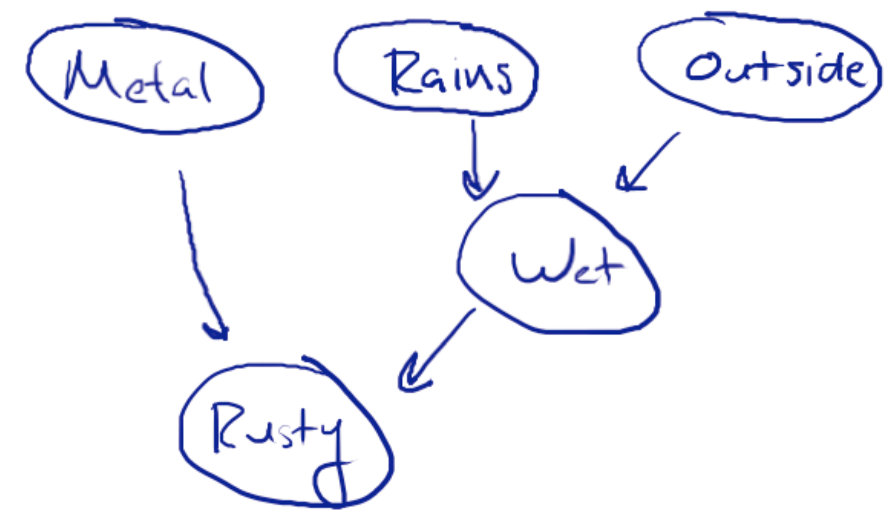
- Intuitively:

Son of explaining away

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 (W ***outside*** conditioning set):

