

Bayesian Networks

Doug Downey
EECS 349 Machine Learning

Answering Queries: Summing Out

		Intelligence = i^1		Intelligence= i^2	
		Time= t^1	Time= t^2	Time= t^1	Time= t^2
Grade	g^1	0.05	0.02	0.15	0.03
	g^2	0.14	0.14	0.05	0.0
	g^3	0.10	0.25	0.01	0.02

$P(\text{Grade} \mid \text{Time} = t^1)?$

$$\sum_{v \in \text{Val}(\text{Intelligence})} P(\text{Grade}, \text{Intelligence} = v \mid \text{Time} = t^1)$$



Answering Queries: Solved?

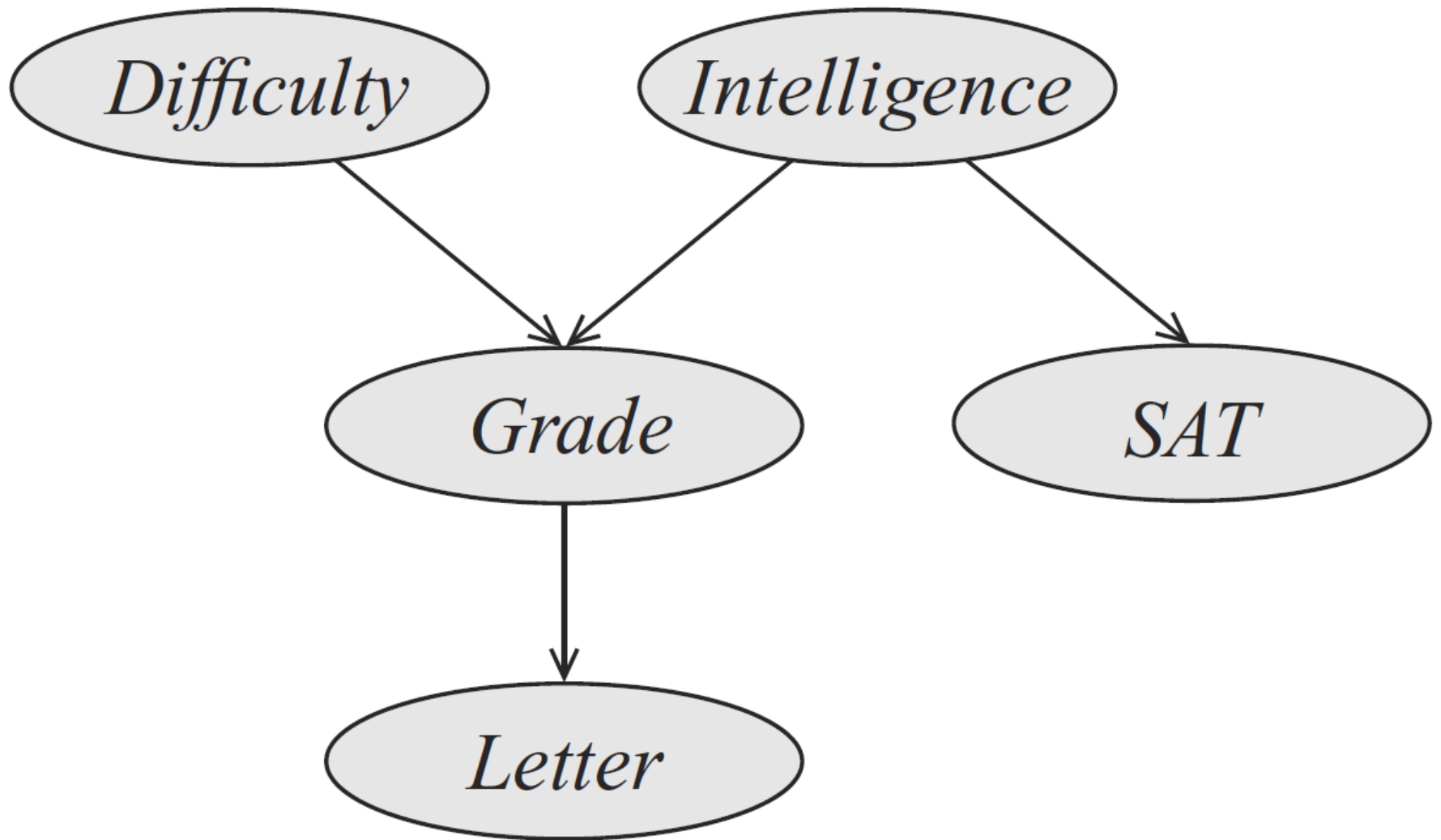
- ▶ Given the joint distribution, we can answer any query by summing
- ▶ ...but, joint distribution has 2^{500} parameters
- ▶ For non-trivial queries (e.g., summing over all possible values of a subset of boolean r.v.s or more), it requires
 - ▶ Way too much **computation** to compute the sum
 - ▶ Way too many **observations** to learn the parameters
 - ▶ Way too much **space** to store the joint distribution

