New Directions in Privacy-preserving Machine Learning

Kamalika Chaudhuri

University of California, San Diego

Sensitive Data

Medical Records

Genetic Data

Search Logs

Google

AOL Violates Privacy

AOL Violates Privacy

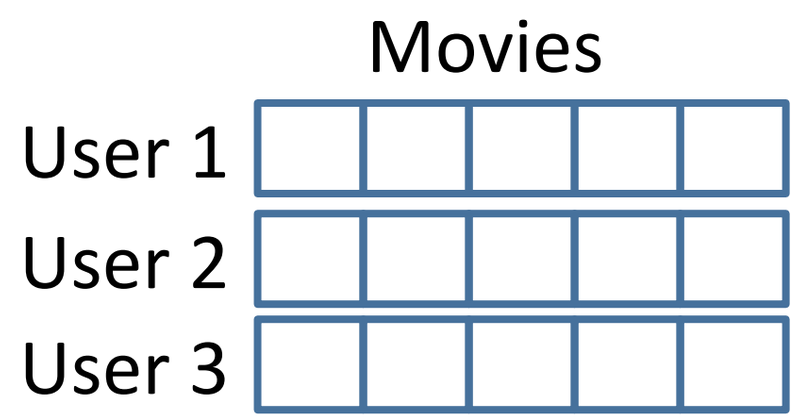
A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on

Netflix Violates Privacy [NS08]



2-8 movie-ratings and dates for Alice reveals:

Whether Alice is in the dataset or not

Alice's other movie ratings

High-dimensional Data is Unique

Example: UCSD Employee Salary Table

Position	Gender	Department	Ethnicity	Salary
Faculty	Female	CSE	SE Asian	-

One employee (Kamalika) fits description!

Simply anonymizing data is **unsafe!**

Disease Association Studies [WLWTZ09]

Cancer

1.00										
.190	1.00									
.216	.251	1.00								
.186	.117	.047	1.00							
.154	.011	.170	.083	1.00						
.190	.140	.102	.095	.139	1.00					
.270	.215	.294	.248	.140	.141	1.00				
.101	.085	.170	.056	.234	.099	.175	1.00			
.239	.071	.163	.111	.161	.093	.199	.157	1.00		
.471	.117	.243	.094	.144	.123	.283	.216	.274	1.00	
.179	.202	.132	.094	.087	.159	.207	.108	.092	.294	1.00

Correlations

Healthy

1.00										
.141	1.00									
.099	.175	1.00								
.093	.199	.157	1.00							
.123	.283	.216	.274	1.00						
.159	.207	.108	.092	.294	1.00					
.088	.152	.075	.163	.156	.220	1.00				
.046	.161	.092	.072	.157	.143	.147	1.00			
.078	.392	.122	.229	.160	.172	.145	.177	1.00		
.045	.155	.135	.139	.110	.048	.126	.104	.169	1.00	
.178	.135	.102	.258	.314	.165	.147	.158	.131	.074	1.00

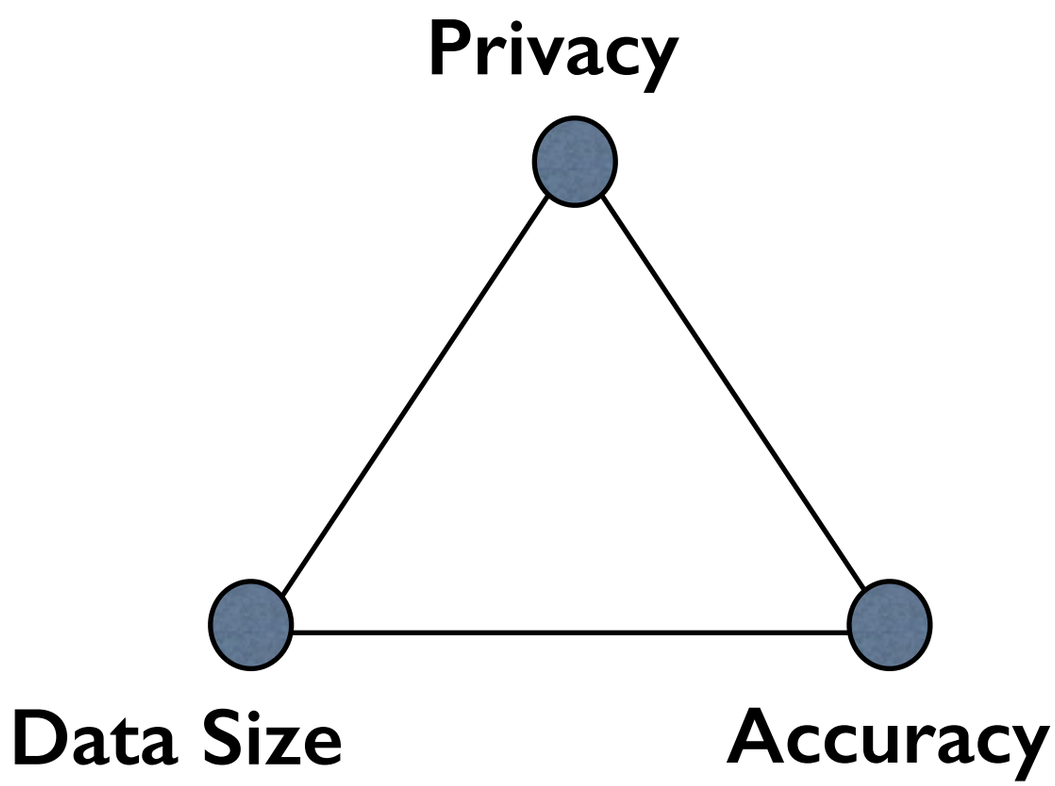
Correlations

Correlation (R^2 values), Alice's DNA reveals:

If Alice is in the **Cancer** set or **Healthy** set

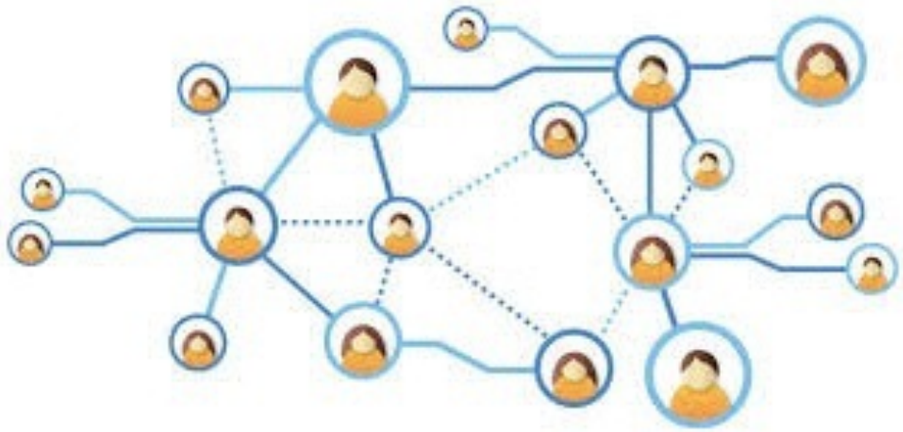
Simply anonymizing data is **unsafe!**

Statistics on small data sets is **unsafe!**

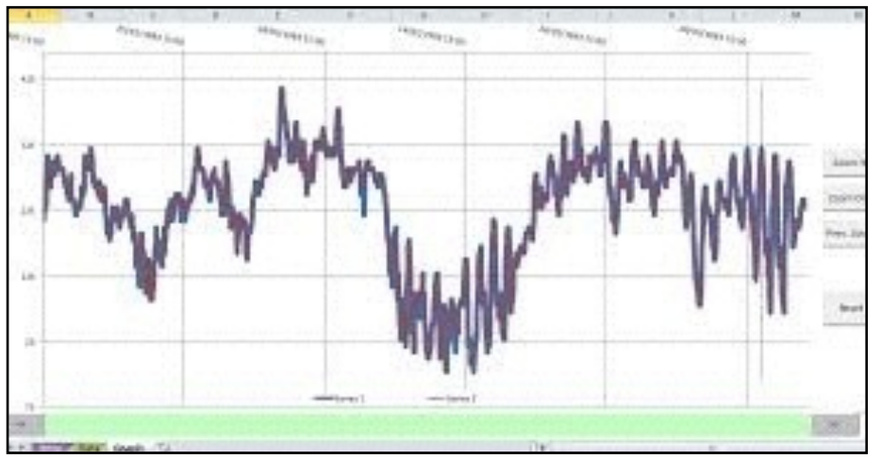


Correlated Data

User information
in social networks



Physical Activity
Monitoring



Why is Privacy Hard for Correlated Data?

Neighbor's information leaks information on user

Talk Agenda:

How do we **learn** from **sensitive** data while still preserving **privacy**?

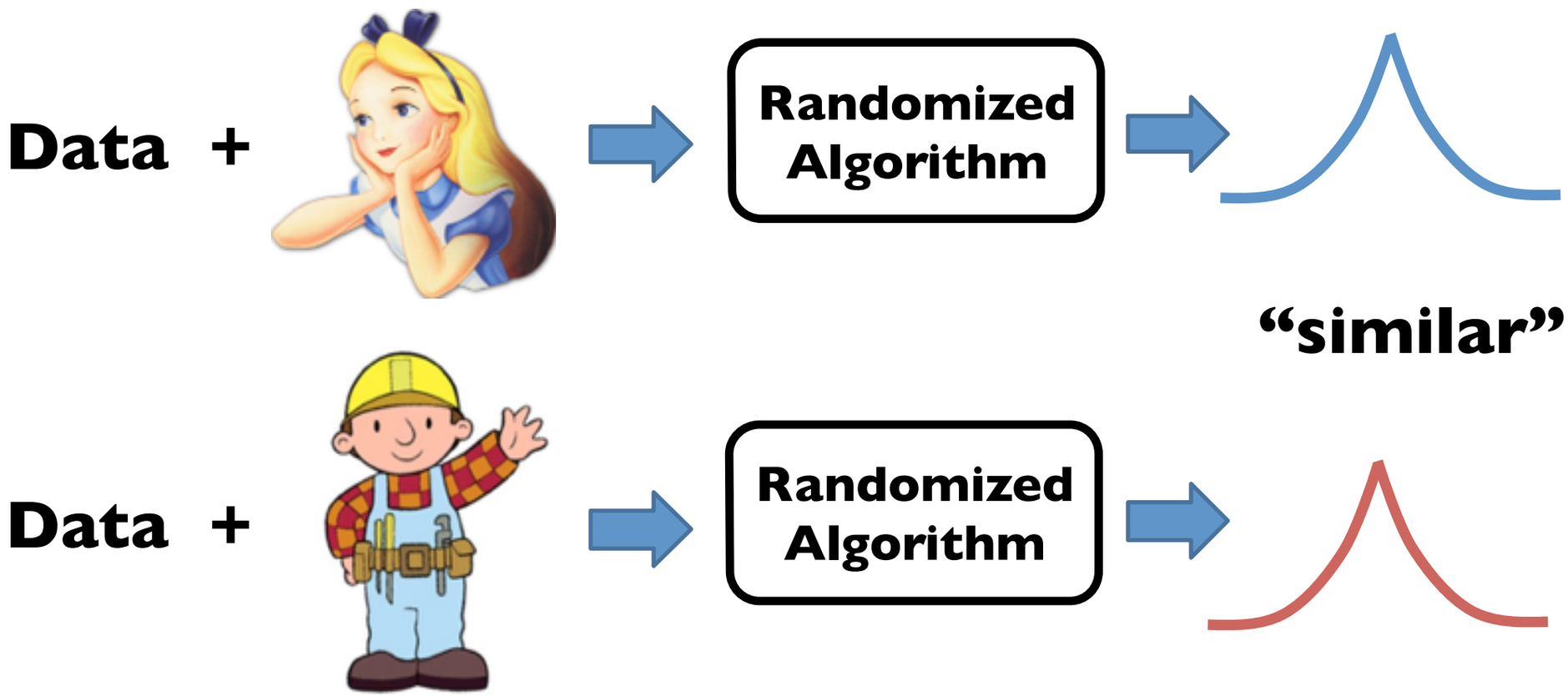
New Directions:

1. Privacy-preserving Bayesian Learning
2. Privacy-preserving statistics on correlated data

Talk Agenda:

1. Privacy for Uncorrelated Data
   - How to define privacy

Differential Privacy [DMNS06]



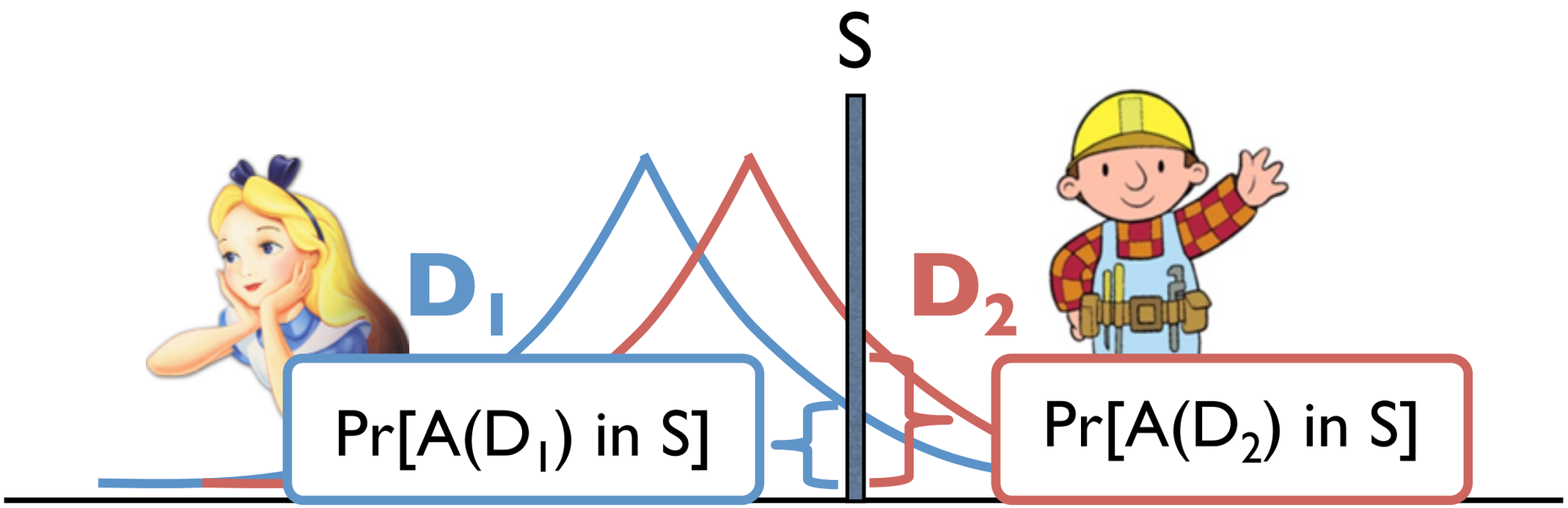
Participation of a single person does not change output

Differential Privacy: Attacker's View

Prior Knowledge + **Algorithm** Output on Data &  = **Conclusion** on 

Prior Knowledge + **Algorithm** Output on Data &  = **Conclusion** on 

Differential Privacy [DMNS06]



For all D_1, D_2 that differ in one person's value, any set S,
If A = ϵ-private randomized algorithm, then:

$$\Pr(A(D_1) \in S) \leq e^{\epsilon} \Pr(A(D_2) \in S)$$

Differential Privacy

1. **Provably strong notion of privacy**

2. **Good approximations for many functions**

   e.g, means, histograms, etc.

Interpretation: Attacker's Hypothesis Test [WZ10, OV13]

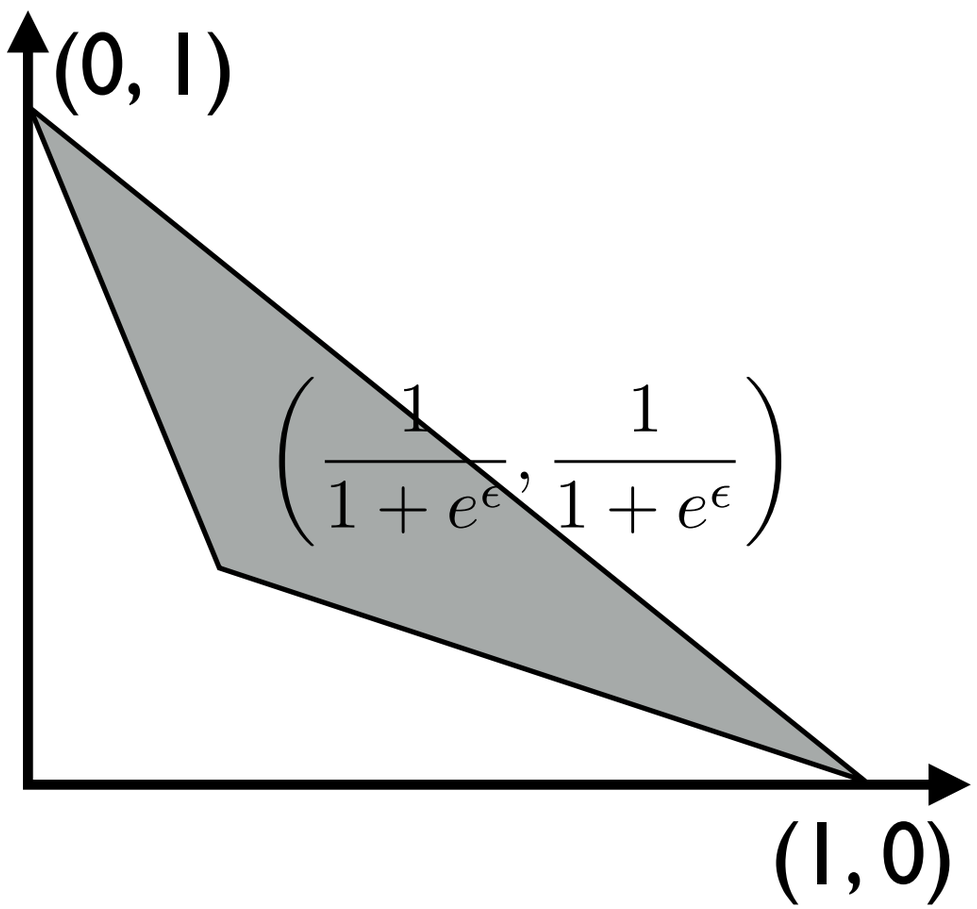
H0: Input to the algorithm = Data + 

H1: Input to the algorithm = Data + 

Failure Events:

False Alarm (FA), Missed Detection (MD)

Interpretation: Attacker's Hypothesis Test [WZ10, OV13]



If algorithm is ϵ-DP

$$\Pr(FA) + e^{\epsilon}\Pr(MD) \geq 1$$

$$e^{\epsilon}\Pr(FA) + \Pr(MD) \geq 1$$

FA = False Alarm

MD = Missed Detection

Talk Agenda:

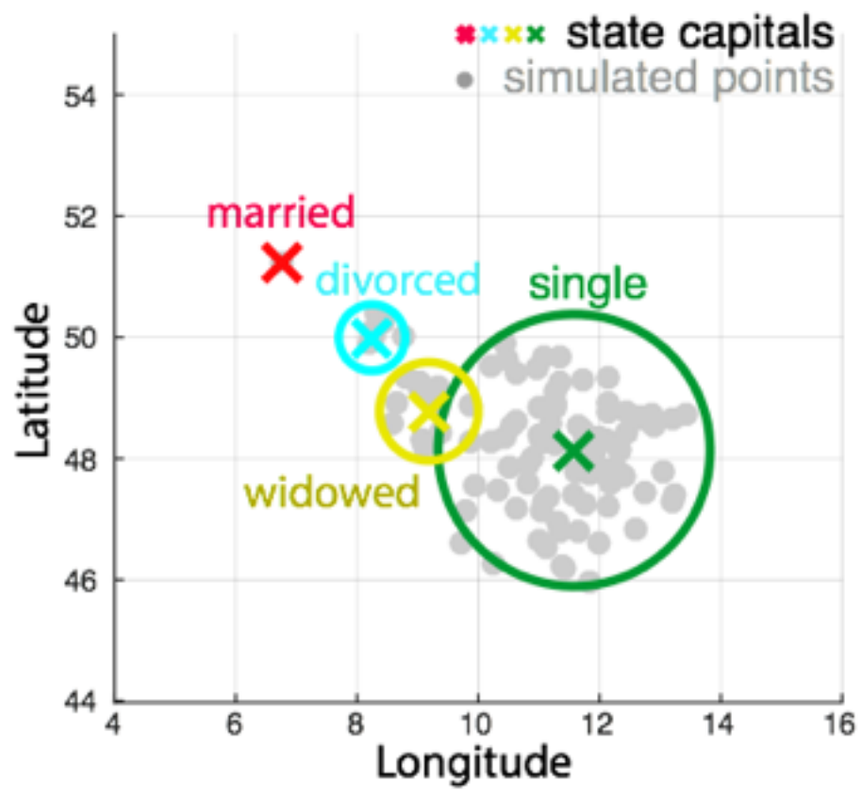
1. Privacy for Uncorrelated Data
   - How to define privacy
   - Privacy-preserving Learning