Longer paths

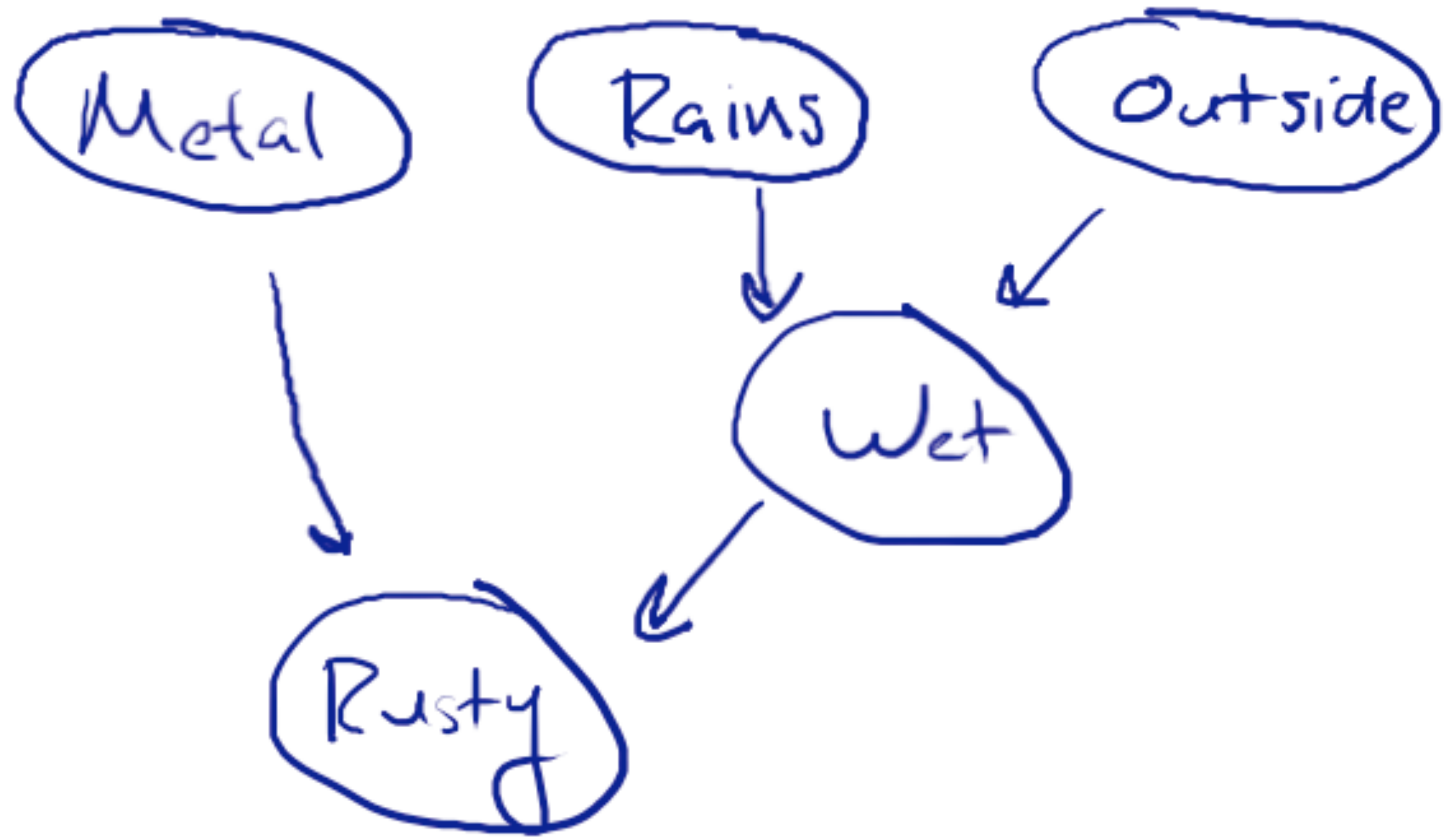


- Node is active if:

  and inactive o/w