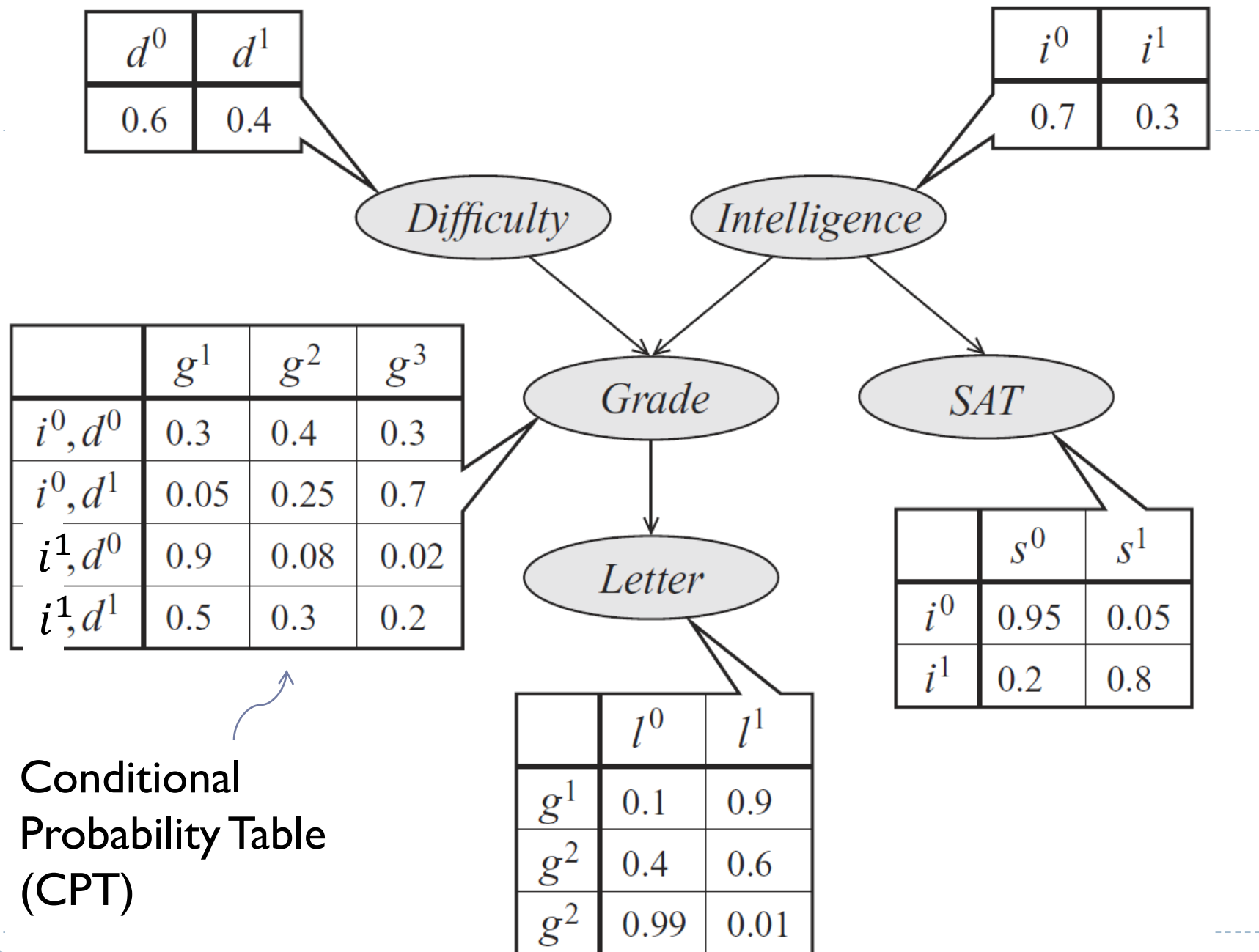


Bayesian Networks

- ▶ A general framework for modeling probability distributions
 - ▶ Expresses conditional independencies
- ▶ Begin with a graph
 - ▶ **Nodes**: Random variables (e.g. attributes, classes)
 - ▶ Directed **Edges**: Causal relationships







What does this wacky thing do?

- ▶ BNs represent the joint distribution compactly
- ▶ You can obtain the BN's probabilities for an event by multiplying the relevant values from each CPT:

$$P(i^1, d^0, g^2, s^1, l^0) = \dots$$



$$P(i^1, d^0, g^2, s^1, l^0)?$$

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

Difficulty

Intelligence

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

Grade

SAT

Letter

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^2	0.99	0.01

What does this wacky thing do?

- ▶ BNs represent the joint distribution compactly
- ▶ You can obtain the BN's probabilities for an event by multiplying the relevant values from each CPT:

$$\begin{aligned} &P(i^1, d^0, g^2, s^1, l^0) \\ &= P(i^1)P(d^0)P(g^2|i^1, d^0)P(s^1|i^1)P(l^0|g^2) \\ &= 0.3 \cdot 0.6 \cdot 0.08 \cdot 0.8 \cdot 0.4 = 0.004608 \end{aligned}$$



Building a Bayes Net

- ▶ Create a node for each attribute or class variable
- ▶ Connect nodes with causal edges
 - ▶ How? Domain knowledge
(or learn from data – more on this in 395/495 PGMs course)
- ▶ Obtain CPTs
 - ▶ How? Use **data**, or write from domain knowledge



Bayes Net Advantages

- ▶ **Compactness**

- ▶ Our “student” network has **15** independent parameters
- ▶ Vs. how many for a full joint distribution table?

- ▶ **Ease of inference**

- ▶ (more on this later)



Think / Pair / Share

What's an upper-bound on the number of parameters in a Bayes Net?

| Think
Start

|
End

Think / Pair / Share

What's an upper-bound on the number of parameters in a Bayes Net?

| Pair
Start

|
End

Think / Pair / Share

What's an upper-bound on the number of parameters in a Bayes Net?

Share

From Graphs to Independencies

- ▶ The Bayes Net encodes **independencies**
 - ▶ Independencies are what allow BN compactness
- ▶ Question:

Which independencies are encoded in a given BN graph?



Global Semantics

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Pa}(X_i))$$



Local Independences

- ▶ Each node is conditionally independent of its non-descendants given its parents.
- ▶ Theorem:
Local Independences $\Leftrightarrow$ Global Semantics



What does the graph look like...