Example 1: Flu Test

Predicts flu or not, based on patient symptoms

Trained on sensitive patient data

Example 2: Clustering Abortion Data



Given data on abortion locations, cluster by location while preserving privacy of individuals

Bayesian Learning

Bayesian Learning

Data $X = \{ x_1, x_2, \ldots \}$
Model Class Θ

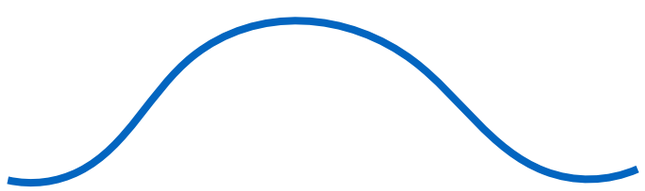
$\}$ Related through likelihood $p(x|\theta)$

Bayesian Learning

Data $X = \{ x_1, x_2, \ldots \}$
Model Class Θ

}

Related through likelihood $p(x|\theta)$



+

Prior $\pi(\theta)$

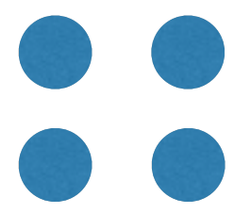
Bayesian Learning

Data X = { $x_1, x_2, \ldots$ }
Model Class Θ

}

Related through likelihood $p(x|\theta)$

+



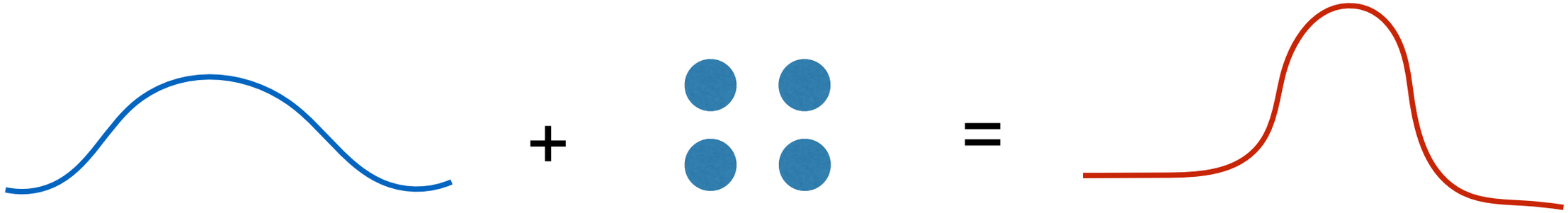
Prior $\pi(\theta)$

Data X

Bayesian Learning

Data X = { $x_1, x_2, \ldots$ }
Model Class Θ

} Related through likelihood $p(x|\theta)$

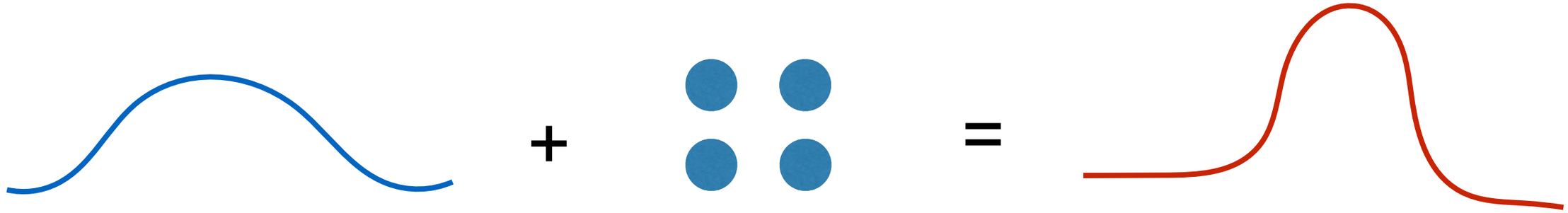


Prior $\pi(\theta)$ + Data X = Posterior $p(\theta|X)$

Bayesian Learning

Data X = { $x_1, x_2, \ldots$ }
Model Class Θ

} Related through likelihood $p(x|\theta)$



Prior $\pi(\theta)$ + Data X = Posterior $p(\theta|X)$

Goal: Output posterior (approx. or samples)

Example: Coin tosses

$X = \{ H, T, H, H \ldots \}$

$\Theta = [0, 1]$

likelihood:

$$p(x|\theta) = \theta^x (1-\theta)^{1-x}$$

Example: Coin tosses

X = { H, T, H, H… }

Θ = [0, 1]

likelihood:

$$p(x|\theta) = \theta^x (1-\theta)^{1-x}$$

+

Prior

$$\pi(\theta) = 1$$

Example: Coin tosses

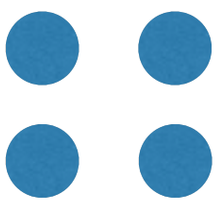
X = { H, T, H, H… }

Θ = [0, 1]

likelihood:

$$p(x|\theta) = \theta^x (1-\theta)^{1-x}$$

+



Prior

$$\pi(\theta) = 1$$

Data X

(h H, t T)

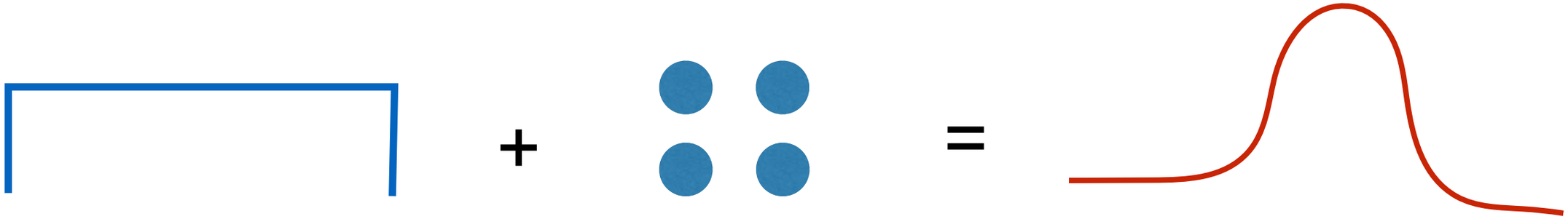
Example: Coin tosses

X = { H, T, H, H… }

Θ = [0, 1]

likelihood:

$$p(x|\theta) = \theta^x (1-\theta)^{1-x}$$



Prior

$$\pi(\theta) = 1$$

Data X

(h H, t T)

Posterior

$$p(\theta|x) \propto \theta^h (1-\theta)^t$$

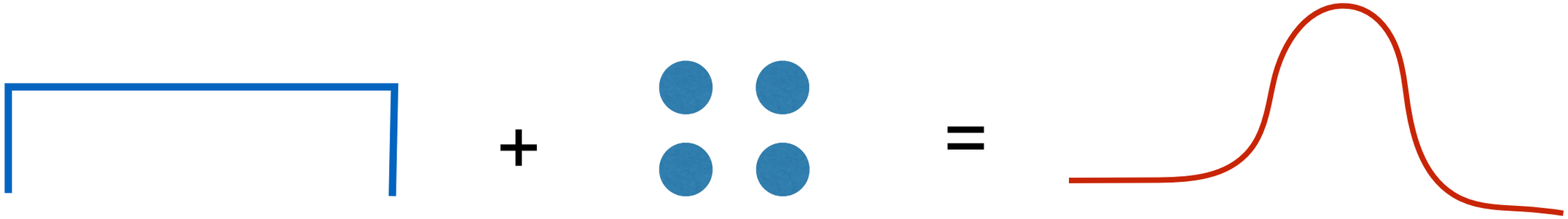
Example: Coin tosses

X = { H, T, H, H… }

Θ = [0, 1]

likelihood:

$$p(x|\theta) = \theta^x (1-\theta)^{1-x}$$



Prior

$$\pi(\theta) = 1$$

Data X

(h H, t T)

Posterior

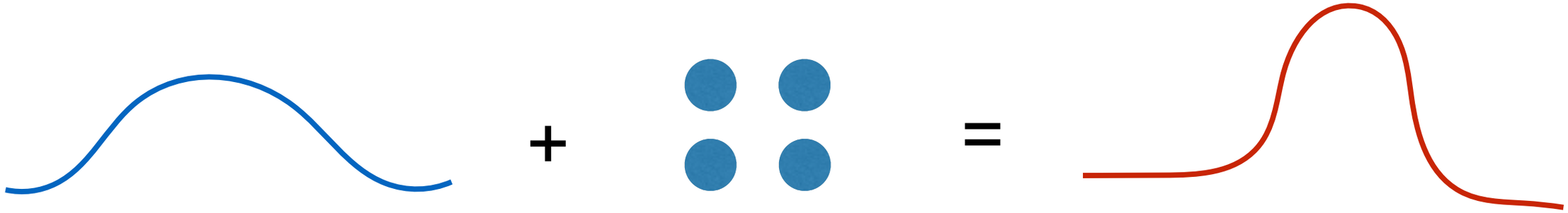
$$p(\theta|x) \propto \theta^h (1-\theta)^t$$

In general, θ is more complex (classifiers, etc)

Private Bayesian Learning

Data X = { $x_1, x_2, \ldots$ }
Model Class Θ

} Related through likelihood $p(x|\theta)$

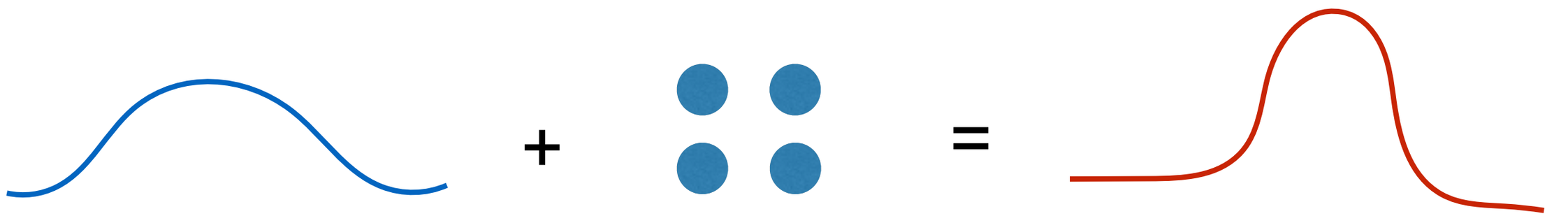


Prior $\pi(\theta)$ + Data X = Posterior $p(\theta|X)$

Private Bayesian Learning

Data X = { $x_1, x_2, \ldots$ }
Model Class Θ

} Related through likelihood $p(x|\theta)$



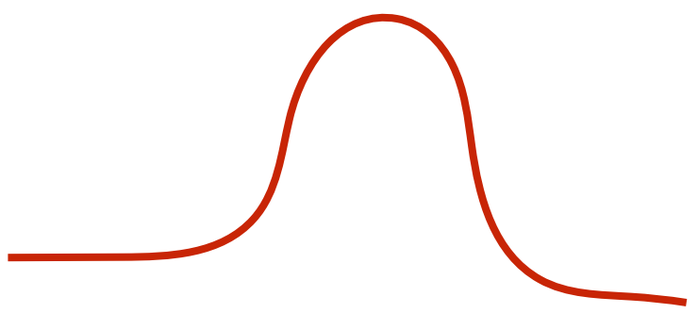
Prior $\pi(\theta)$ Data X Posterior $p(\theta|X)$