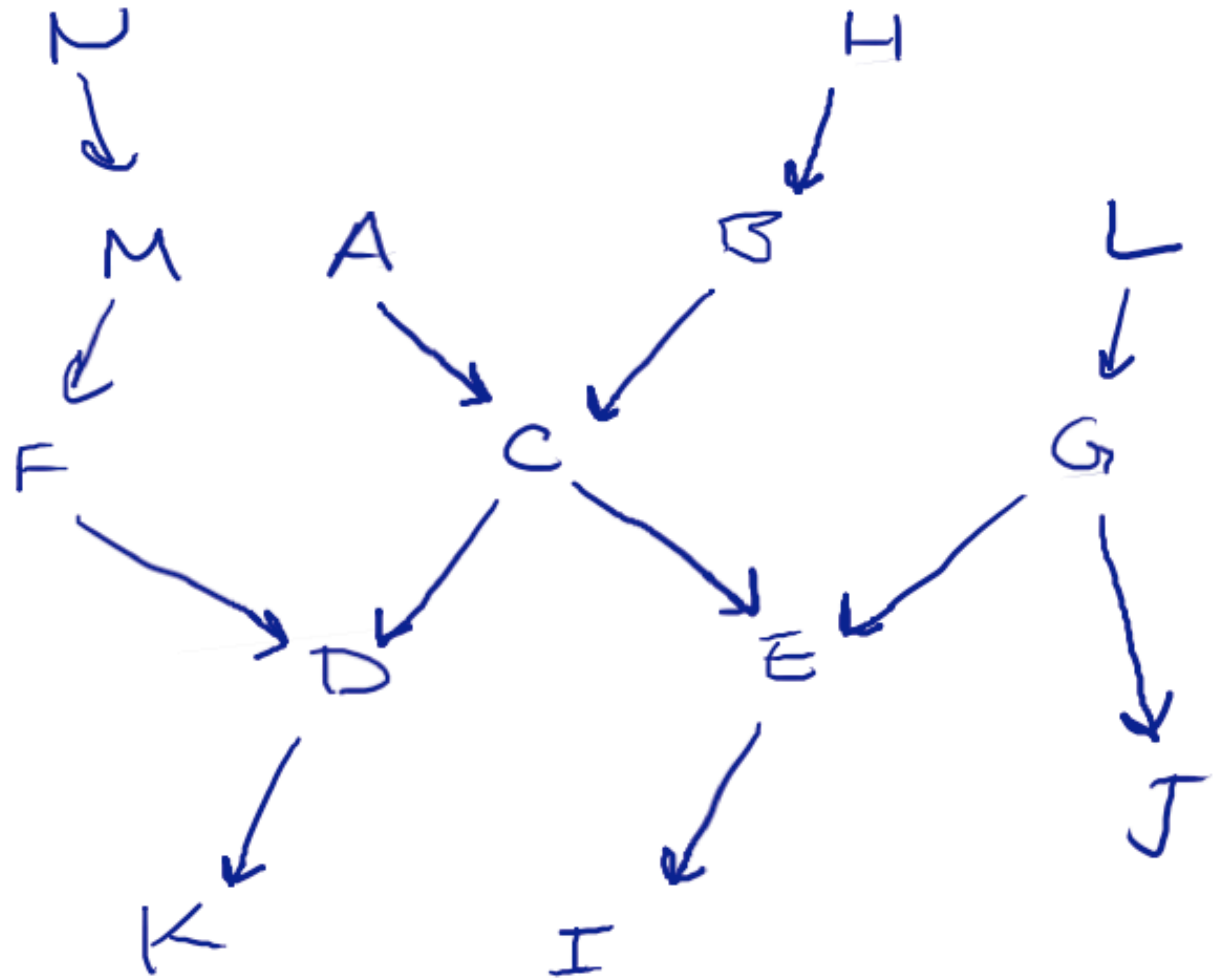
- Path is active if intermediate nodes are