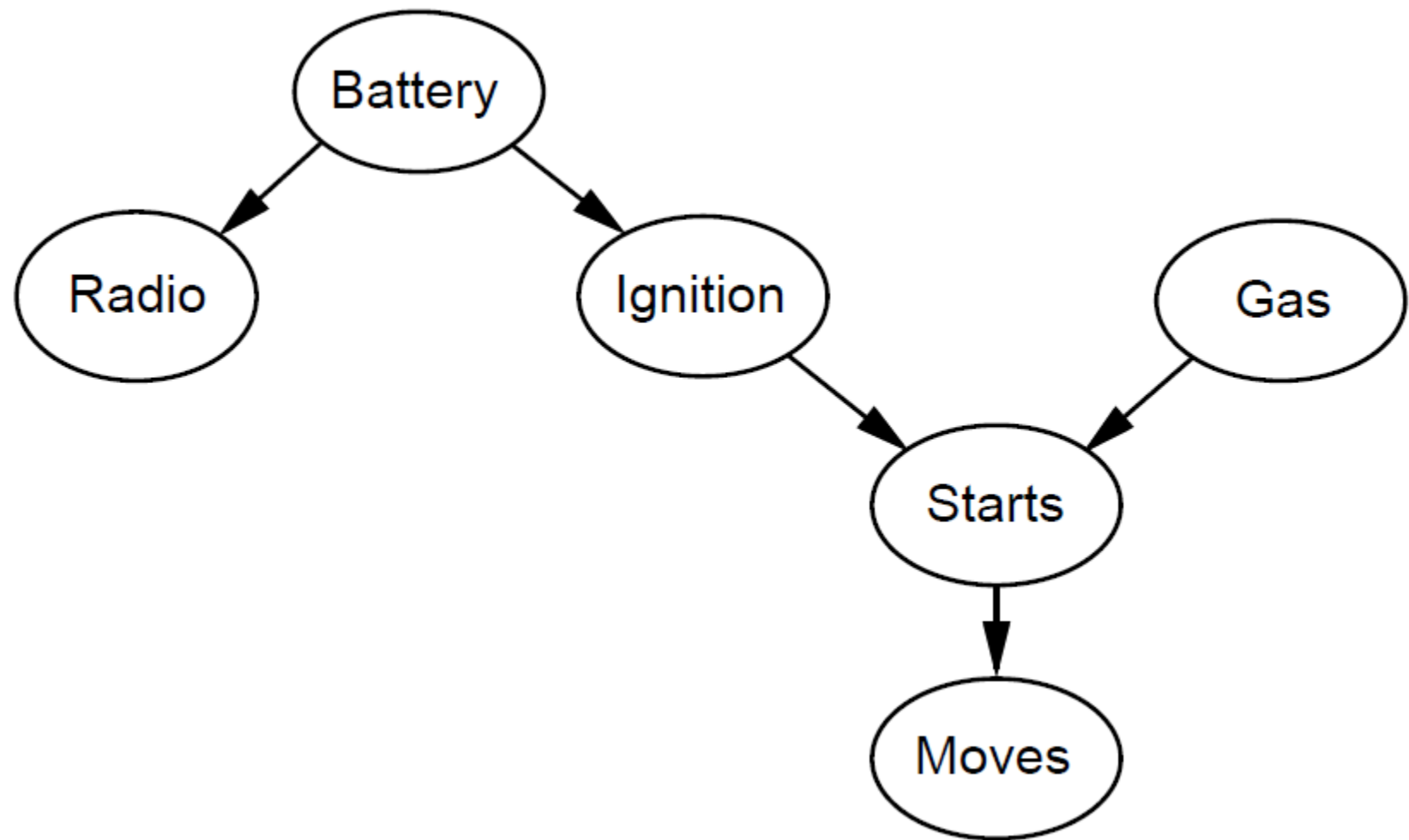
- ▶ No independence?
- ▶ All variables independent?
- ▶ Common Cause? Common Effect?
 - ▶ Correlation $\neq$ causation
 - ▶ “Explaining away”

