Bayesian Networks
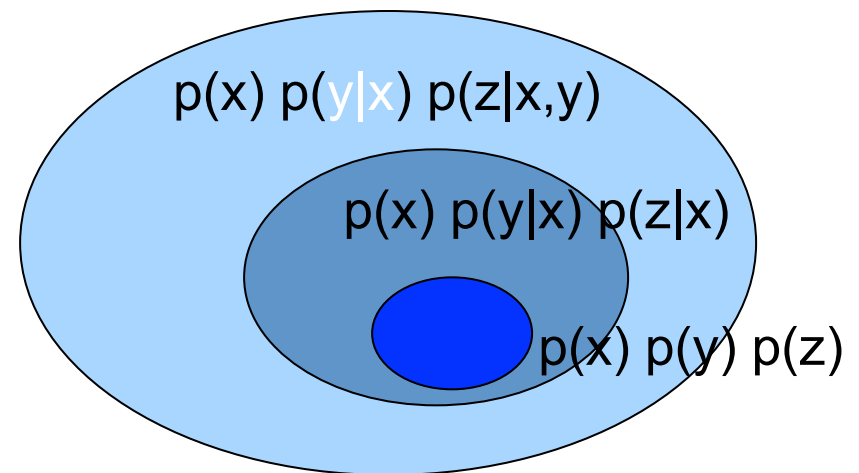
Introduction to Graphical Models

Prof. Alexander Ihler

Conditional independence

- Recall: chain rule of probability
  - p(x,y,z) = p(x) p(y|x) p(z|x,y)

- *Some* of these models will be conditionally independent
  - e.g.,   p(x,y,z) = p(x) p(y|x) p(z|x)

- *Some* models may have even *more* independence
  - E.g.,   p(x,y,z) = p(x) p(y) p(z)

p(x) p(y|x) p(z|x,y)
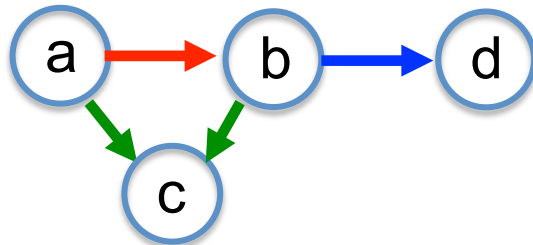
p(x) p(y|x) p(z|x)

p(x) p(y) p(z)

Bayesian networks

- Directed graphical model
- Nodes associated with variables
- "Draw" independence in conditional probability expansion
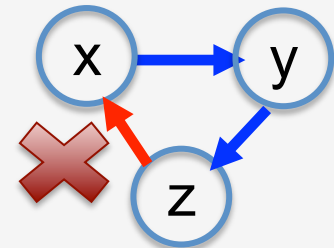 - Parents in graph are the RHS of conditional

- Ex: $p(x, y, z) = p(x)\ p(y \mid x)\ p(z \mid y)$
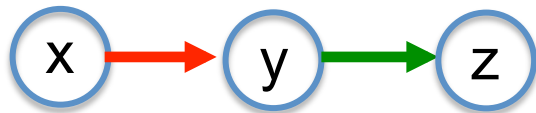
- Ex: $p(a, b, c, d) = p(a)\ p(b \mid a)\ p(c \mid a, b)\ p(d \mid b)$

*Graph must be **acyclic***

Corresponds to an order over the variables (chain rule)

Example

- Consider the following 5 binary variables:
  - B = a burglary occurs at your house
  - E = an earthquake occurs at your house
  - A = the alarm goes off
  - J = John calls to report the alarm
  - M = Mary calls to report the alarm

  - What is P(B | M, J) ? (for example)

  - We can use the full joint distribution to answer this question
    - Requires 2^5 = 32 probabilities

    - Can we use prior domain knowledge to come up with a Bayesian network that requires fewer probabilities?

Constructing a Bayesian network

- Order the variables in terms of causality (may be a partial order)
    - e.g., $\{\, E,\ B\,\} \longrightarrow \{\, A\,\} \longrightarrow \{\, J,\ M\,\}$

- Now, apply the chain rule, and simplify based on assumptions

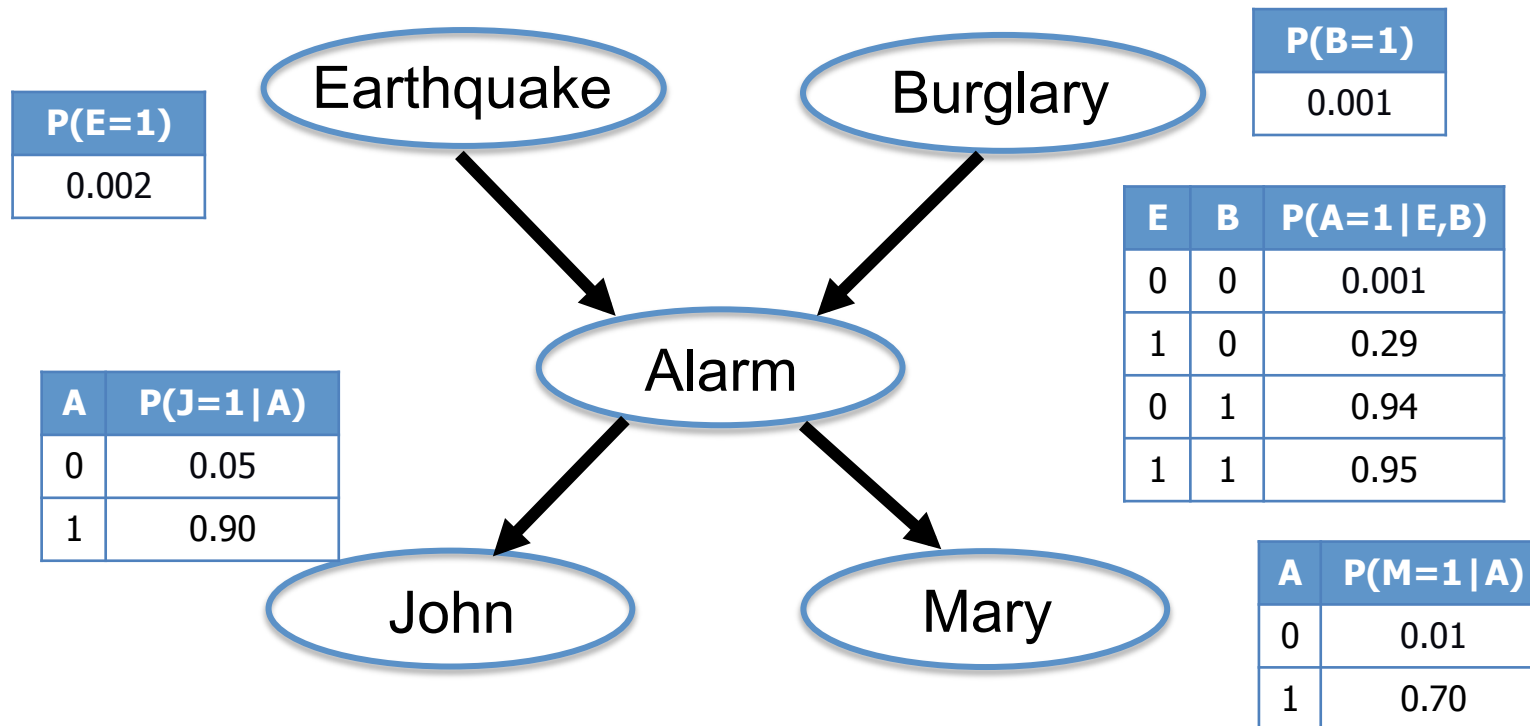$$p(J, M, A, E, B) = p(E, B)\ p(A\,|\,E, B)\ p(J, M\,|\,A, E, B)$$
$$= p(E)\ p(B)\ p(A\,|\,E, B)\ p(J, M\,|\,A)$$
$$= p(E)\ p(B)\ p(A\,|\,E, B)\ p(J\,|\,A)\ p(M\,|\,A)$$