Another example



Markov blanket

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



Learning Bayes nets

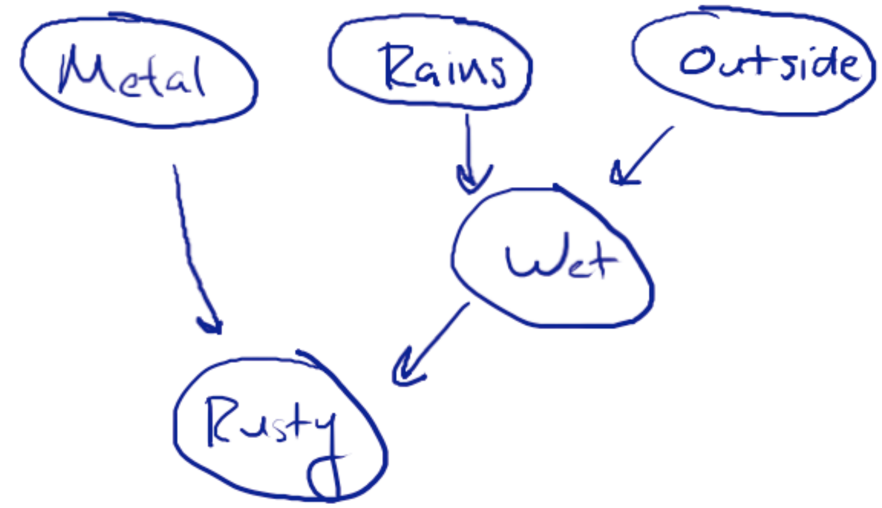
P(M) =

P(Ra) =

P(O) =

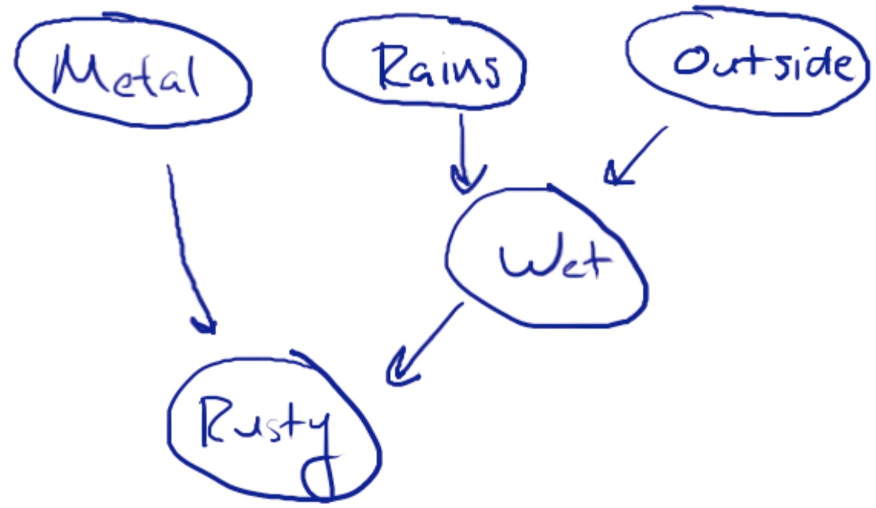
P(W | Ra, O) =

P(Ru | 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

Laplace smoothing



P(M) =

P(Ra) =

P(O) =

P(W | Ra, O) =

P(Ru | 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!

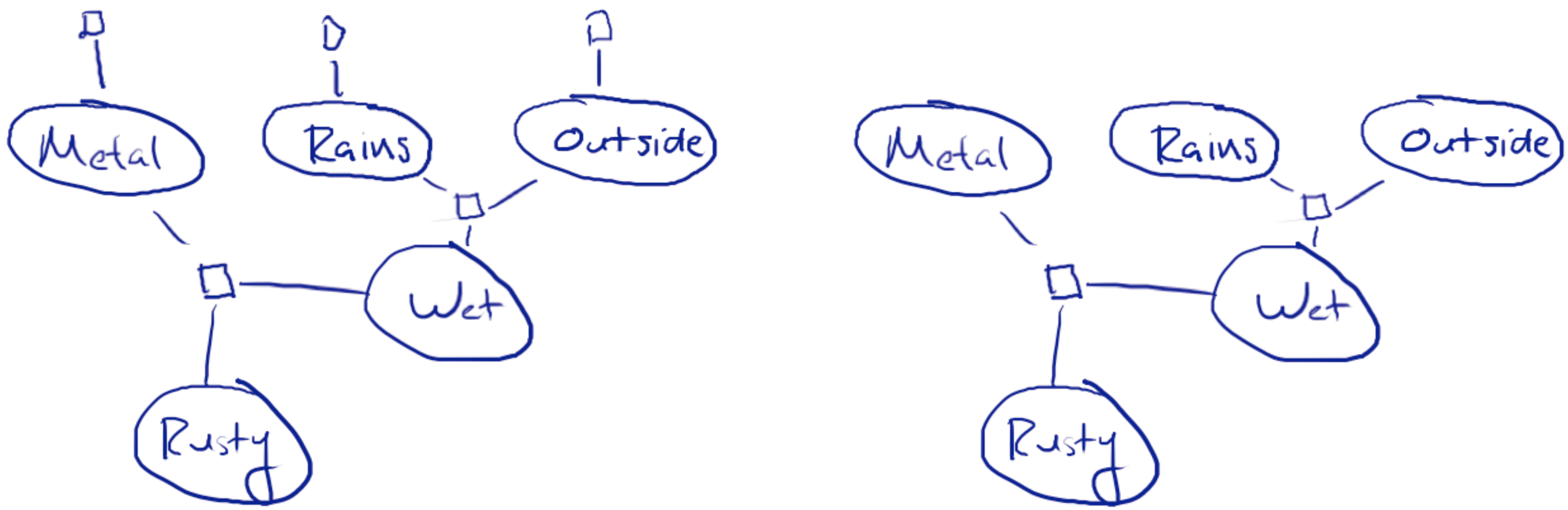
Factor graphs

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

Rusty robot: factor graph

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

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
- In general, P(A, B, …) =
- Z =