Goal: Output **private approx.** to posterior

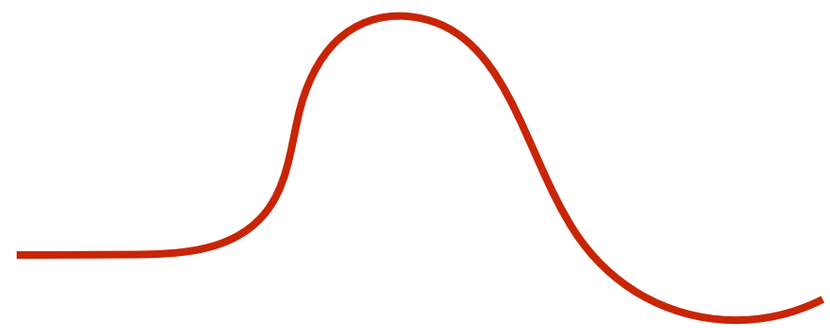
How to make posterior private?

Option 1: Direct posterior sampling [Detal14]

Not private unless under restrictive conditions



$p(\theta|D)$



$p(\theta|D')$

How to make posterior private?

Option 2: Sample from truncated posterior at high temperature [WFS15]

Disadvantage:
Intractable - technically privacy only on convergence
Needs more data/subjects

Our Work: Exponential Families

Exponential family distributions:

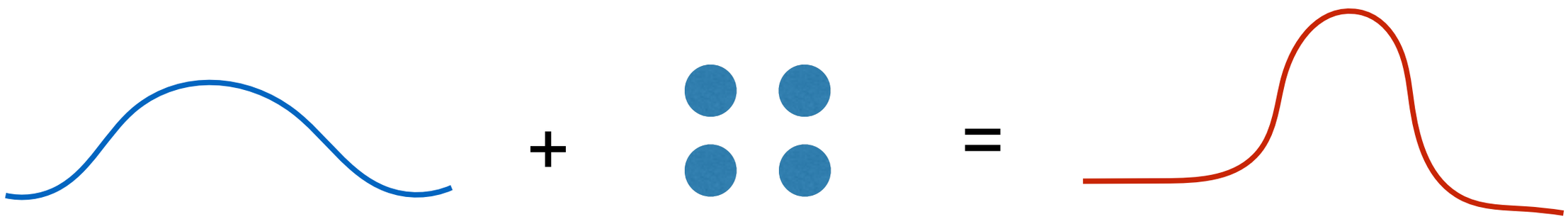
$$p(x|\theta) = h(x)e^{\theta^\top T(x) - A(\theta)}$$

where T is a sufficient statistic

Includes many common distributions like Gaussians, Binomials, Dirichlets, Betas, etc

Properties of Exponential Families

Exponential families have **conjugate priors**



Prior $\pi(\theta)$ Data X Posterior $p(\theta|X)$

$p(\theta|x)$ is in the same distribution class as $\pi(\theta)$

eg, Gaussians-Gaussians, Beta-Binomial, etc

Sampling from Exponential Families

(Non-private) posterior comes from exp. family:

$$p(\theta|x) \propto e^{\eta(\theta)^\top (\sum_i T(x_i)) - B(\theta)}$$

given data $x_1, x_2, \ldots$

Private Sampling:

1. If T is bounded, add noise to $\sum_i T(x_i)$ to get private version T'

2. Sample from the perturbed posterior:

$$p(\theta|x) \propto e^{\eta(\theta)^\top T' - B(\theta)}$$

Performance

- Theoretical Guarantees
- Experiments

Theoretical Guarantees

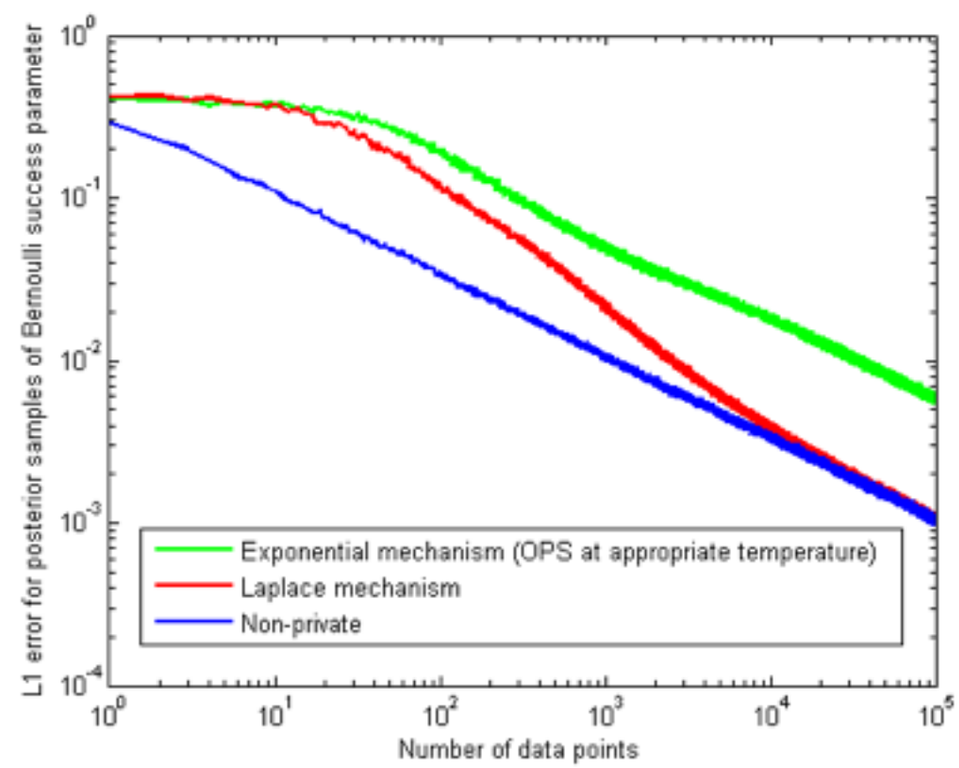
Performance Measure: Asymptotic Relative Efficiency

(Lower = more sample efficient for large n)

Non-private: 2

Our Method: 2

[WFS15]: $\max(2, 1 + 1/\epsilon)$



Experiments - Task

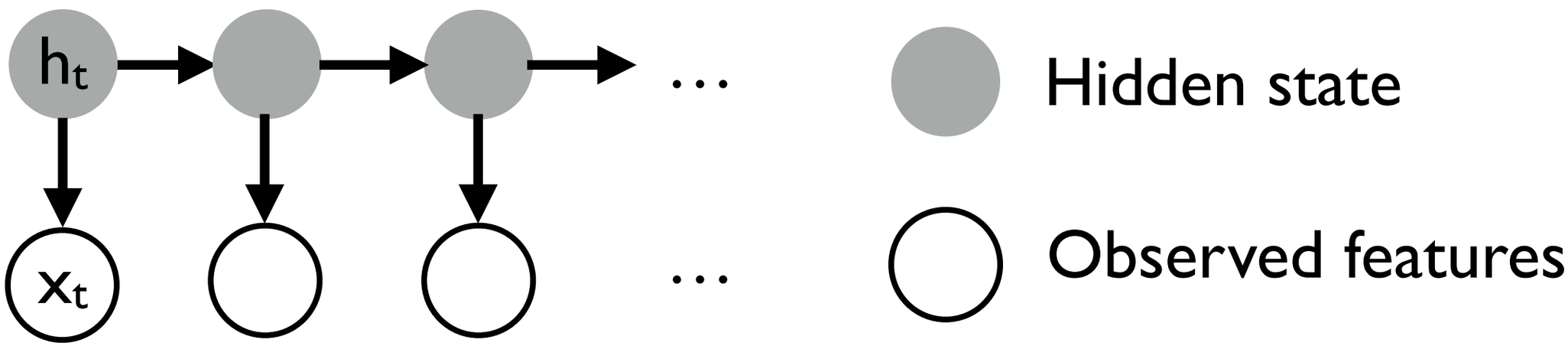
Task: Time series clustering of events in Wikileaks war logs while preserving event-level privacy

Data: War-log entries - Afghanistan (75K), Iraq (390K)

Goal: Cluster entries in each region based on features (casualty counts, enemy/friendly fire, explosive hazards, etc…)