- These assumptions are reflected in the
  graph structure of the Bayesian network
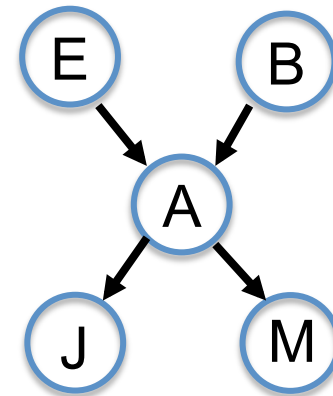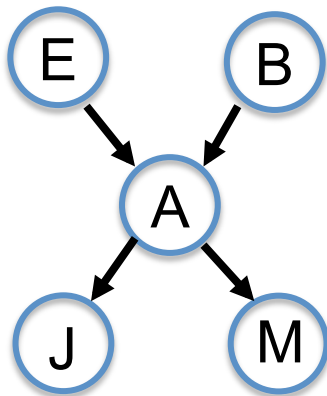
Constructing a Bayesian network

- Given $p(J, M, A, E, B) = p(E)\ p(B)\ p(A\,|\,E, B)\ p(J\,|\,A)\ p(M\,|\,A)$

- Define probabilities: 1   +   1   +   4   +   2   +   2

- Where do these come from?
  - Expert knowledge; estimate from data; some combination

P(E=1)
0.002

P(B=1)
0.001

E	B	P(A=1|E,B)
0	0	0.001
1	0	0.29
0	1	0.94
1	1	0.95

A	P(J=1|A)
0	0.05
1	0.90

A	P(M=1|A)
0	0.01
1	0.70

Earthquake   Burglary

Alarm

John   Mary

Constructing a Bayesian network

- Joint distribution

Full joint distribution:
 2^5 = 32 probabilities

Structured distribution:
   specify 10 parameters

E	B	A	J	M	P(...)
0	0	0	0	0	.93674
1	0	0	0	0	.00133
0	1	0	0	0	.00006
1	1	0	0	0	.00000
0	0	1	0	0	.00003
1	0	1	0	0	.00002
0	1	1	0	0	.00003
1	1	1	0	0	.00000
0	0	0	1	0	.04930
1	0	0	1	0	.00007
0	1	0	1	0	.00000
1	1	0	1	0	.00000
0	0	1	1	0	.00027
1	0	1	1	0	.00016
0	1	1	1	0	.00025
1	1	1	1	0	.00000

E	B	A	J	M	P(...)
0	0	0	0	1	.00946
1	0	0	0	1	.00001
0	1	0	0	1	.00000
1	1	0	0	1	.00000
0	0	1	0	1	.00007
1	0	1	0	1	.00004
0	1	1	0	1	.00007
1	1	1	0	1	.00000
0	0	0	1	1	.00050
1	0	0	1		.00000
0	1	0	1	1	.00000
1	1	0	1	1	.00000
0	0	1	1	1	.00063
1	0	1	1	1	.00037
0	1	1	1	1	.00059
1	1	1	1	1	.00000