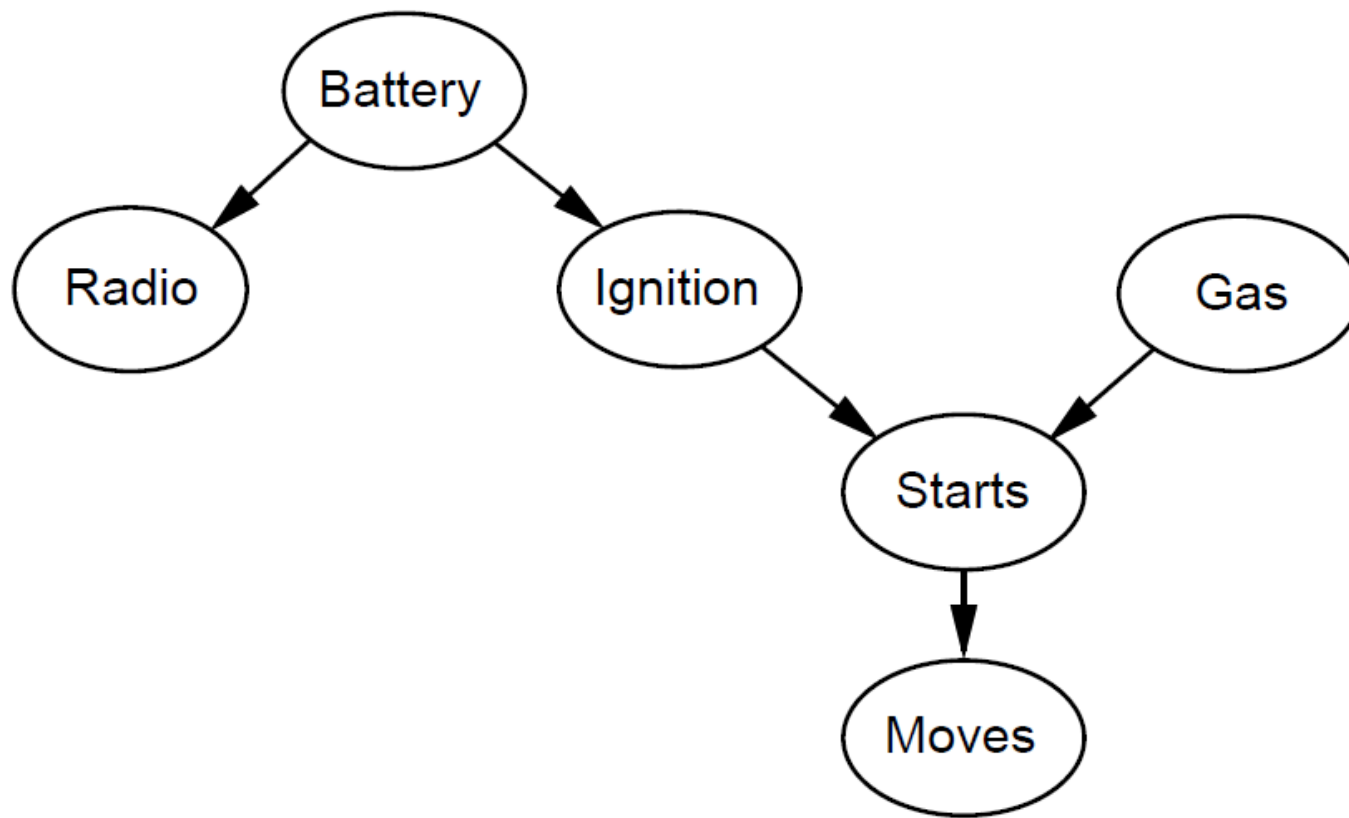


Active Trails and D-separation

- ▶ Two nodes in G are d-separated unless there is an active trail between them
- ▶ An *Active Trail* between nodes X and Y given evidence nodes $\mathbf{E}$ is any path between X and Y such that
 - ▶ For any v-structure $(A \Rightarrow C \Leftarrow B)$ on the path, either C or one of its descendants is in $\mathbf{E}$
 - ▶ No other nodes on the path are in $\mathbf{E}$