Experiments - Model

Hidden Markov Model for each region



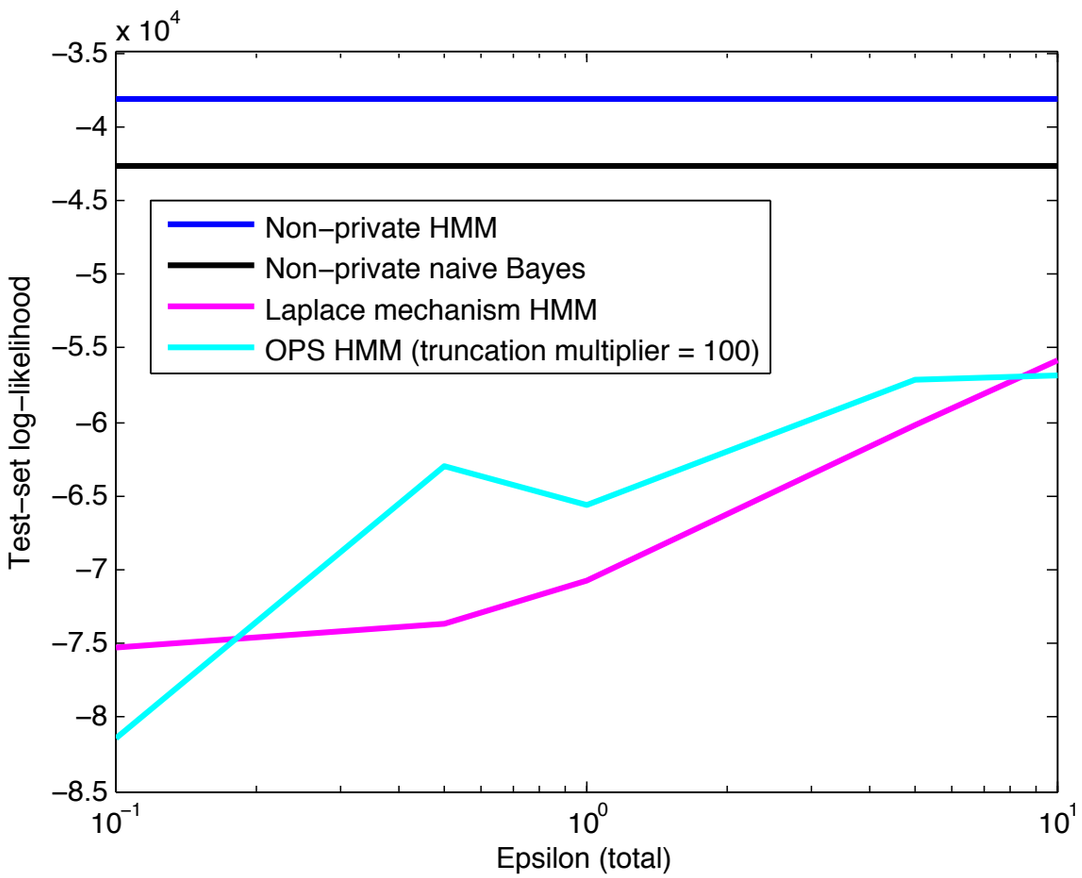
Discrete states (h_t) and observations (x_t)

Transition parameters T: $T_{ij} = P(h_{t+1} = i \mid h_t = j)$

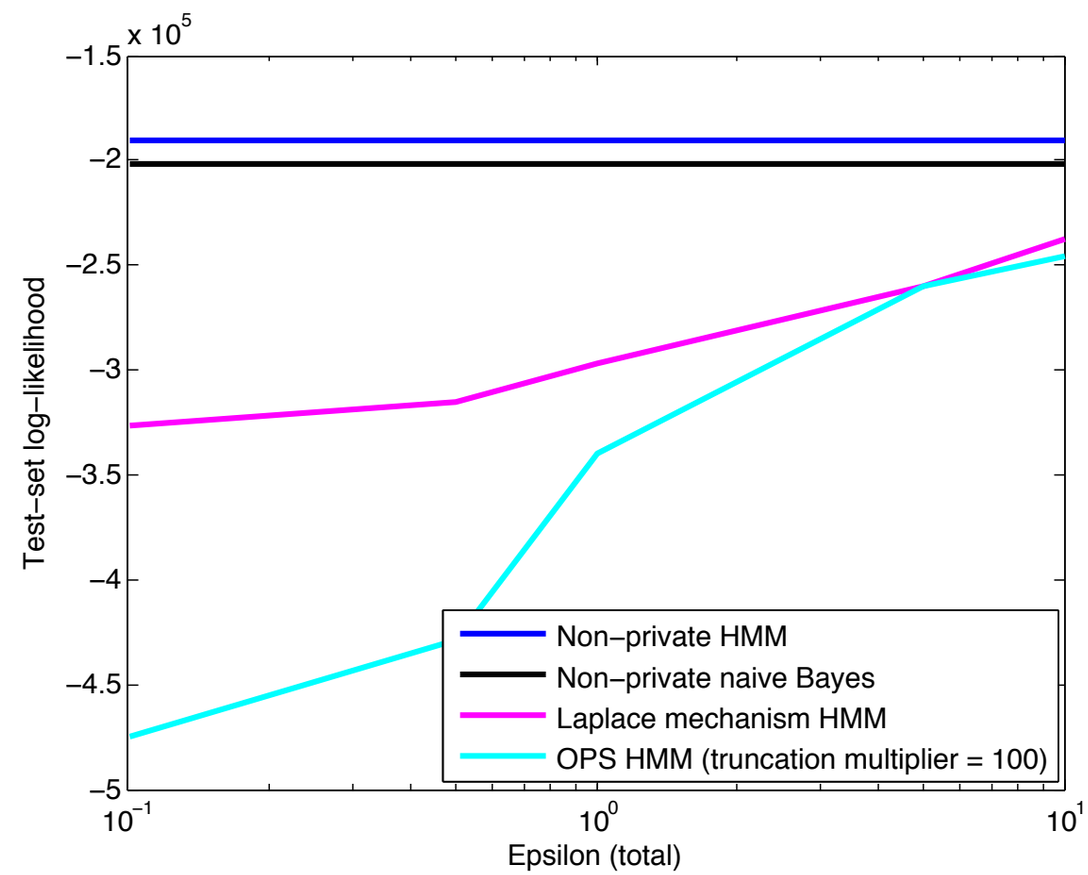
Emission parameters O: where $O_{ij} = P(x_t = i \mid h_t = j)$

Goal: Sample from posterior P(O| data)
(in the exponential family)