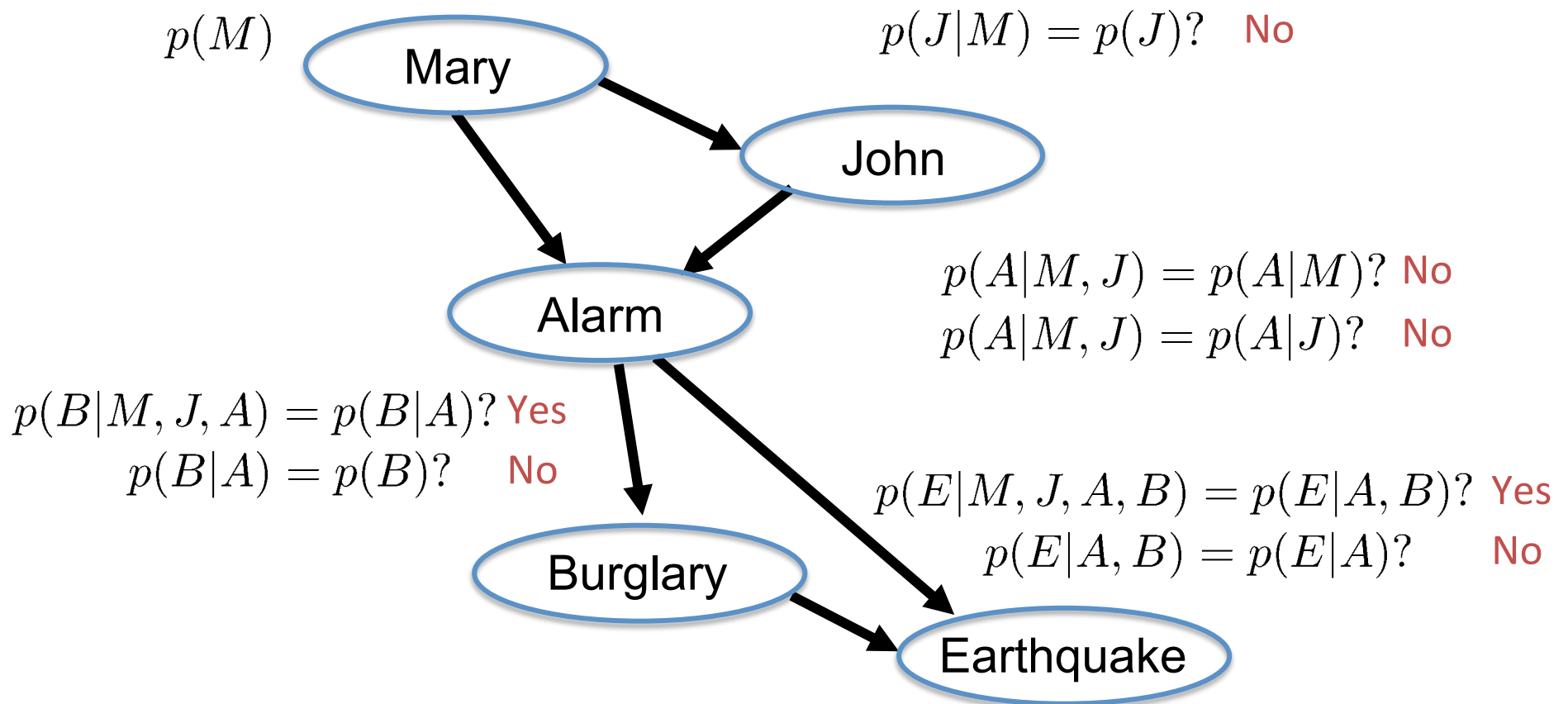# Alarm network [Beinlich et al., 1989]

The "alarm" network: 37 variables, 509 parameters (rather than $2^{37} = 10^{11}$!)
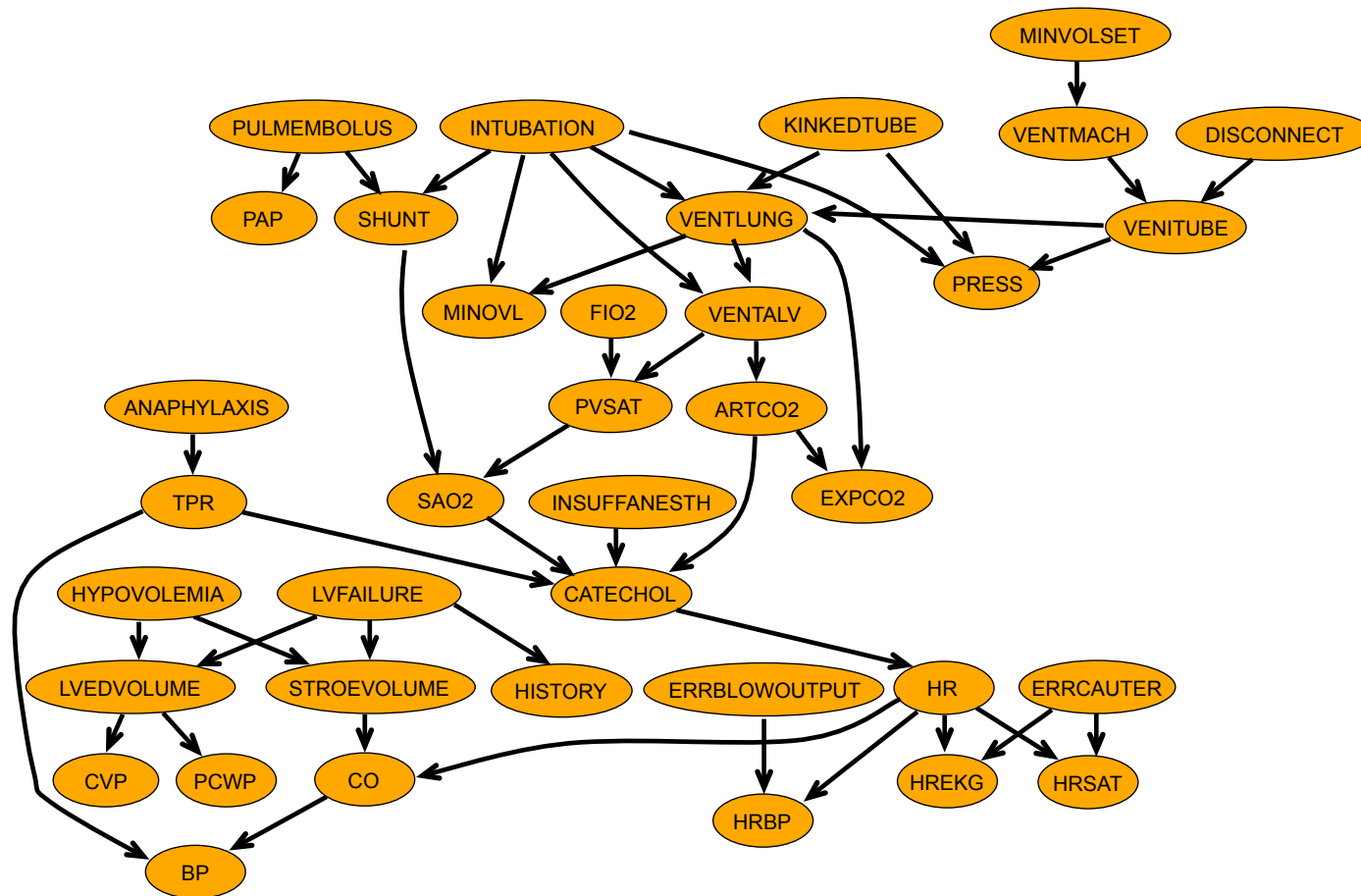
Network structure and ordering

