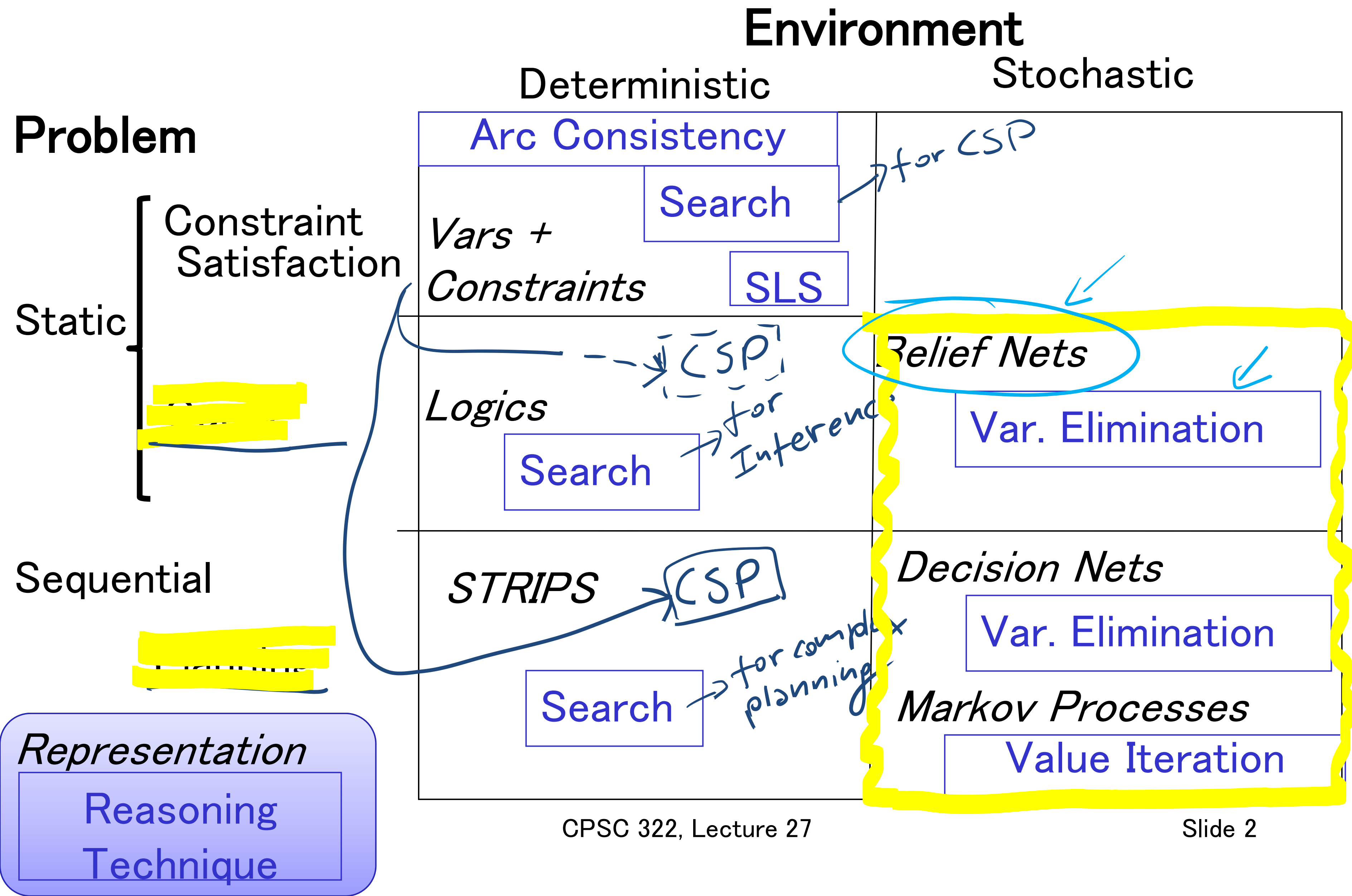
Reasoning Under Uncertainty: Belief Networks

Computer Science cpsc322, Lecture 27

(Textbook Chpt 6.3)

June, 15, 2017

Big Picture: R&R systems



Key points Recap

- We model the environment as a set of ….random vars
  $X_1 \ldots X_n$ JPD $P(X_1 \ldots X_n)$
- Why the joint is not an adequate representation ?

"Representation, reasoning and learning" are "exponential" in ….. ~~#~~ vars

Solution: Exploit marginal&**conditional** independence

$\boxed{P(X|Y)} = P(X)$ $\boxed{P(X|YZ)} = P(X|Z)$

But how does independence allow us to simplify the joint?

CHAIN RULE!

Lecture Overview

- **Belief Networks**
  - **Build sample BN**
  - Intro Inference, Compactness, Semantics
  - More Examples

Belief Nets: Burglary Example

There might be a **burglar** in my house

B

The **anti-burglar alarm** in my house may go off

A

I have an agreement with two of my neighbors, **John** and **Mary**, that they **call** me if they hear the alarm go off when I am at work

M J

Minor earthquakes may occur and sometimes the set off the alarm.

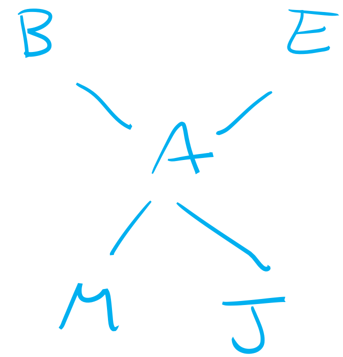
E

Variables: B A M J E $n=5$

Joint has 2^5-1 entries/probs 2^n-1

Belief Nets: Simplify the joint

- Typically order vars to reflect causal knowledge (i.e., causes *before effects*)
  - A burglar (B) can set the alarm (A) off
  - An earthquake (E) can set the alarm (A) off
  - The alarm can cause Mary to call (M)
  - The alarm can cause John to call (J)



$$P(\underline{B, E}, A, M, J)$$

marginal indep.

conditional indep.