Experiments - Results



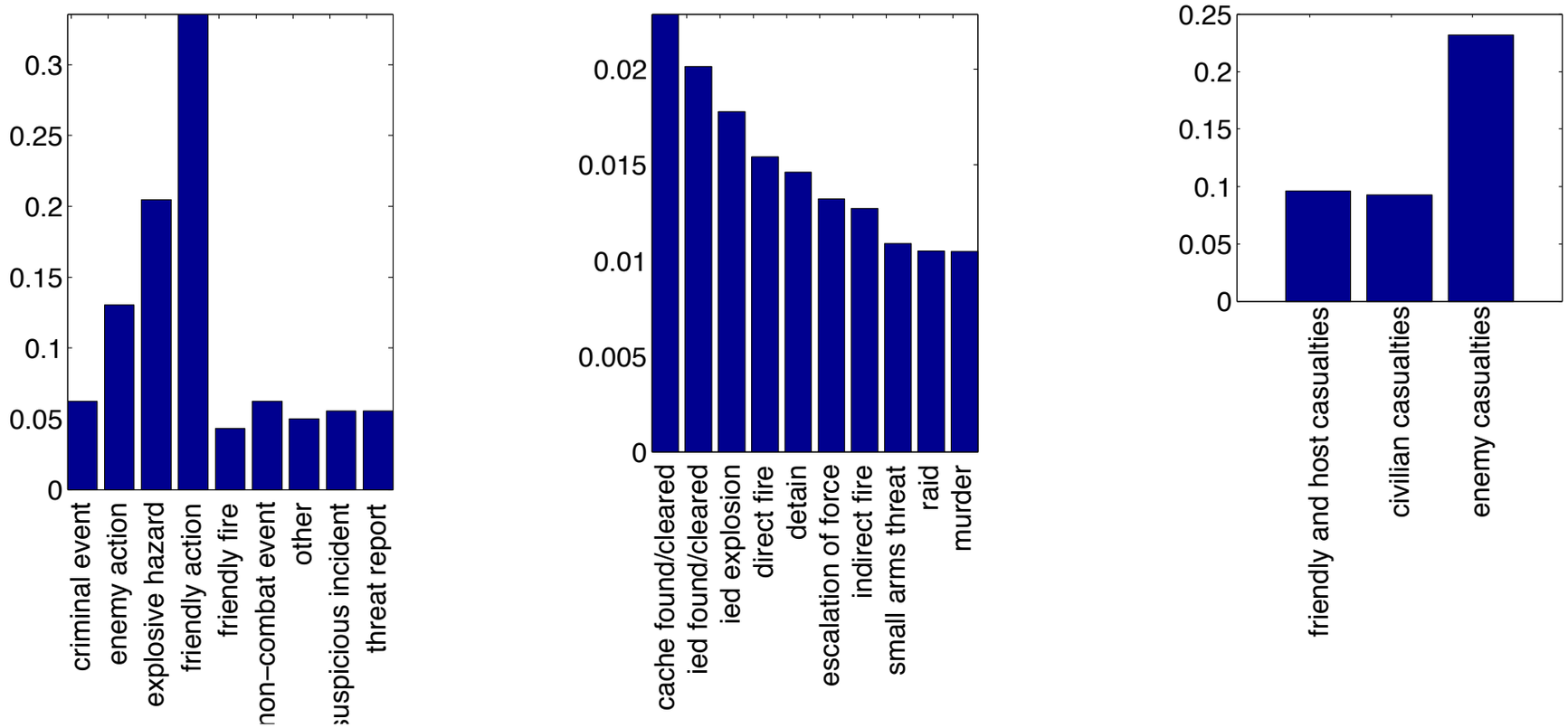
Afghanistan



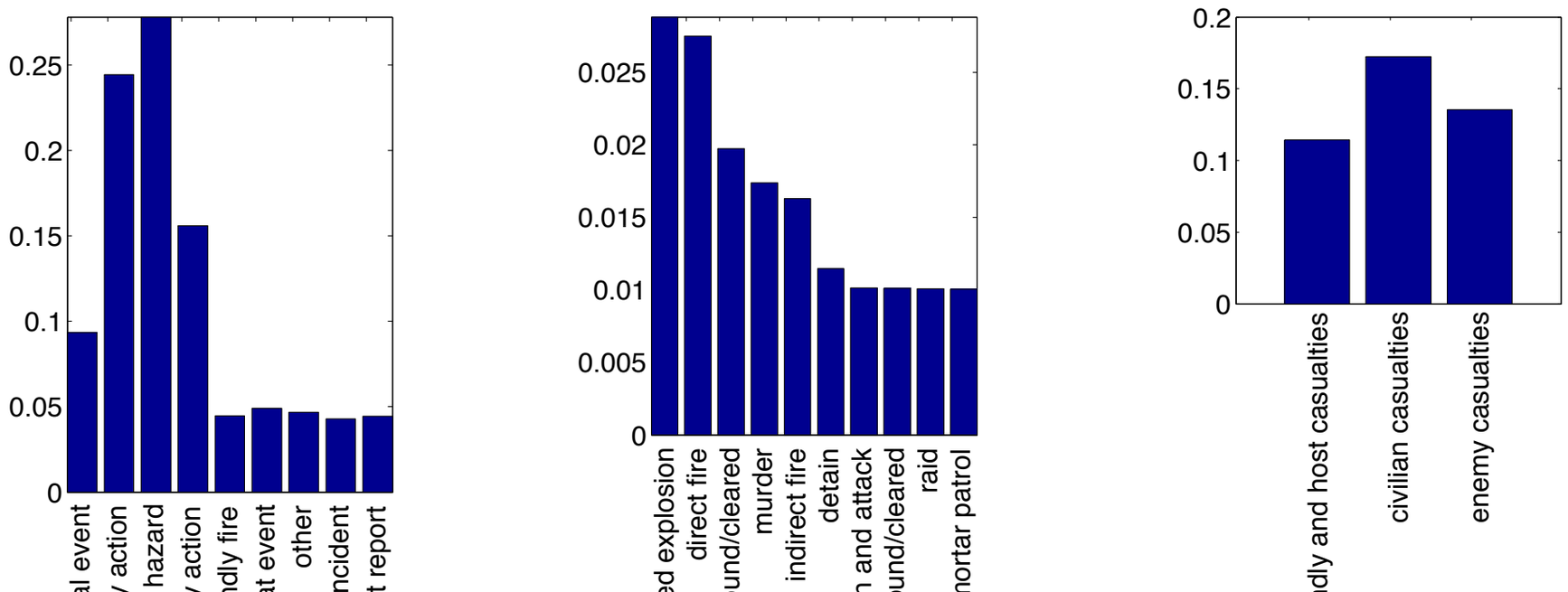
Iraq

Experiments - States

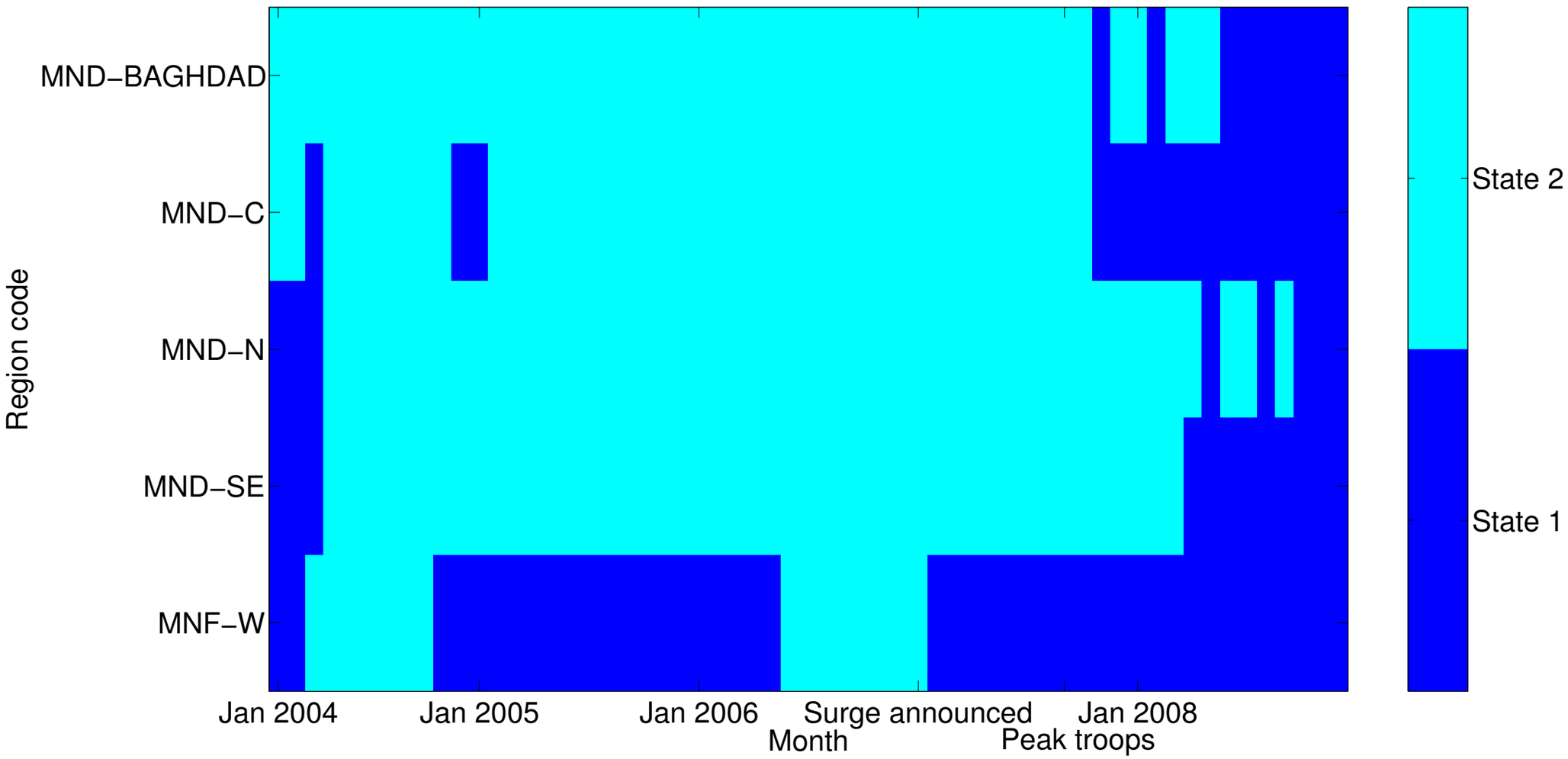
Iraq
State 1



Iraq
State 2



Experiments - Clustering



Conclusion

New method for private posterior sampling from exponential families

Open Problems:

1. Private sampling from more complex posteriors

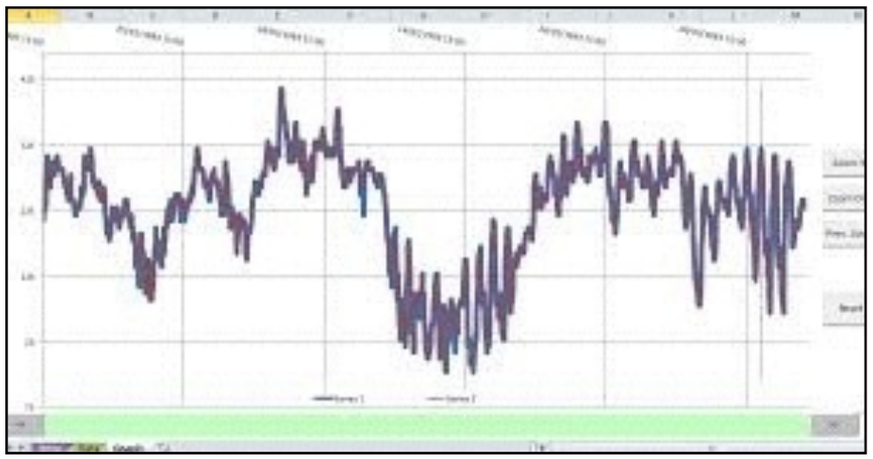
2. Private versions of other Bayesian posterior approximation schemes (variational Bayes, etc)

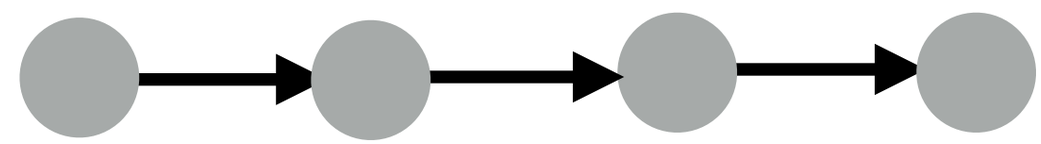
3. Combining Bayesian inference with more relaxed forms of DP (eg, concentrated DP, distributional DP, etc)

Talk Agenda:

1. Privacy for Uncorrelated Data
   - How to define privacy
   - Privacy-preserving Bayesian Learning
2. Privacy for Correlated Data

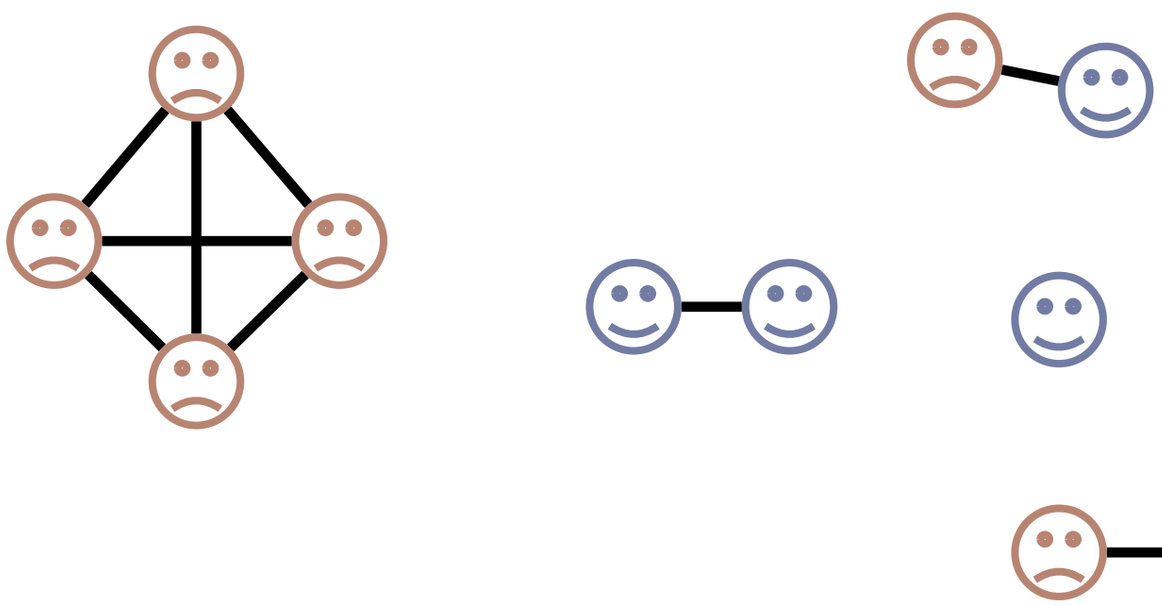
Example 1: Activity Monitoring





Share aggregate data on physical activity with doctor or provider, while hiding activity at each specific time

Example 2: Spread of Flu in Network



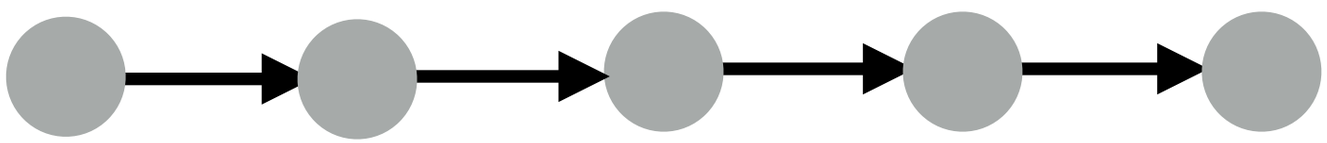
Interaction Network

Publish aggregate statistics, preserve individual privacy

Why is Differential Privacy not Enough for Correlated data?