- The network structure depends on the conditioning order
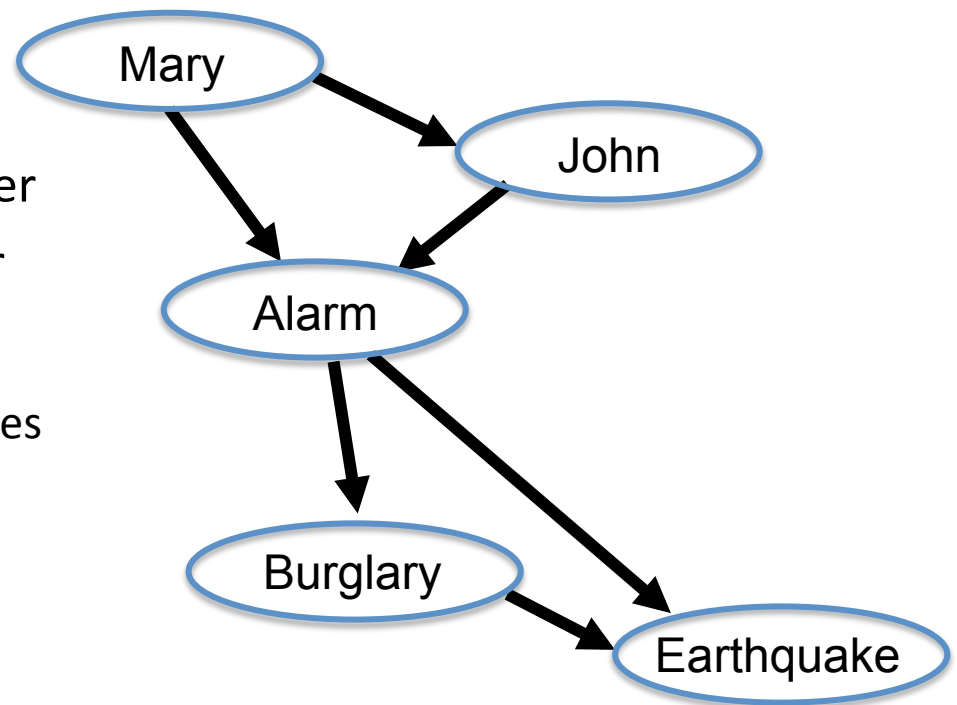 - Suppose we choose ordering M, J, A, B, E

$p(M)$ **Mary**

$p(J|M) = p(J)?$ No

John

$p(A|M, J) = p(A|M)?$ No
$p(A|M, J) = p(A|J)?$ No

Alarm

$p(B|M, J, A) = p(B|A)?$ Yes
$p(B|A) = p(B)?$ No

$p(E|M, J, A, B) = p(E|A, B)?$ Yes
$p(E|A, B) = p(E|A)?$ No

Burglary

Earthquake

Network structure and ordering

- The network structure depends on the conditioning order
  - Suppose we choose ordering M, J, A, B, E

- "Non-causal" ordering
  - Deciding independence is harder
  - Selecting probabilities is harder
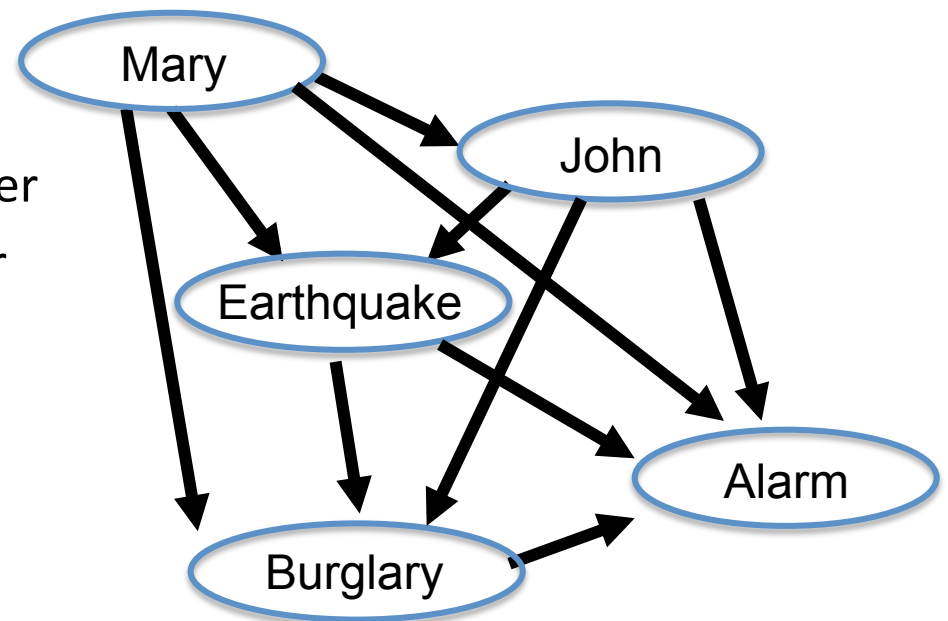 - Representation is less efficient

    1 + 2 + 4 + 2 + 4 = 13 probabilities

Network structure and ordering

- The network structure depends on the conditioning order
  - Suppose we choose ordering M, J, A, B, E

- "Non-causal" ordering
  - Deciding independence is harder
 - Selecting probabilities is harder
  - Representation is less efficient

  - Some orders may not reveal any independence!

$$p(J, M, A, E, B) = p(M)\, p(J|M)\, p(E|M, J)\, p(B|M, J, E)\, p(A|M, J, E, B)$$

Reasoning in Bayesian networks

- Suppose we observe J
  - Observing J makes A more likely
  - A being more likely makes B more likely

- Suppose we observe A
  - Makes M more likely

- Observe A and J?
  - J doesn't add anything to M
  - Observing A makes J, M independent