- Apply Chain Rule

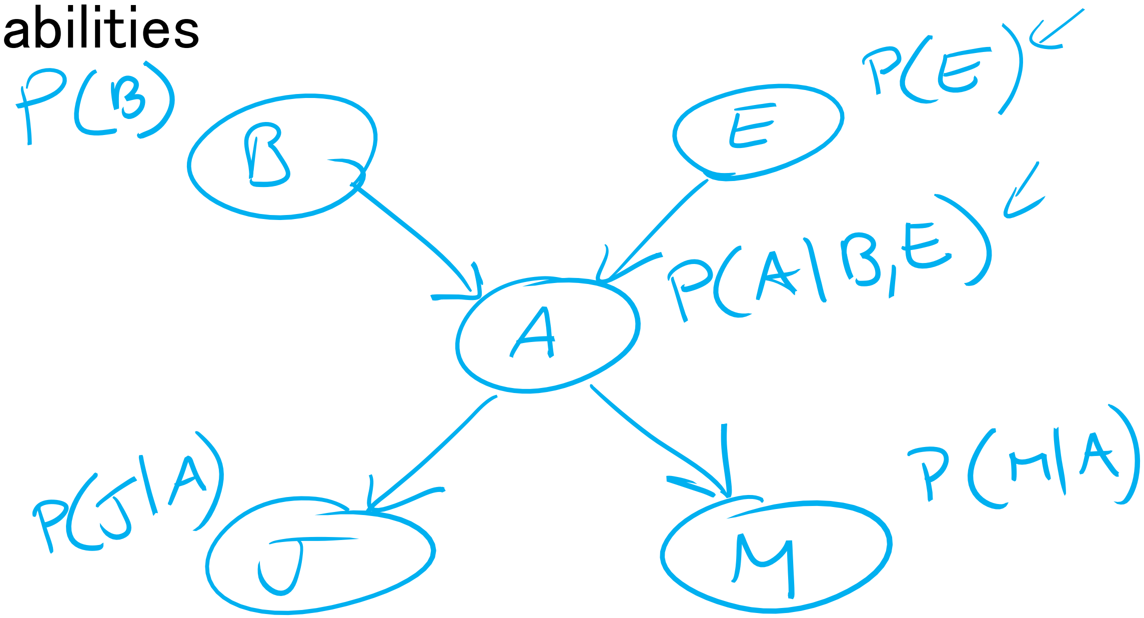
$$\underline{P(B)}\ \underline{P(E \mid \cancel{B})}\ P(A \mid B, E)\ P(M \mid A, \cancel{E}, \cancel{B})\ \underline{P(J \mid \cancel{M}, A, \cancel{E}, \cancel{B})}$$

- Simplify according to marginal&conditional independence

Belief Nets: Structure + Probs

$$P(B) * P(E) * \underline{P(A \mid \boxed{B, E})} * P(M \mid A) * P(J \mid A)$$

- Express remaining dependencies as a network
  - Each var is a node
  - For each var, the conditioning vars are its parents
  - Associate to each node corresponding conditional probabilities



- Directed Acyclic Graph (DAG)

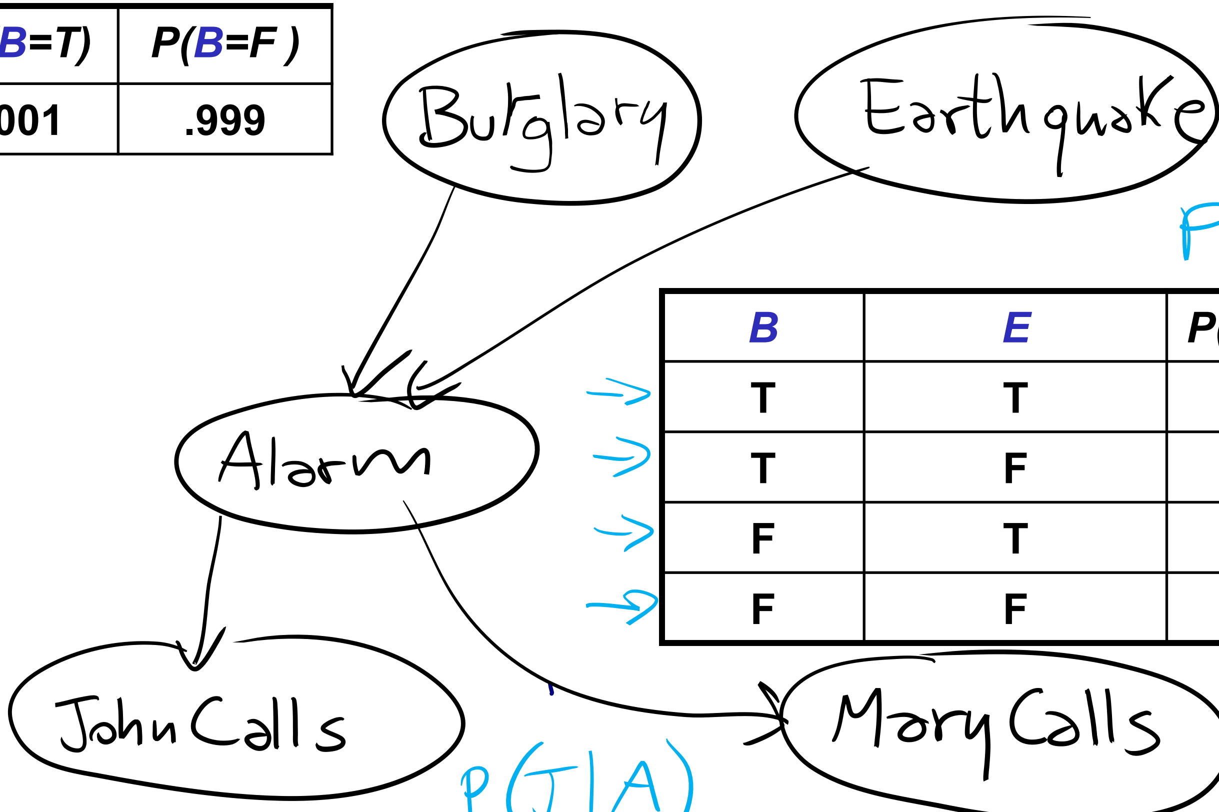
Burglary: complete BN

P(B)