Example: Activity Monitoring

$D = (x_1, .., x_T)$, x_t = activity at time t

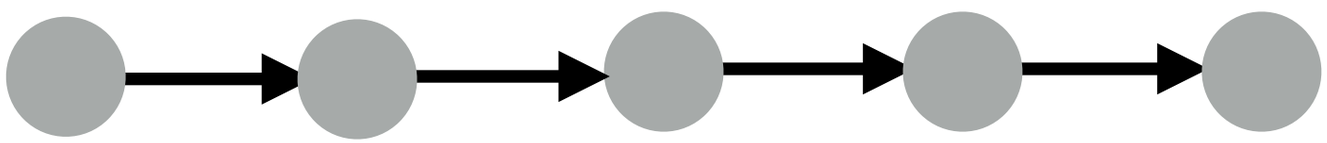


Correlation Network

Goal: (1) Publish activity histogram
(2) Prevent adversary from knowing activity at t

Example: Activity Monitoring

$D = (x_1, .., x_T)$, x_t = activity at time t

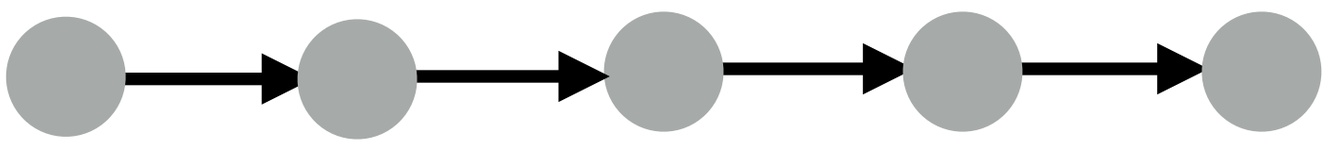


Correlation Network

1-DP: Output histogram of activities + noise with stdev 1

Example: Activity Monitoring

$D = (x_1, .., x_T)$, x_t = activity at time t



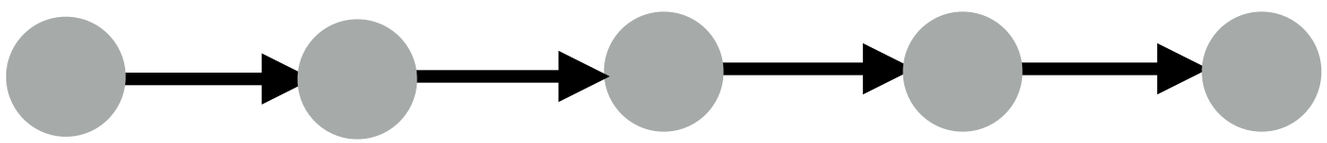
Correlation Network

1-DP: Output histogram of activities + noise with stdev 1

Not enough - activities across time are highly correlated!

Example: Activity Monitoring

$D = (x_1, .., x_T)$, x_t = activity at time t



Correlation Network

1-Group DP: Output histogram of activities + noise with stdev T

Too much noise - no utility!

Talk Agenda:

1. Privacy for Uncorrelated Data
   - How to define privacy
   - Privacy-preserving Classification

2. Privacy for Correlated Data
   - How to define privacy

Pufferfish Privacy [KM12]

Secret Set S

S: Information to be protected

e.g: Alice's age is 25, Bob has a disease

Pufferfish Privacy [KM12]

Secret Set S

Secret Pairs Set Q

Q: Pairs of secrets we want to be indistinguishable

e.g: (Alice's age is 25, Alice's age is 40)

(Bob is in dataset, Bob is not in dataset)

Pufferfish Privacy [KM12]

Secret Set S	Secret Pairs Set Q	Distribution Class Θ

Θ: A set of distributions that plausibly generate the data

e.g: (connection graph G, disease transmits w.p $[0.1, 0.5]$)

(Markov Chain with transition matrix in set **P**)

May be used to model correlation in data

Pufferfish Privacy [KM12]

Secret Set S

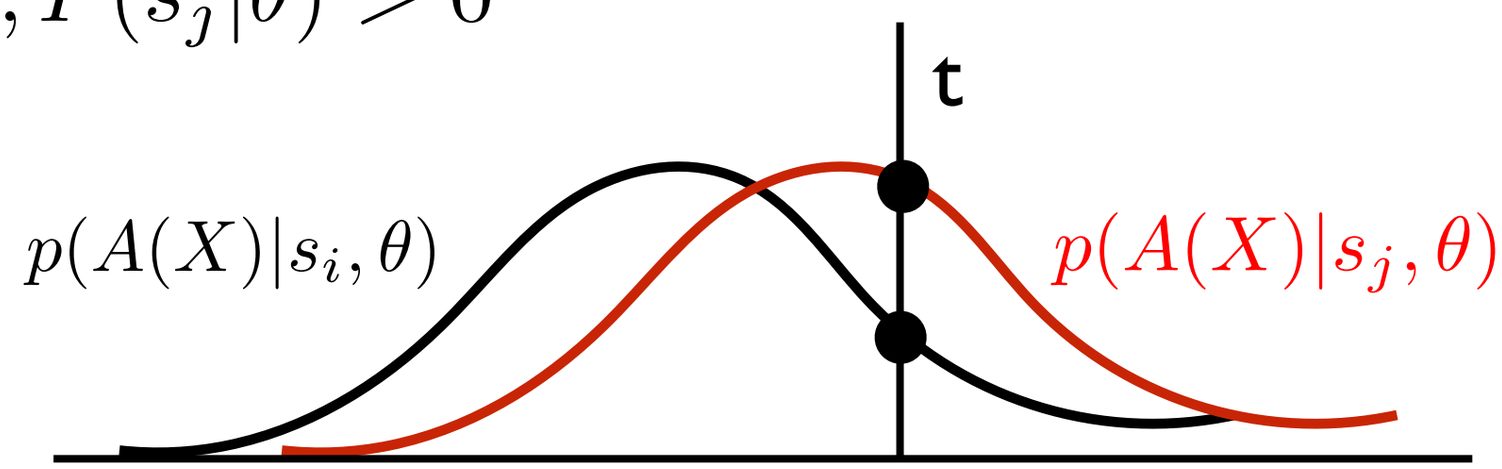
Secret Pairs Set Q

Distribution Class Θ

An algorithm A is ϵ-Pufferfish private with parameters (S, Q, Θ) if for all (s_i, s_j) in Q, for all $\theta \in \Theta$, $X \sim \theta$, all t,

$$p_{\theta,A}(A(X) = t | s_i, \theta) \leq e^{\epsilon} \cdot p_{\theta,A}(A(X) = t | s_j, \theta)$$

whenever $P(s_i|\theta), P(s_j|\theta) > 0$



Pufferfish Generalizes DP [KM12]

Theorem: Pufferfish = Differential Privacy when:

$S = \{ \; s_{i,a} :=$ Person i has value a, for all i, all a in domain X $\}$

$Q = \{ \; (s_{i,a} \; s_{i,b})$, for all i and (a, b) pairs in X x X $\}$

$\Theta = \{$ Distributions where each person i is independent $\}$

Pufferfish Generalizes DP [KM12]

Theorem: Pufferfish = Differential Privacy when:

$S = \{ \; s_{i,a} :=$ Person i has value a, for all i, all a in domain X $\}$

$Q = \{ \; (s_{i,a} \; s_{i,b})$, for all i and (a, b) pairs in X x X $\}$

$\Theta = \{$ Distributions where each person i is independent $\}$

Theorem: No utility possible when:

$\Theta = \{$ All possible distributions $\}$

Talk Agenda:

1. Privacy for Uncorrelated Data
   - How to define privacy
   - Privacy-preserving Classification

2. Privacy for Correlated Data
   - How to define privacy
   - Privacy-preserving Statistics

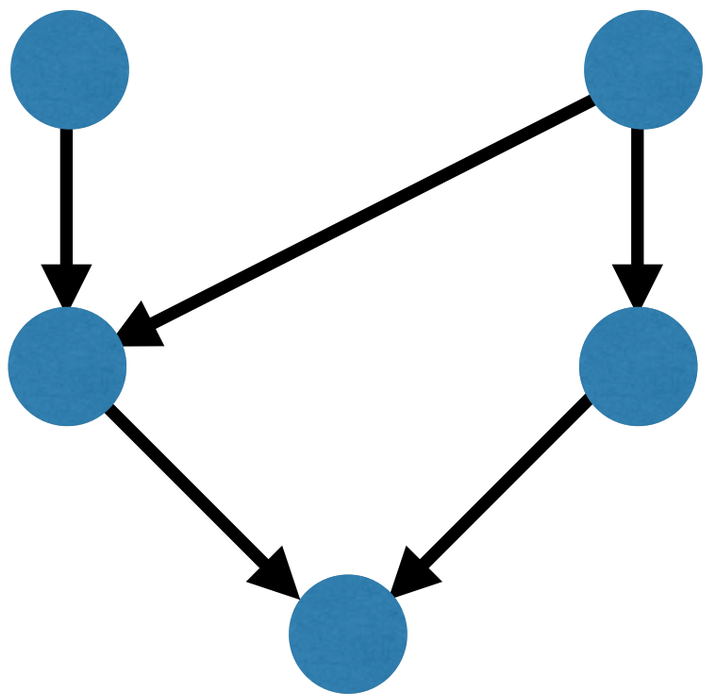
How to get Pufferfish privacy?

Special case mechanisms [KM12, HMD12]

Is there a **more general** Pufferfish mechanism for a large class of correlated data?