Ex: image segmentation

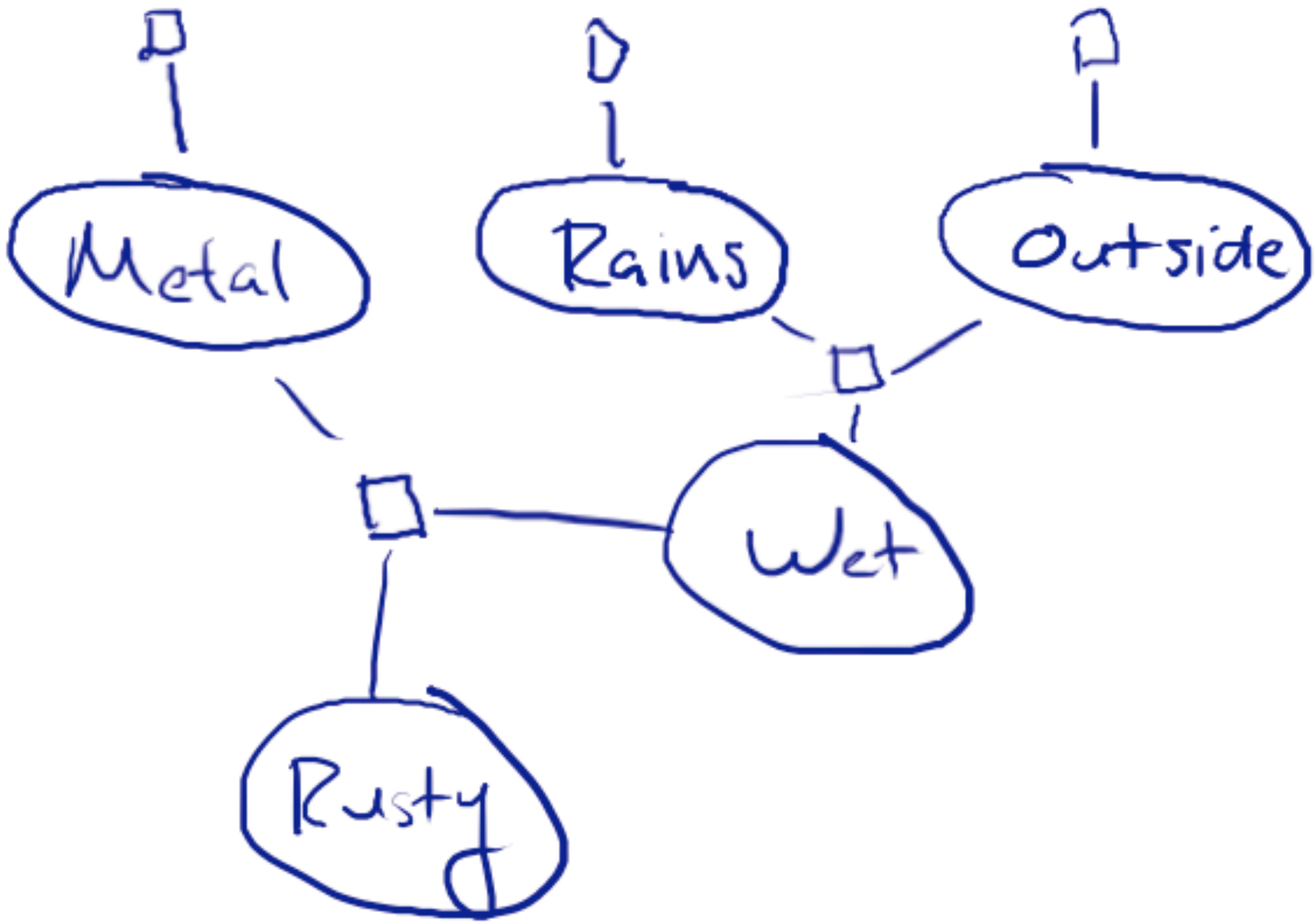
Factor graph → Bayes net

- Conversion possible, but more involved
  - Each representation can handle ***any*** distribution
- Without adding nodes:
- Adding nodes:

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



Modeling independence

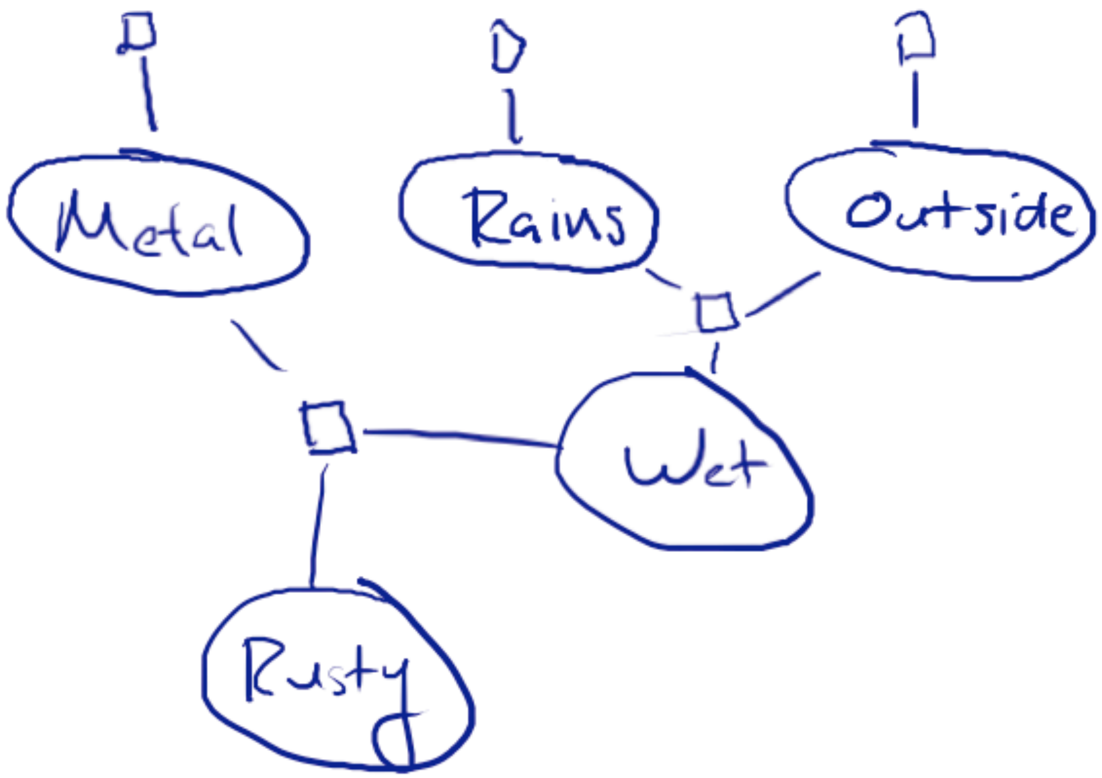
- Take a Bayes net, list the (conditional) independences
- Convert to a factor graph, list the (conditional) independences
- Are they the same list?
- What happened?

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

Inference

$$P(M, Ra, O, W, Ru) = \phi_1(M)\,\phi_2(Ra)\,\phi_3(O)\,\phi_4(Ra, O, W)\,\phi_5(M, W, Ru) / Z$$