P(B=T)	P(B=F)
.001	.999

P(E)

P(E=T)	P(E=F)
.002	.998



P(A | B,E)

B	E	P(A=T | B,E)	P(A=F | B,E)
T	T	.95	.05
T	F	.94	.06
F	T	.29	.71
F	F	.001	.999

P(J | A)

A	P(J=T | A)	P(J=F | A)
T	.90	.10
F	.05	.95

P(M | A)

A	P(M=T | A)	P(M=F | A)
T	.70	.30
F	.01	.99

call for any other reasons

Lecture Overview

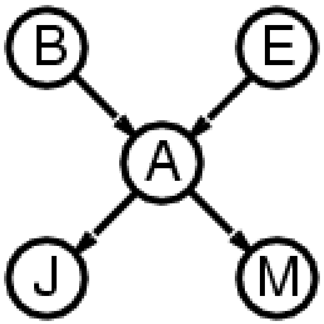
- **Belief Networks**
  - Build sample BN
  - **Intro Inference, Compactness, Semantics**
  - More Examples

Burglary Example: Bnets inference

Our BN can answer any probabilistic query that can be answered by processing the joint!

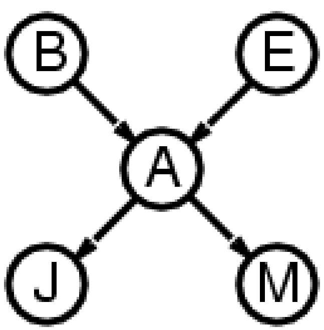
(Ex1) I'm at work,

- neighbor John calls to say my alarm is ringing,
- neighbor Mary doesn't call.
- No news of any earthquakes.
- Is there a burglar?



(Ex2) I'm at work, Try this

- Receive message that neighbor John called ,
- News of minor earthquakes.
- Is there a burglar?



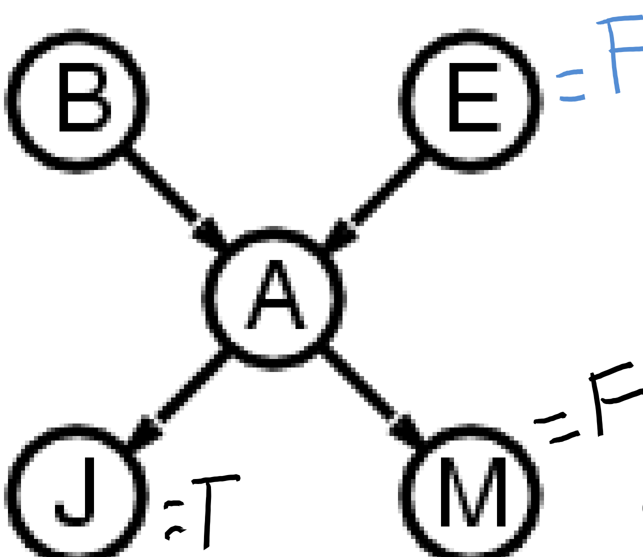
Set digital places to monitor to 5

Burglary Example: Bnets inference

Our BN can answer any probabilistic query that can be answered by processing the joint!

(Ex1) I'm at work,

- neighbor John calls to say my alarm is ringing,
- neighbor Mary doesn't call.
- **No news of any earthquakes.**
- Is there a burglar?



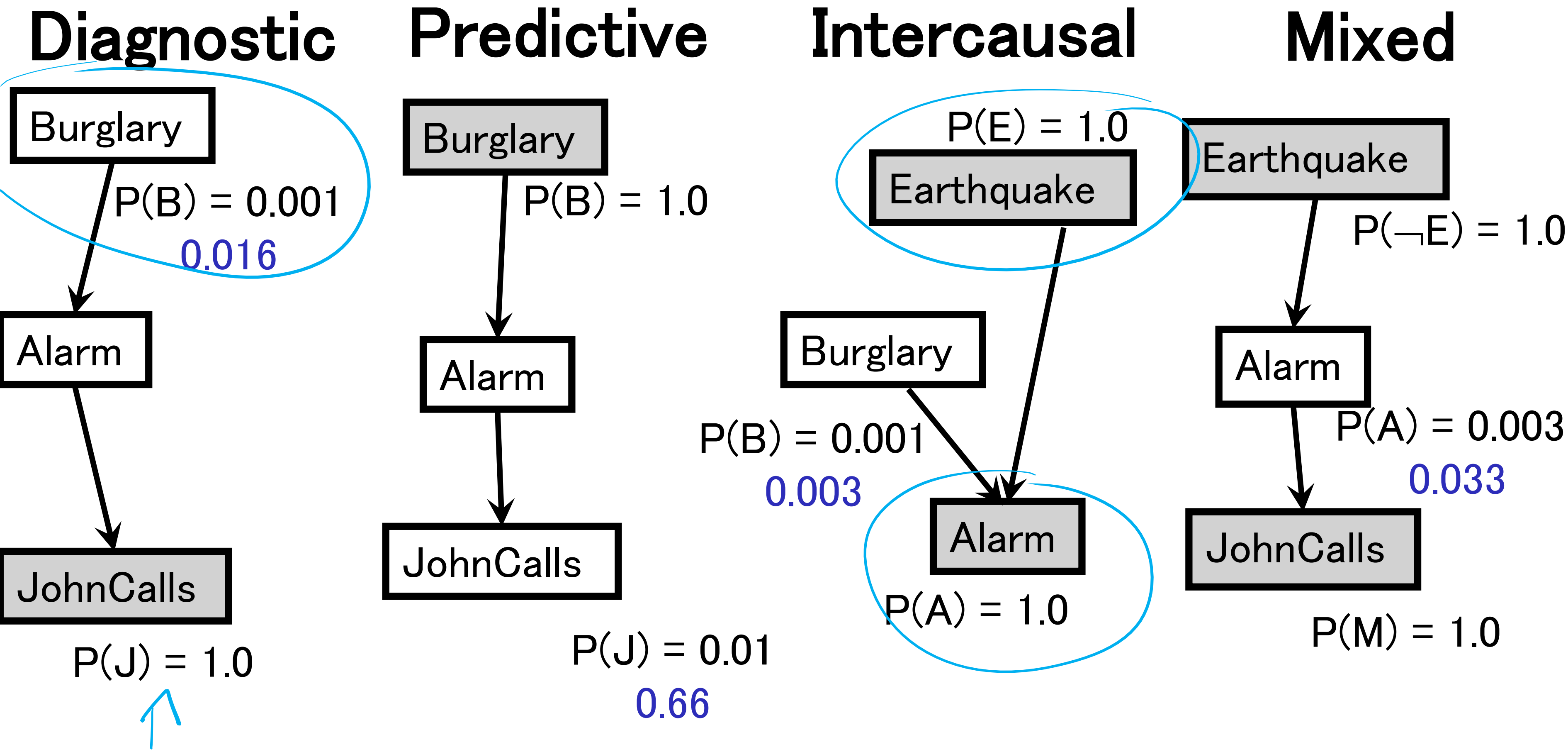
The probability of Burglar will:

A. Go down

B. Remain the same

C. Go up

Bayesian Networks – Inference Types



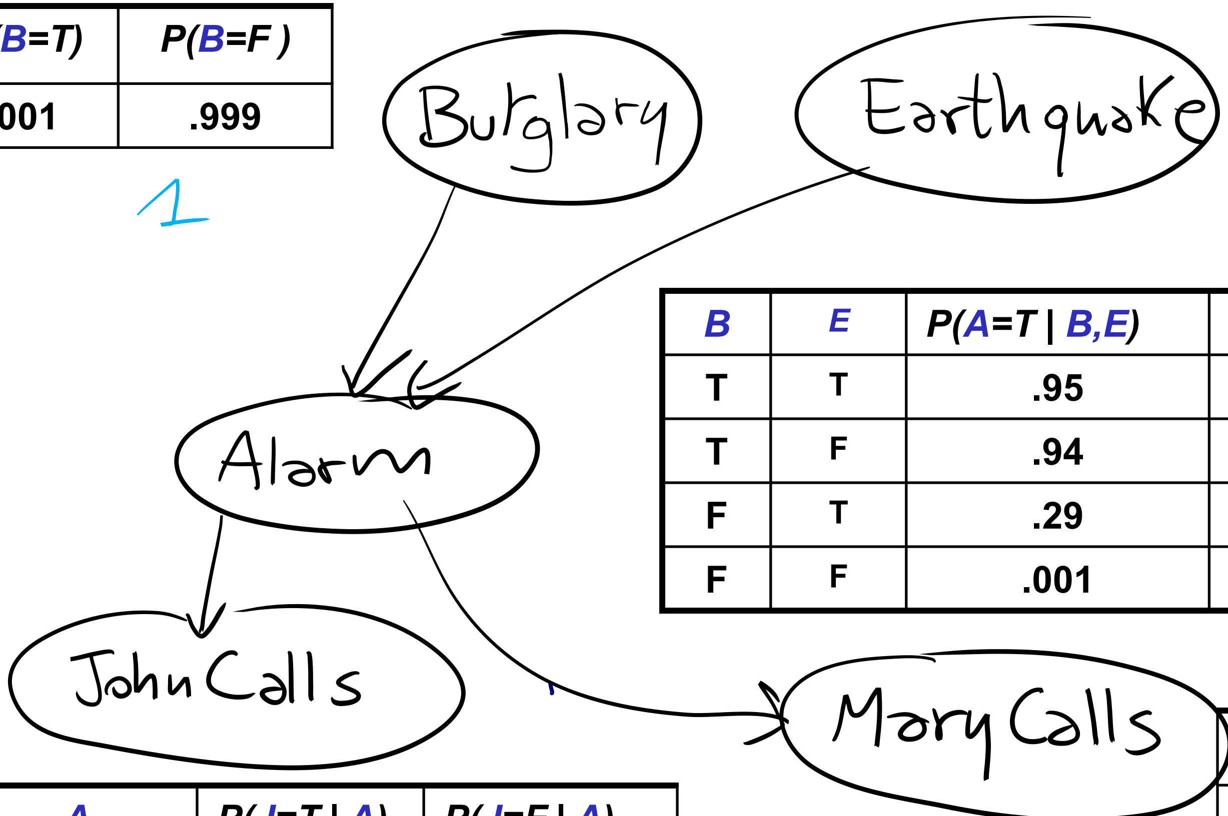
BNnets: Compactness

P(B=T)	P(B=F)
.001	.999

1

P(E=T)	P(E=F)
.002	.998

1



B	E	P(A=T | B,E)	P(A=F | B,E)
T	T	.95	.05
T	F	.94	.06
F	T	.29	.71
F	F	.001	.999

4

A	P(J=T | A)	P(J=F | A)
T	.90	.10
F	.05	.95

2

A	P(M=T | A)	P(M=F | A)
T	.70	.30
F	.01	.99

2

BNet

$2 + 2 + 4 + 1 + 1 = 10$

$|JPD| = 2^5 - 1$

BNets: Compactness

Conditional Probability Table

In General:

A CPT for boolean X_i with k boolean parents has 2^k rows for the combinations of parent values

Each row requires **one number** p_i for $X_i = true$ (the number for $X_i = false$ is just $1-p_i$)

for each node

If **each variable** has no more than k **parents**, the complete network requires $O(n 2^k)$ numbers

For $k \ll n$, this is a substantial improvement,

- the numbers required grow linearly with n, vs. $O(2^n)$ for the full joint distribution

BNets: Construction General Semantics

The full joint distribution can be defined as the product of conditional distributions:

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}) \text{ (chain rule)}$$

Simplify according to marginal&conditional independence

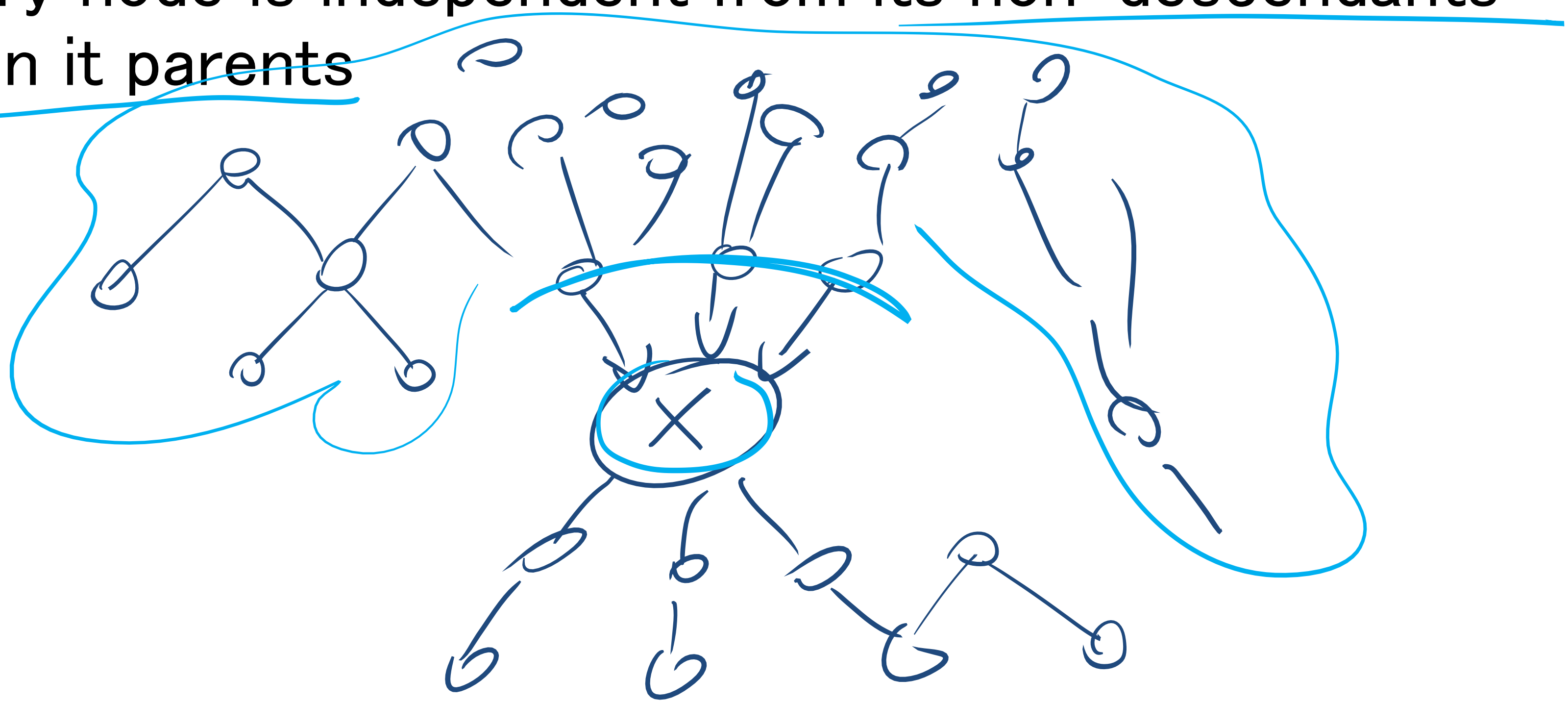
- Express remaining dependencies as a network
  - Each var is a node
  - For each var, the conditioning vars are its parents
  - Associate to each node corresponding conditional probabilities

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

BNets: Construction General Semantics (cont')

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

- Every node is independent from its non-descendants given it parents



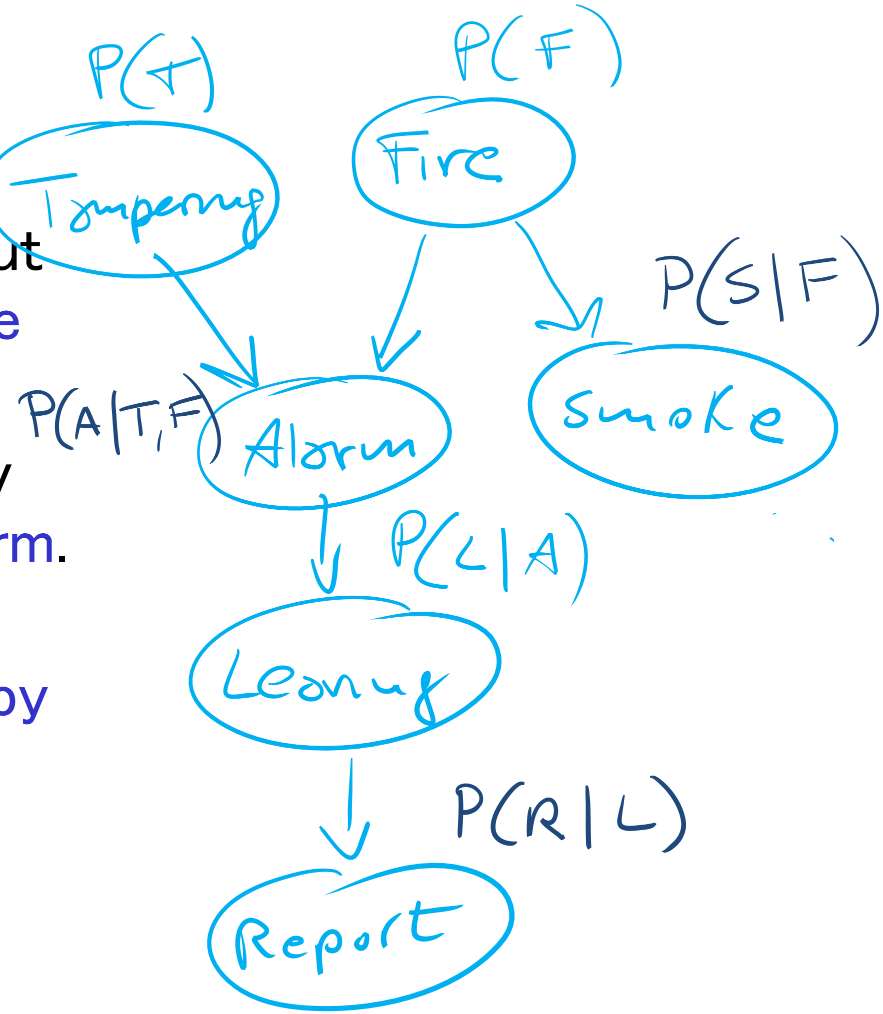
Lecture Overview

- **Belief Networks**
  - Build sample BN
  - Intro Inference, Compactness, Semantics
  - **More Examples**