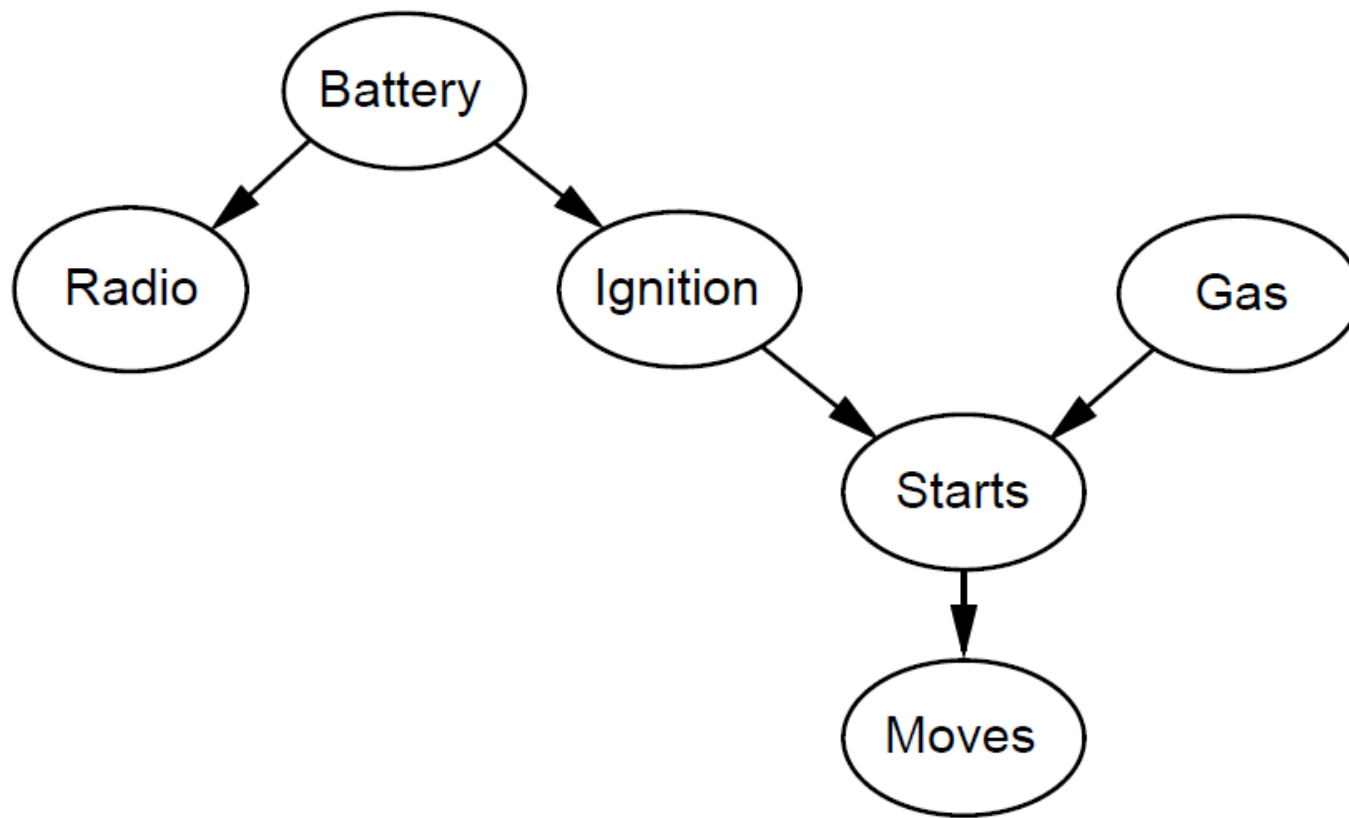
$(\text{Gas} \perp \text{Radio})?$ $(\text{Radio} \perp \text{Ignition})?$