$\phi_1(M) =$

T	0.9
F	0.1

$\phi_2(Ra) =$

T	0.7
F	0.3

$\phi_3(O) =$

T	0.2
F	0.8

$\phi_4(Ra, O, W) =$

TTT	0.9
TTF	0.1
TFT	0.1
TFF	0.9
FTT	0.1
FTF	0.9
FFT	0.1
FFF	0.9

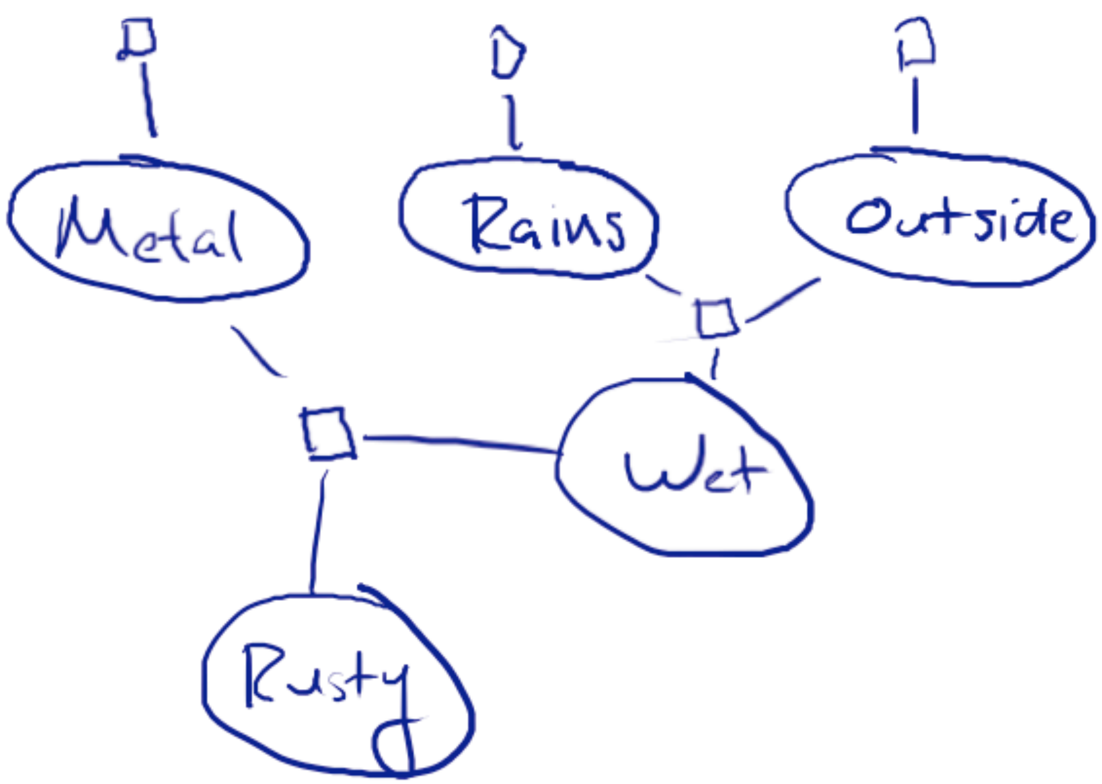
$\phi_5(M, W, Ru) =$

TTT	0.8
TTF	0.2
TFT	0.1
TFF	0.9
FTT	0
FTF	1
FFT	0
FFF	1

- Typical Q: given Ra=F, Ru=T, what is P(W)?

Incorporate evidence

$$P(M, Ra, O, W, Ru) = \phi_1(M)\,\phi_2(Ra)\,\phi_3(O)\,\phi_4(Ra, O, W)\,\phi_5(M, W, Ru) / Z$$



$\phi_1(M) =$

T	0.9
F	0.1

$\phi_2(Ra) =$

T	0.7
F	0.3

$\phi_3(O) =$

T	0.2
F	0.8

$\phi_4(Ra, O, W) =$

TTT	0.9
TTF	0.1
TFT	0.1
TFF	0.9
FTT	0.1
FTF	0.9
FFT	0.1
FFF	0.9

$\phi_5(M, W, Ru) =$

TTT	0.8
TTF	0.2
TFT	0.1
TFF	0.9
FTT	0
FTF	1
FFT	0
FFF	1

Condition on Ra=F, Ru=T

Eliminate nuisance nodes

$$P(M, \cancel{Ra}, O, W, \cancel{Ru}) = \phi_1(M)\,\cancel{\phi_2(Ra)}\,\phi_3(O)\,\phi_4(\cancel{Ra}, O, W)\,\phi_5(M, W, \cancel{Ru})\,/\,Z$$

- Remaining nodes: M, O, W
- Query: P(W)
- So, O&M are nuisance—marginalize away
- Marginal =

Elimination order

$$\sum_{M} \sum_{O} \phi_1(M)\,\phi_3(O)\,\phi_4(O,W)\,\phi_5(M,W)\,/\,Z$$

- Sum out the nuisance variables in turn
- Can do it in any order, but some orders may be easier than others
- Let's do O, then M

$\phi_3(O) =$

T	0.2
F	0.8

$\phi_4(O,W) =$

TT	0.1
TF	0.9
FT	0.1
FF	0.9

One last elimination

$\phi_1(M) =$

T	0.9
F	0.1

$\phi_6(W) =$

T	0.1
F	0.9

$\phi_5(M, W) =$

T	T	0.8
T	F	0.1
F	T	0
F	F	0