- How can we read independence directly from the graph?
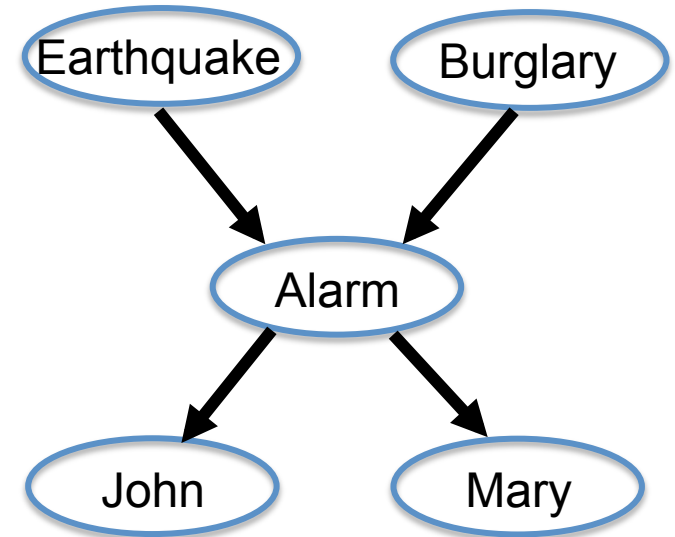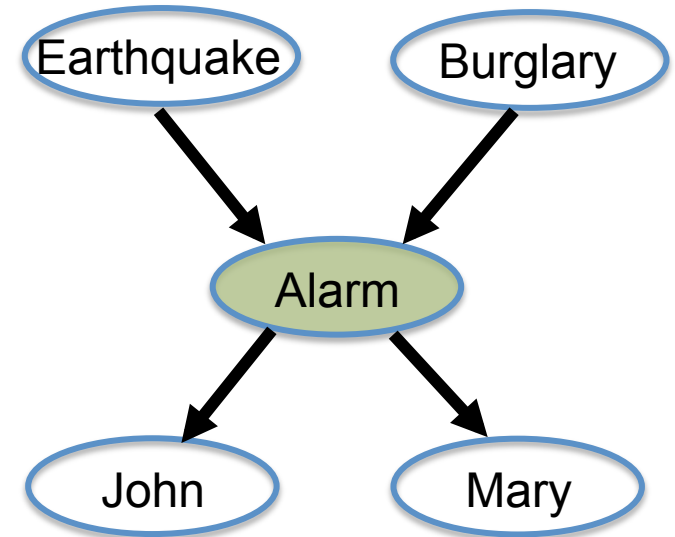
Reasoning in Bayesian networks

- How are J,M related given A?
  - P(M) = 0.0117
  - P(M|A) = 0.7
  - P(M|A,J) = 0.7
  - Conditionally independent

    (we actually know this by construction!)

- Proof:

$$p(J, M | a) \propto \sum_{e,b} p(e) \ p(b) \ p(a|e,b) \ p(J|a) \ p(M|a)$$

$$= \left(\sum_{e,b} p(e, b, a) \right) \ p(J|a) \ p(M|a)$$

$$= p(a) \ p(J|a) \ p(M|a)$$

$$= c_a \ f_a(J) \ g_a(M)$$
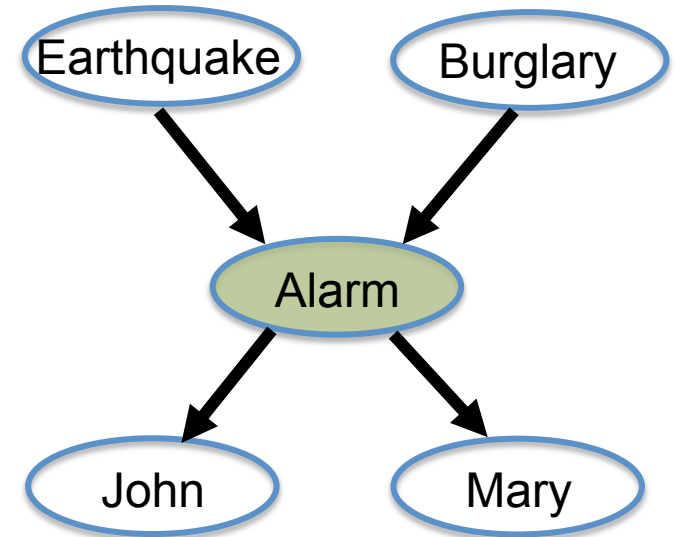
Reasoning in Bayesian networks

- How are J,B related given A?
  - P(B)   = 0.001
  - P(B|A)   = 0.3735
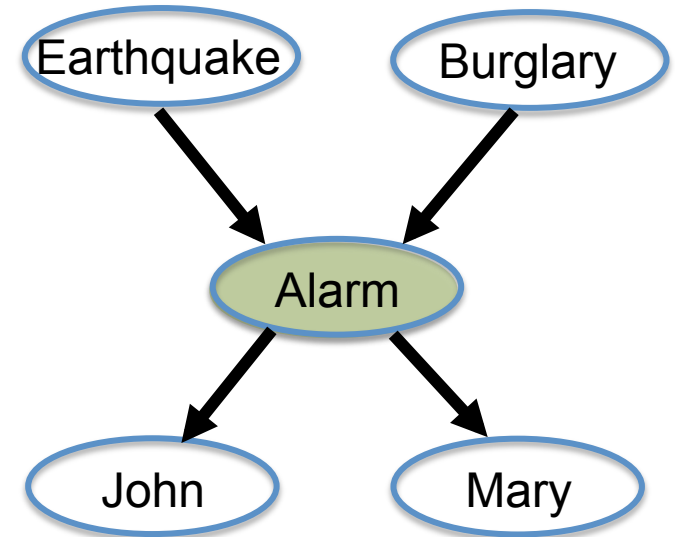 - P(B|A,J)   = 0.3735
  - Conditionally independent

- Proof:

$$p(J, B | a) \propto \sum_{e,m} p(e)\ p(B)\ p(a|e, B)\ p(J|a)\ p(m|a)$$

$$= \left(\sum_{e} p(e, B, a) \right) p(J|a) \left(\sum_{m} p(m|a) \right)$$

$$= p(B, a)\ p(J|a)$$

$$= f_a(B)\ g_a(J)$$

# Reasoning in Bayesian networks

- How are E,B related?
  - P(B)   = 0.001
  - P(B|E)   = 0.001
  - (Marginally) independent

- What about given A?
  - P(B|A)   = 0.3735
  - P(B|A,E)   = 0.0032
  - Not conditionally independent!
  - The "causes" of A become coupled by observing its value
  - Sometimes called "explaining away"
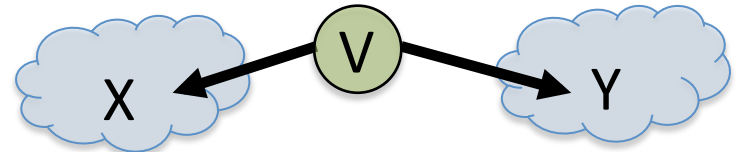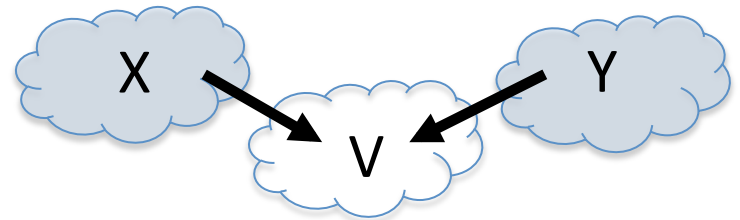
D-Separation

- Prove sets X,Y independent given Z?