Other Examples: Fire Diagnosis (textbook Ex. 6.10)

Suppose you want to diagnose whether there is a fire in a building

- you receive a noisy report about whether everyone is leaving the building.
- if everyone is leaving, this may have been caused by a fire alarm.
- if there is a fire alarm, it may have been caused by a fire or by tampering
- if there is a fire, there may be smoke raising from the bldg.



Other Examples (cont')

- Make sure you explore and understand the **Fire Diagnosis** example (we'll expand on it to study Decision Networks)

- **Electrical Circuit** example (textbook ex 6.11)

- **Patient's wheezing and coughing** example (ex. 6.14)

- Several other examples on

Realistic BNet: Liver Diagnosis

Source: Onisko et al., 1999

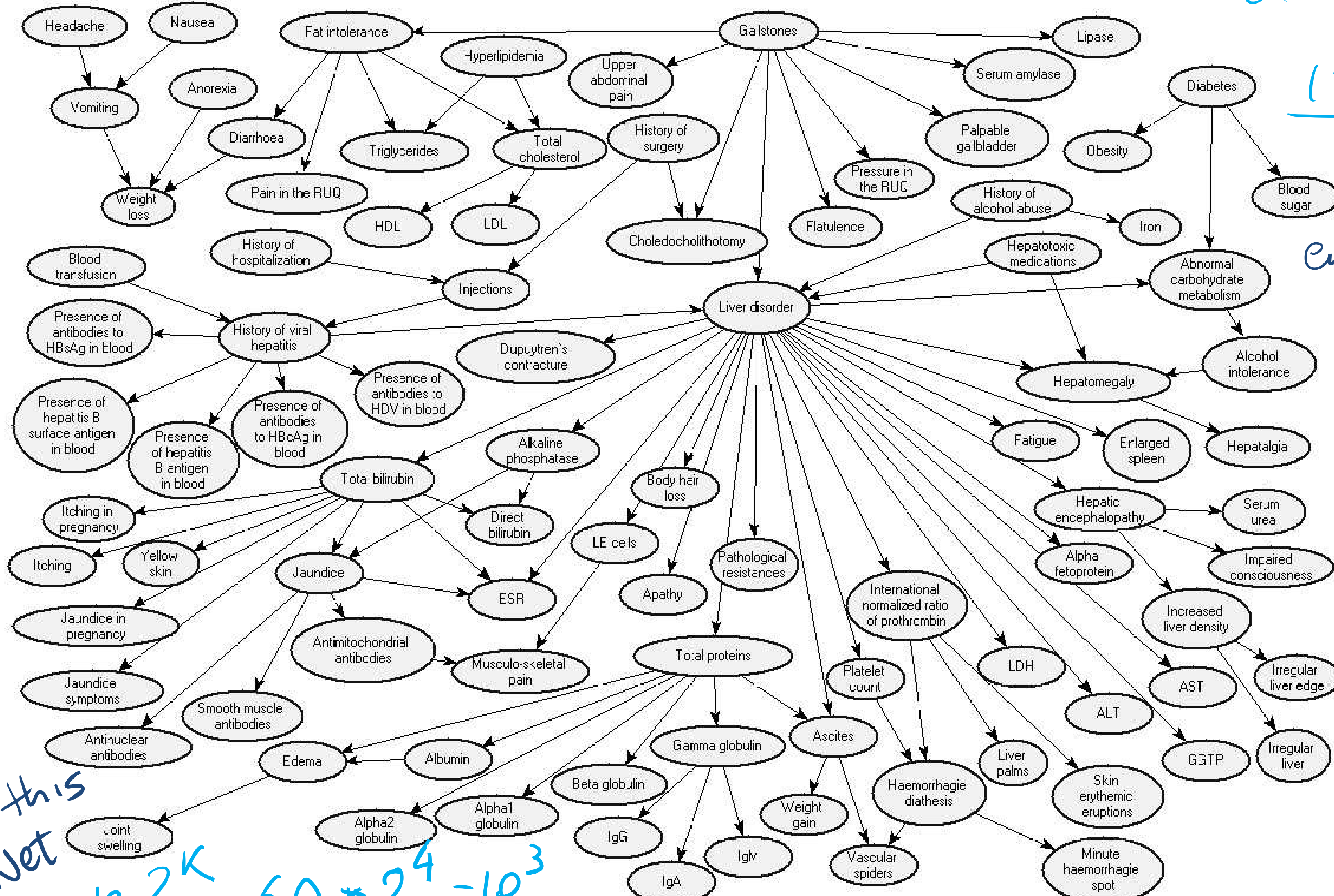
~60 nodes

$n = 60$ $\sim 2^{60}$

JPD

$\approx (2^{10})^6$

$\approx 10^{18}$ entries



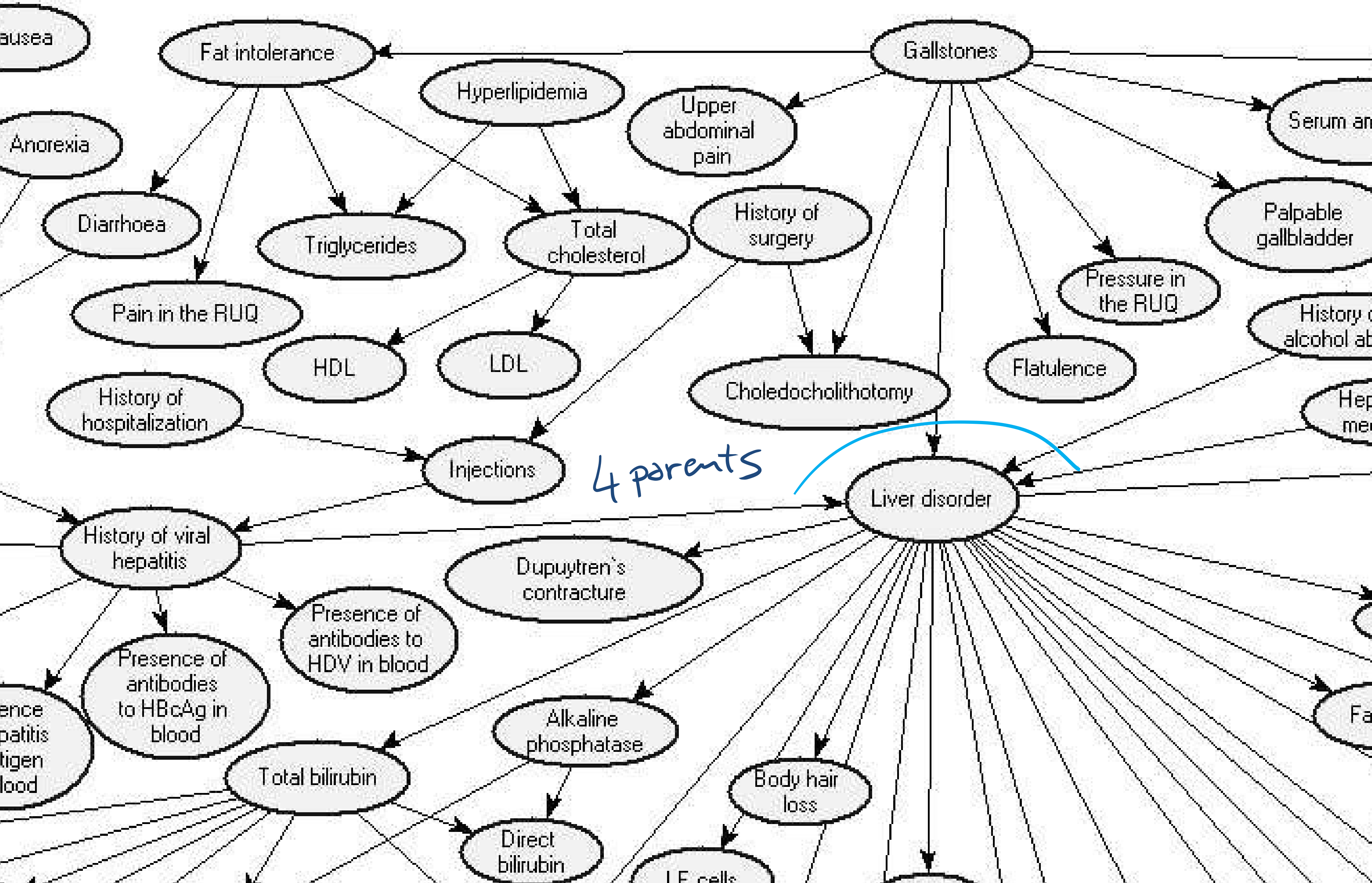
for this BNet

$\sim 2^K$

$60 * 2^4 = 10^3$

Realistic BNet: Liver Diagnosis

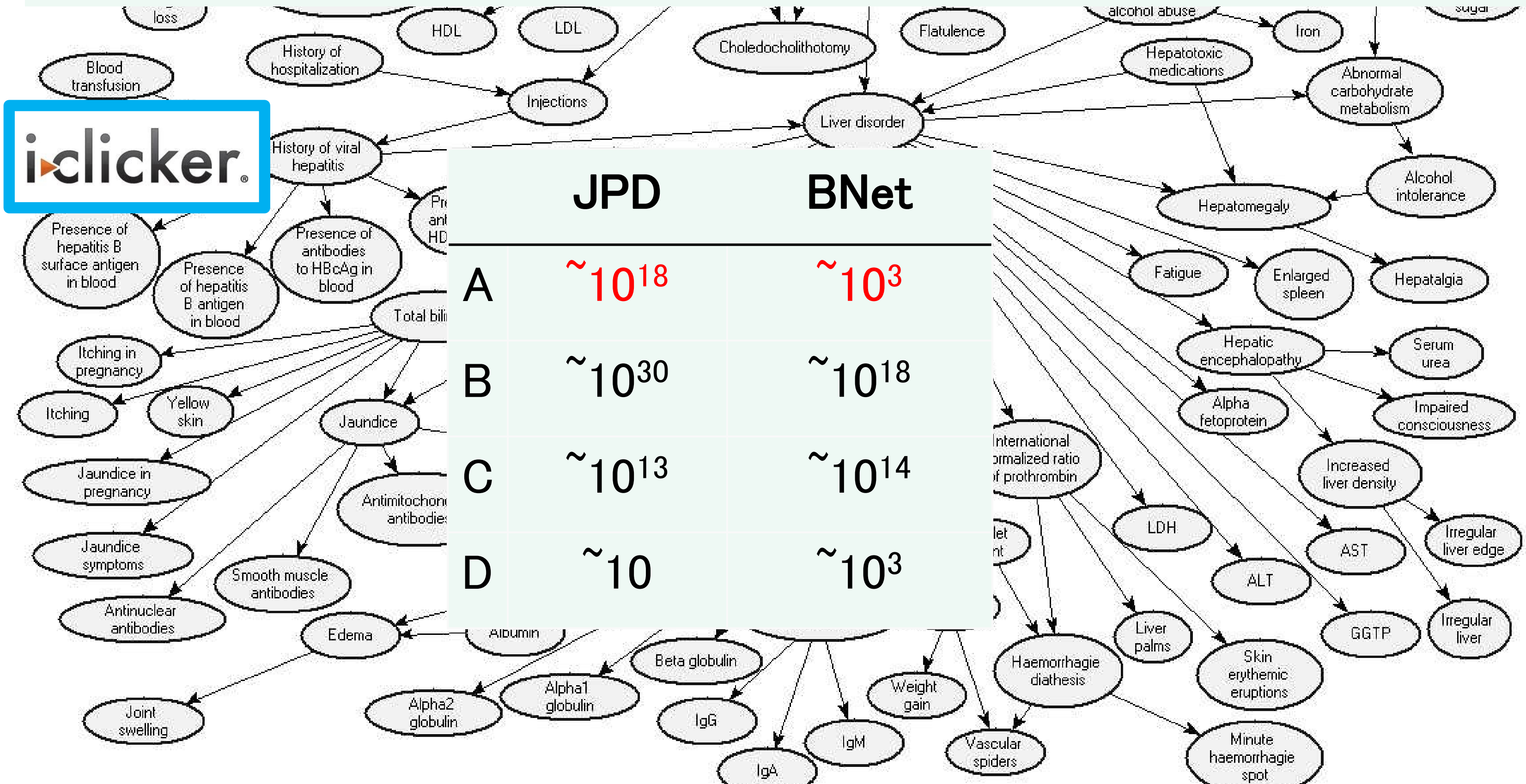
Source: Onisko et al., 1999



Realistic BNet: Liver Diagnosis

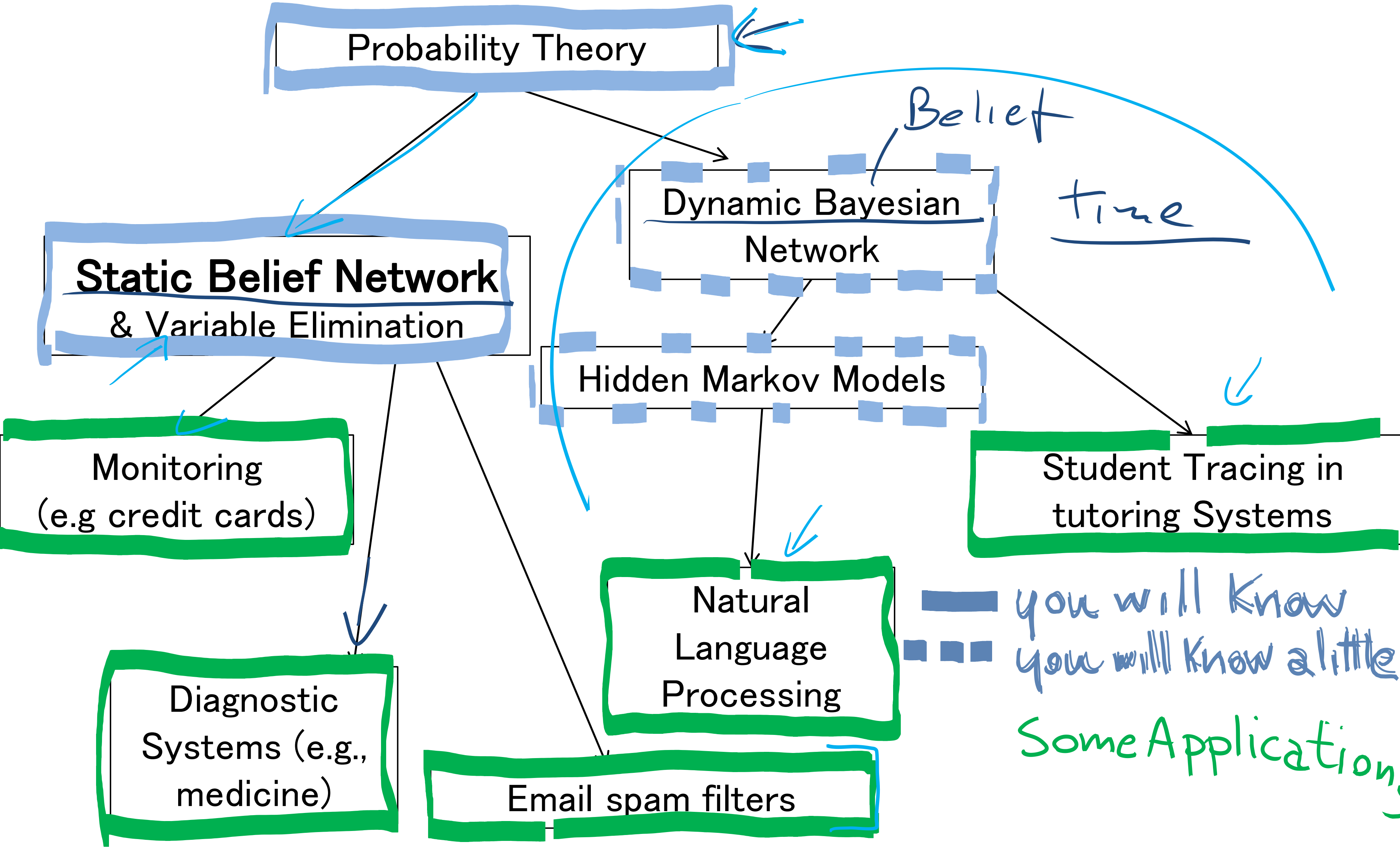
Source: Onisko et al., 1999

Assuming there are ~60 nodes in this Bnet with max number of parents =4; and assuming all nodes are binary, how many numbers are required for the JPD vs BNet



	JPD	BNet
A	$\sim 10^{18}$	$\sim 10^{3}$
B	$\sim 10^{30}$	$\sim 10^{18}$
C	$\sim 10^{13}$	$\sim 10^{14}$
D	~ 10	$\sim 10^{3}$

Answering Query under Uncertainty



Learning Goals for today's class

You can:

Build a Belief Network for a simple domain

Classify the types of inference

Diagnostic, Predictive, Intercausal, Mixed

Compute the representational saving in terms on number of probabilities required

Next Class (Wednesday!)

Bayesian Networks Representation

- **Additional Dependencies** encoded by BNets
- More **compact representations** for CPT
- Very simple but extremely useful Bnet (**Bayes Classifier**)

Belief network summary

- A belief network is a directed acyclic graph (DAG) that effectively expresses independence assertions among random variables.
- The parents of a node X are those variables on which X directly depends.
- Consideration of causal dependencies among variables typically help in constructing a Bnet