$(\text{Radio} \perp \text{Ignition} \mid \text{Battery})?$ $(\text{Gas} \perp \text{Radio} \mid \text{Moves})?$

| Think

Start

| End



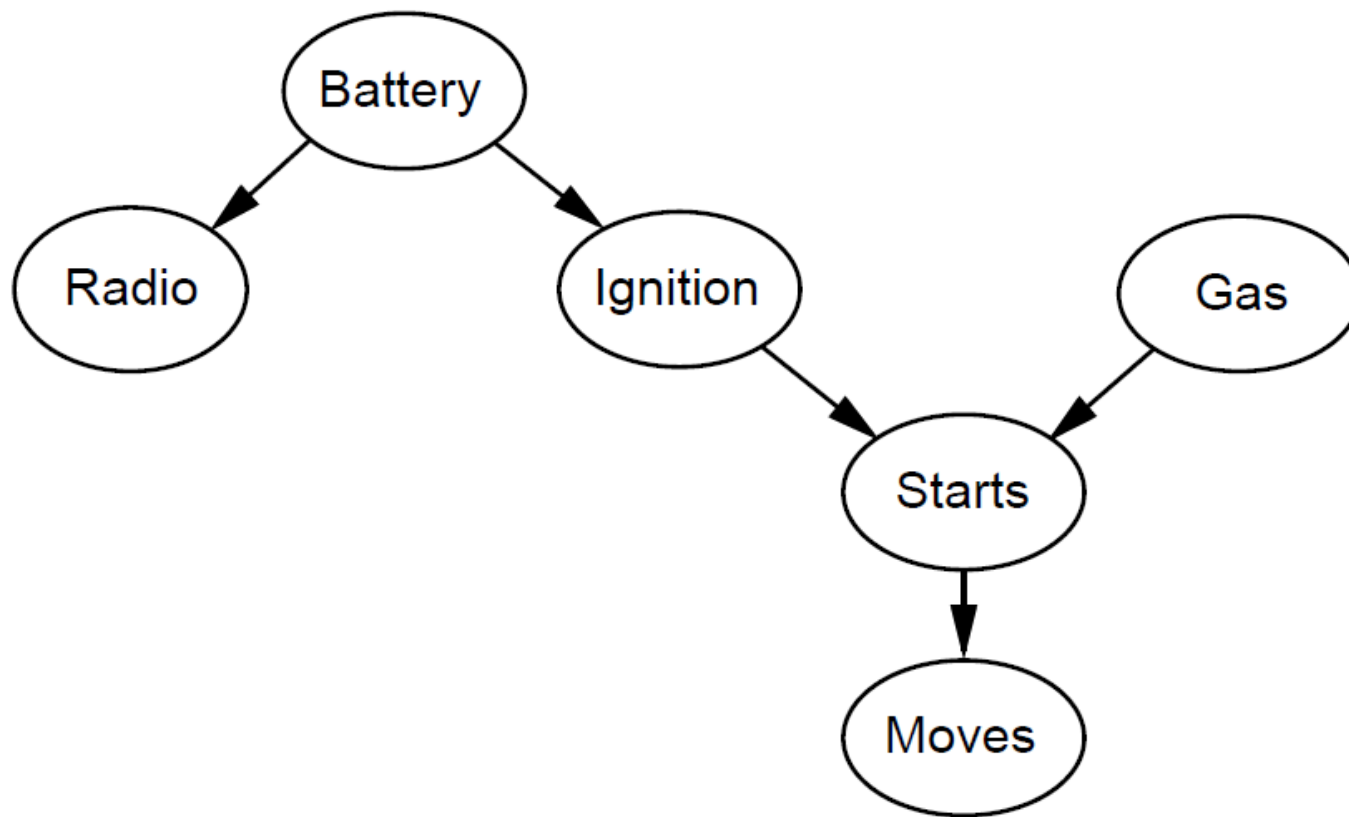
$(\text{Gas} \perp \text{Radio})? (\text{Radio} \perp \text{Ignition})?$

$(\text{Radio} \perp \text{Ignition} \mid \text{Battery})? (\text{Gas} \perp \text{Radio} \mid \text{Moves})?$

| Pair

Start

| End



$(\text{Gas} \perp \text{Radio})?$ $(\text{Radio} \perp \text{Ignition})?$

$(\text{Radio} \perp \text{Ignition} \mid \text{Battery})?$ $(\text{Gas} \perp \text{Radio} \mid \text{Moves})?$

Share