- Check all *undirected* paths from X to Y

- A path is "inactive" if it passes through:

  (1) A "chain" with an observed variable
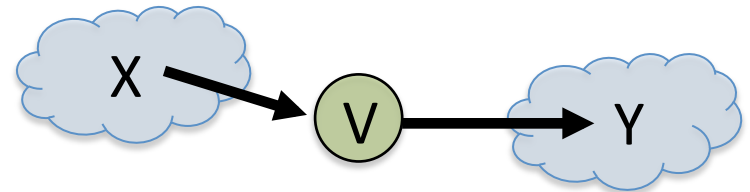
  (2) A "split" with an observed variable

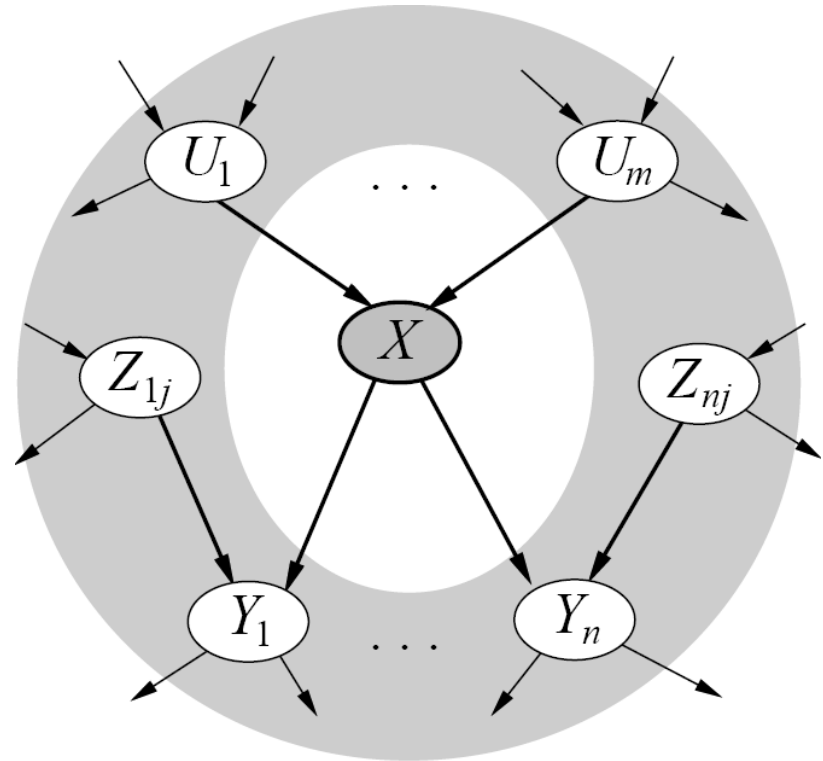 (3) A "vee" with **only unobserved** variables below it

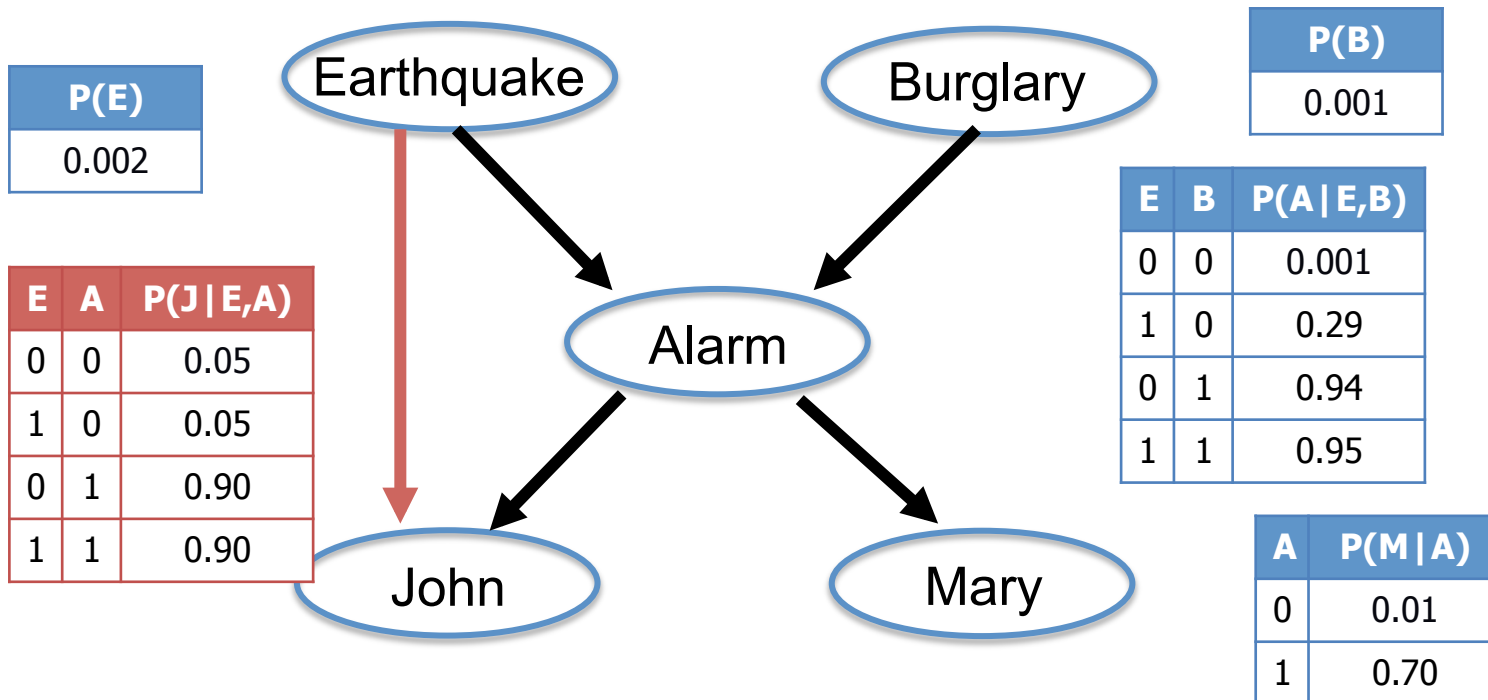- If all paths are inactive, conditionally independent!

Markov blanket

A node is conditionally independent
of all other nodes in the network
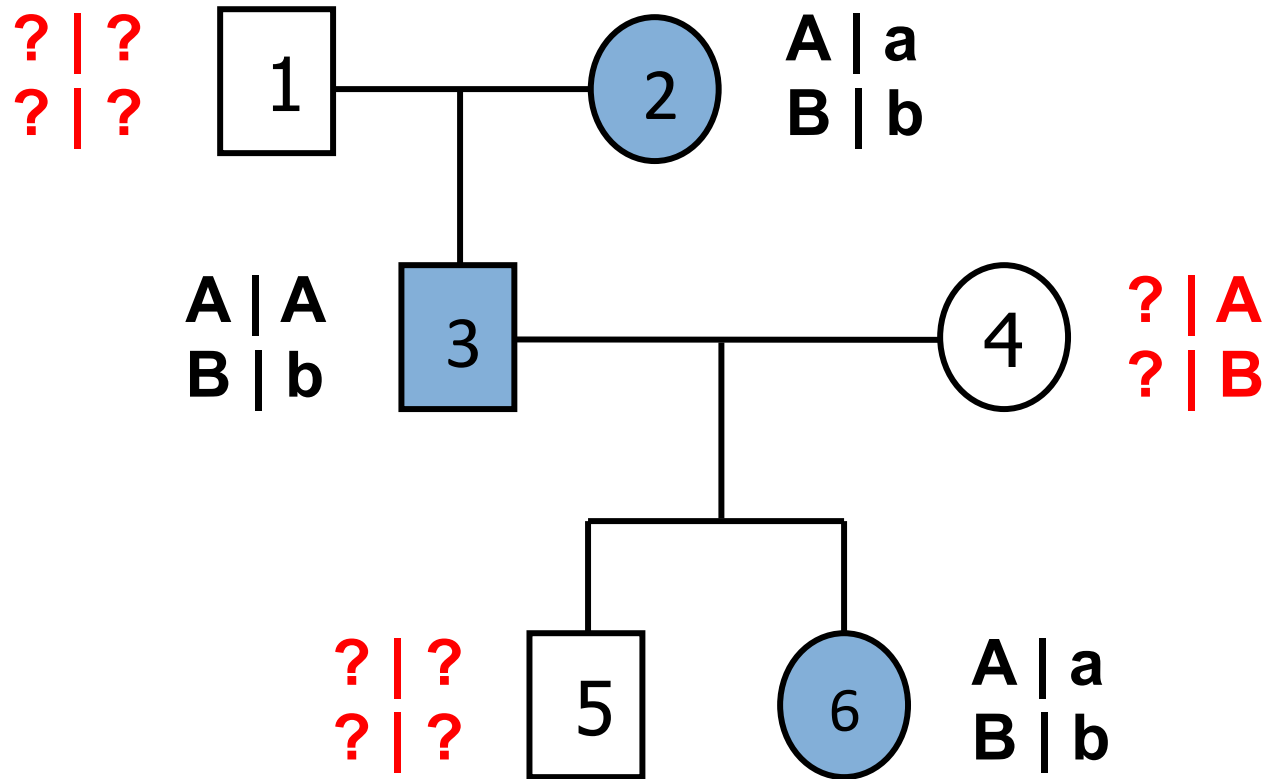given its Markov blanket (in gray)

Graphs and Independence

- Graph structure allows us to infer independence in p(.)
  - X,Y d-separated given Z?

- Adding edges
  - Fewer independencies inferred, but still valid to represent p(.)
  - Complete graph: can represent any distribution p(.)

P(E)
0.002

P(B)
0.001

E	A	P(J|E,A)
0	0	0.05
1	0	0.05
0	1	0.90
1	1	0.90

E	B	P(A|E,B)
0	0	0.001
1	0	0.29
0	1	0.94
1	1	0.95

A	P(M|A)
0	0.01
1	0.70

Earthquake → Alarm
Burglary → Alarm
Alarm → John
Alarm → Mary
Earthquake → John

Genetic linkage analysis

? | ?
? | ?

1

2

A | a
B | b

A | A
B | b

3

4

? | A
? | B

? | ?
? | ?

5

6

A | a
B | b
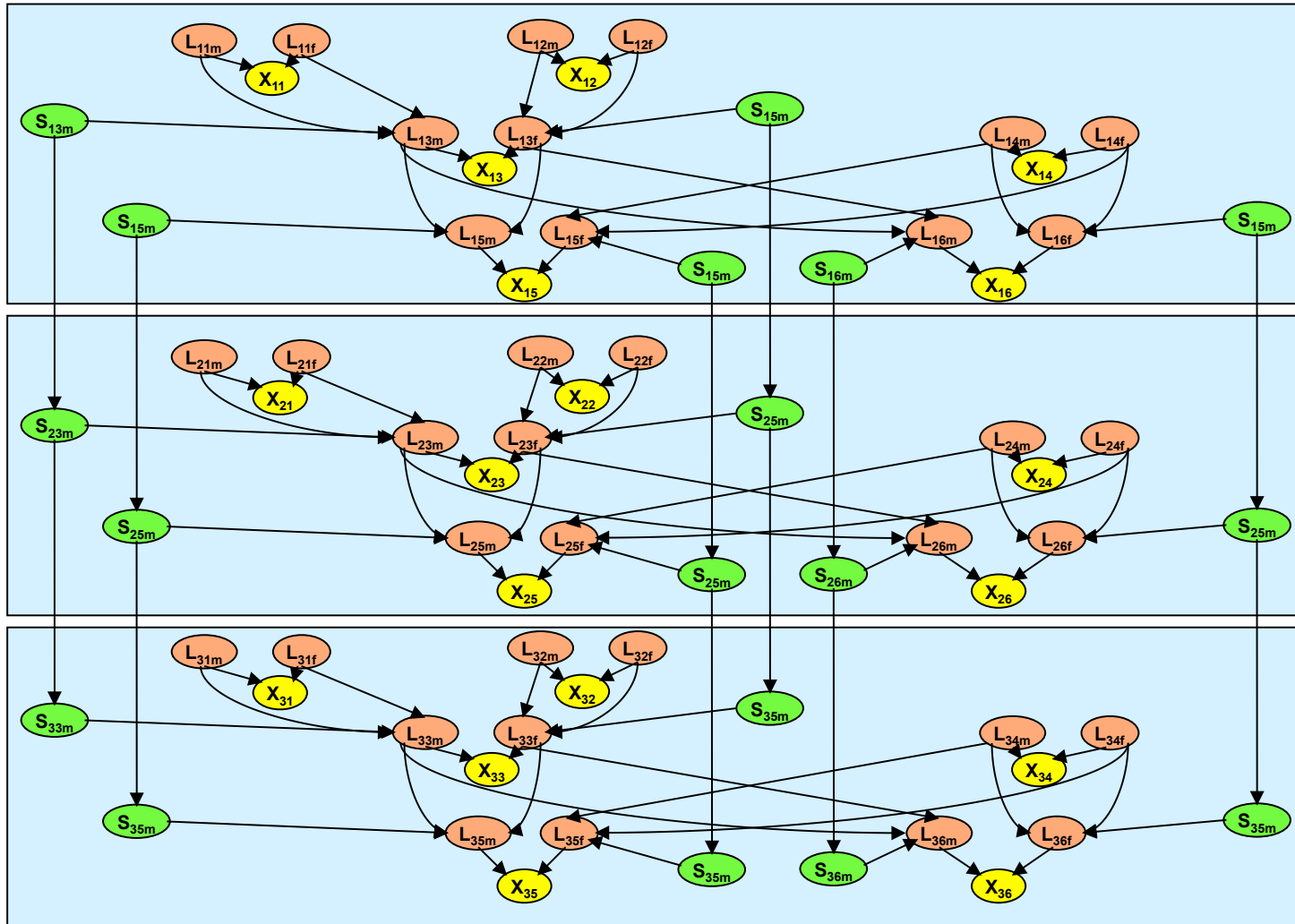
- 6 individuals
- Haplotype: {2, 3}
- Genotype: {6}
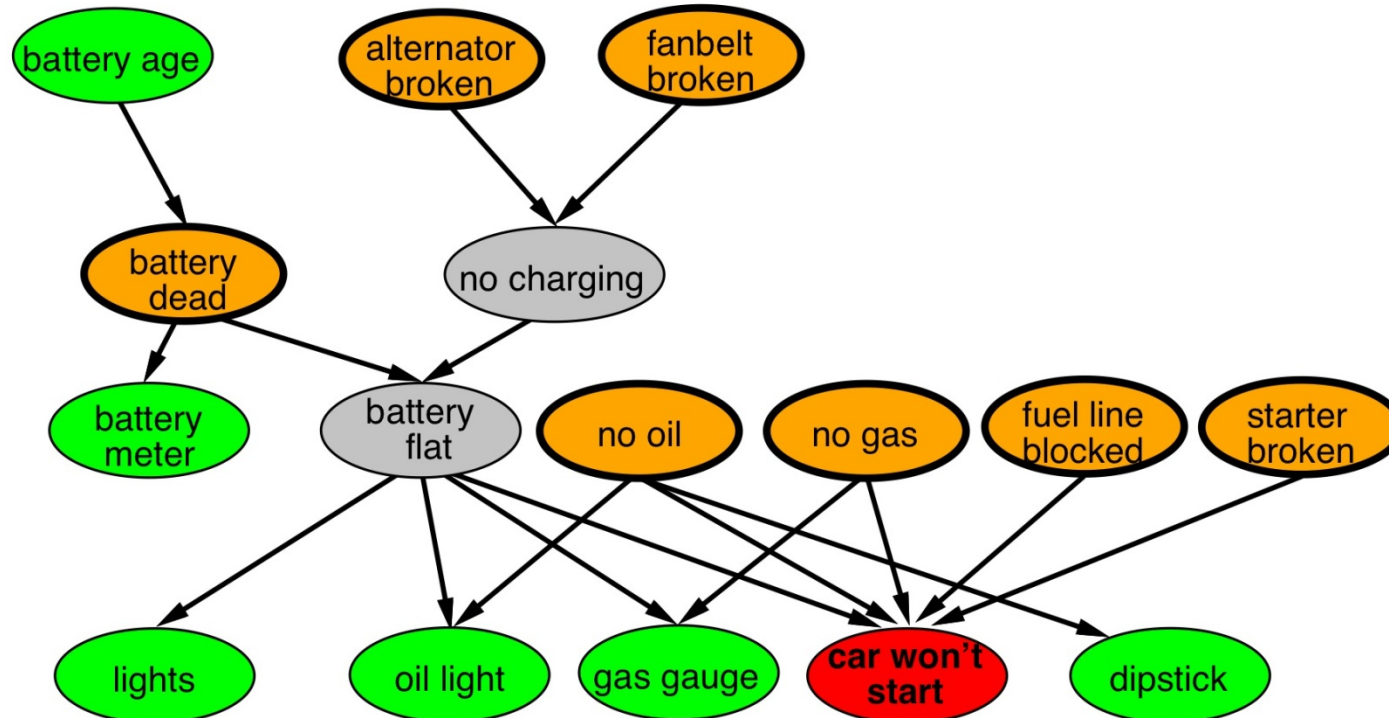- Unknown

Pedigree model: 6 people, 3 markers

Example: Car diagnosis

Initial evidence: car won't start

Testable variables (green), "broken, so fix it" variables (orange)

Hidden variables (gray) ensure sparse structure, reduce parameters

Compact conditional distributions contd.

Noisy-OR distributions model multiple noninteracting causes
1) Parents $U_1 \ldots U_k$ include all causes (can add leak node)
2) Independent failure probability q_i for each cause alone

$$\Rightarrow P(X|U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = 1 - \prod_{i=1}^{j} q_i$$

$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F	**0.0**	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Number of parameters **linear** in number of parents

Summary

- Bayesian networks
  - Directed, acyclic graphs
  - Encode chain rule + conditional independence
  - Efficient representation

- Building a Bayesian network
  - Select ordering
  - Evaluate independence relations
  - Assign probabilities

- Reasoning about independence in the graph
  - D-separation
  - Markov blanket