Our work: Yes, the **Markov Quilt Mechanism**

Correlation Measure: Bayesian Networks



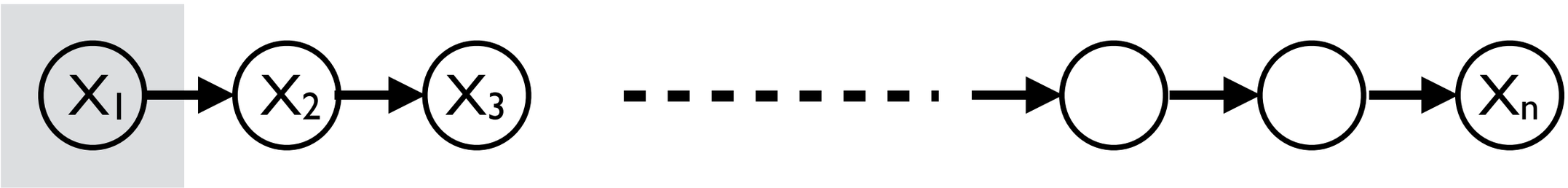
Node: variable

Directed Acyclic Graph

Joint distribution of variables:

$$\Pr(X_1, X_2, \ldots, X_n) = \prod_i \Pr(X_i | \text{parents}(X_i))$$

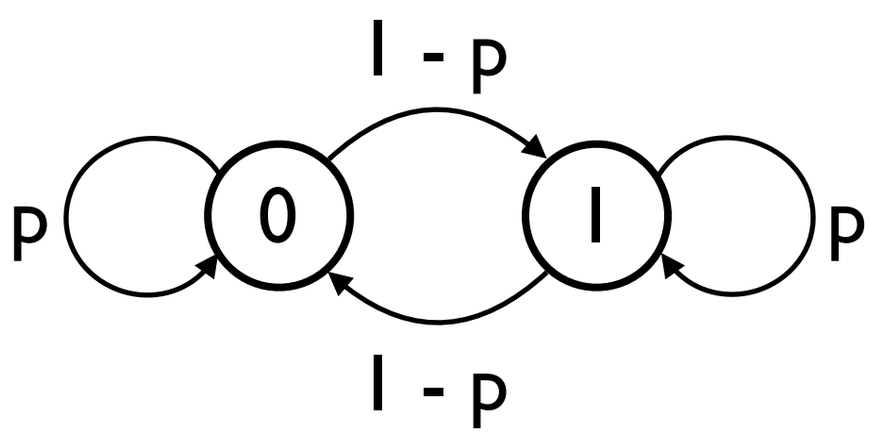
A Simple Example



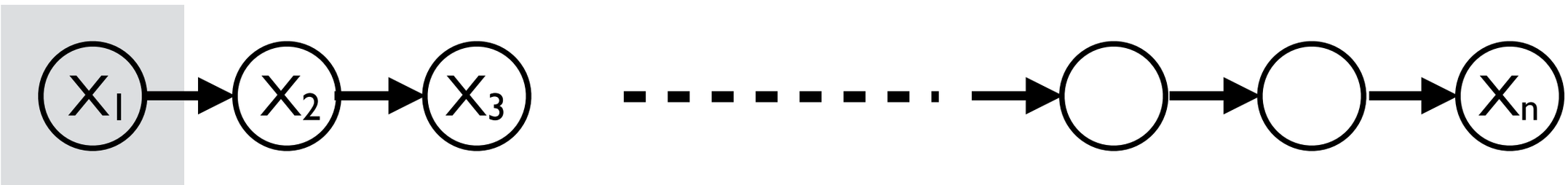
Model:

X_i in {0, 1}

State Transition Probabilities:



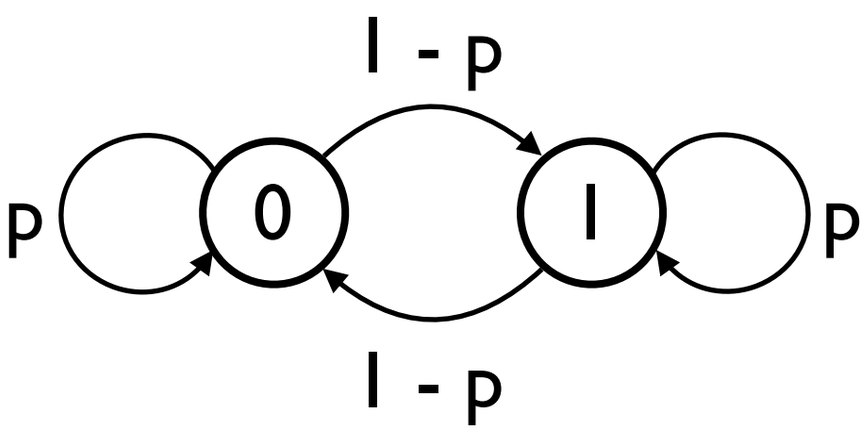
A Simple Example



Model:

X_i in {0, 1}

State Transition Probabilities:

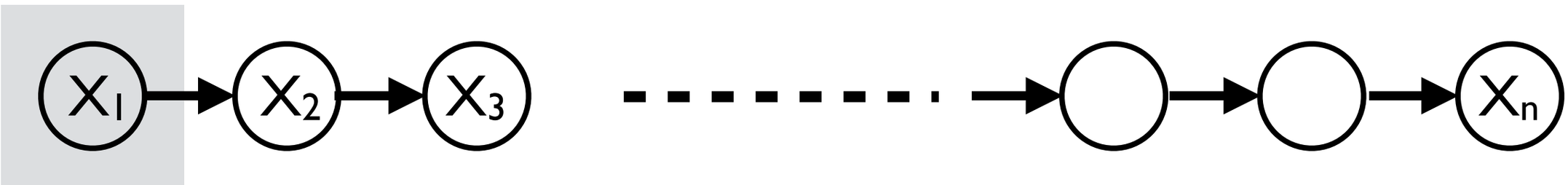


$Pr(X_2 = 0 | X_1 = 0) = p$

$Pr(X_2 = 0 | X_1 = 1) = 1 - p$

….

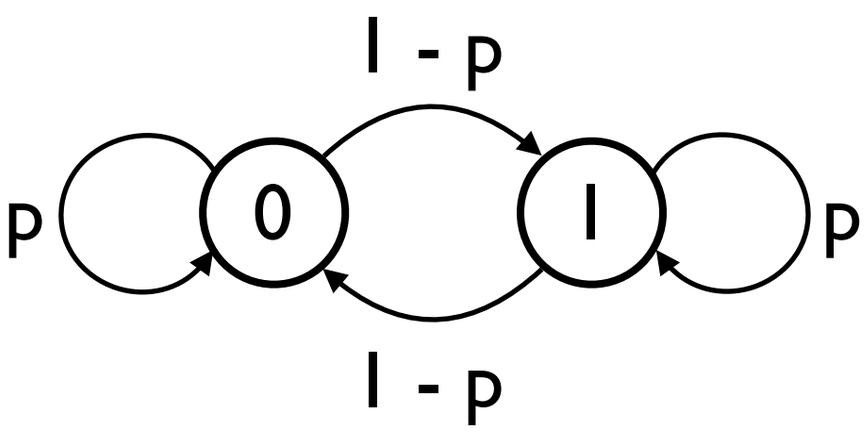
A Simple Example



Model:

X_i in {0, 1}

State Transition Probabilities:



$\Pr(X_2 = 0 \mid X_1 = 0) = p$

$\Pr(X_2 = 0 \mid X_1 = 1) = 1 - p$

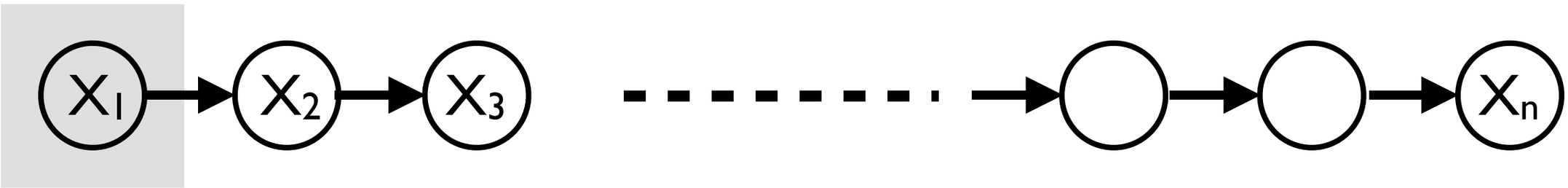
....

$\Pr(X_i = 0 \mid X_1 = 0) = \frac{1}{2} + \frac{1}{2}(2p-1)^{i-1}$

$\Pr(X_i = 0 \mid X_1 = 1) = \frac{1}{2} - \frac{1}{2}(2p-1)^{i-1}$

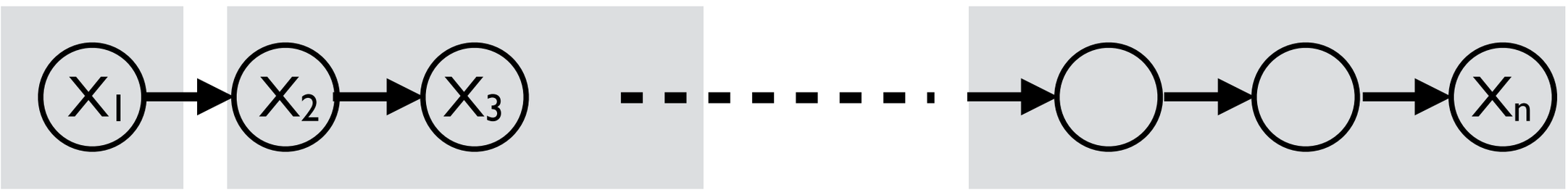
Influence of X_1 diminishes with distance

Algorithm: Main Idea



Goal: Protect X_1

Algorithm: Main Idea

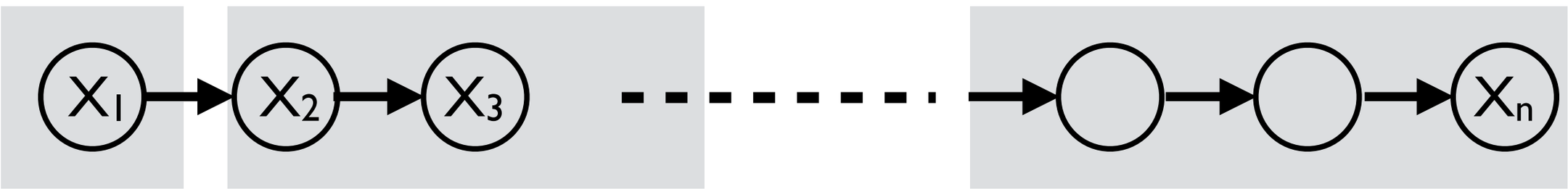


Local nodes
(high correlation)

Rest
(almost independent)

Goal: Protect X_1

Algorithm: Main Idea



Local nodes
(high correlation)

Rest
(almost independent)

Goal: Protect X_1

Add noise to hide local nodes **+** Small correction for rest

Measuring "Independence"

Max-influence of X_i on a set of nodes X_R:

$$e(X_R|X_i) = \max_{a,b} \sup_{\theta \in \Theta} \max_{x_R} \log \frac{\Pr(X_R = x_R | X_i = a, \theta)}{\Pr(X_R = x_R | X_i = b, \theta)}$$

Low $e(X_R|X_i)$ means X_R is almost independent of X_i

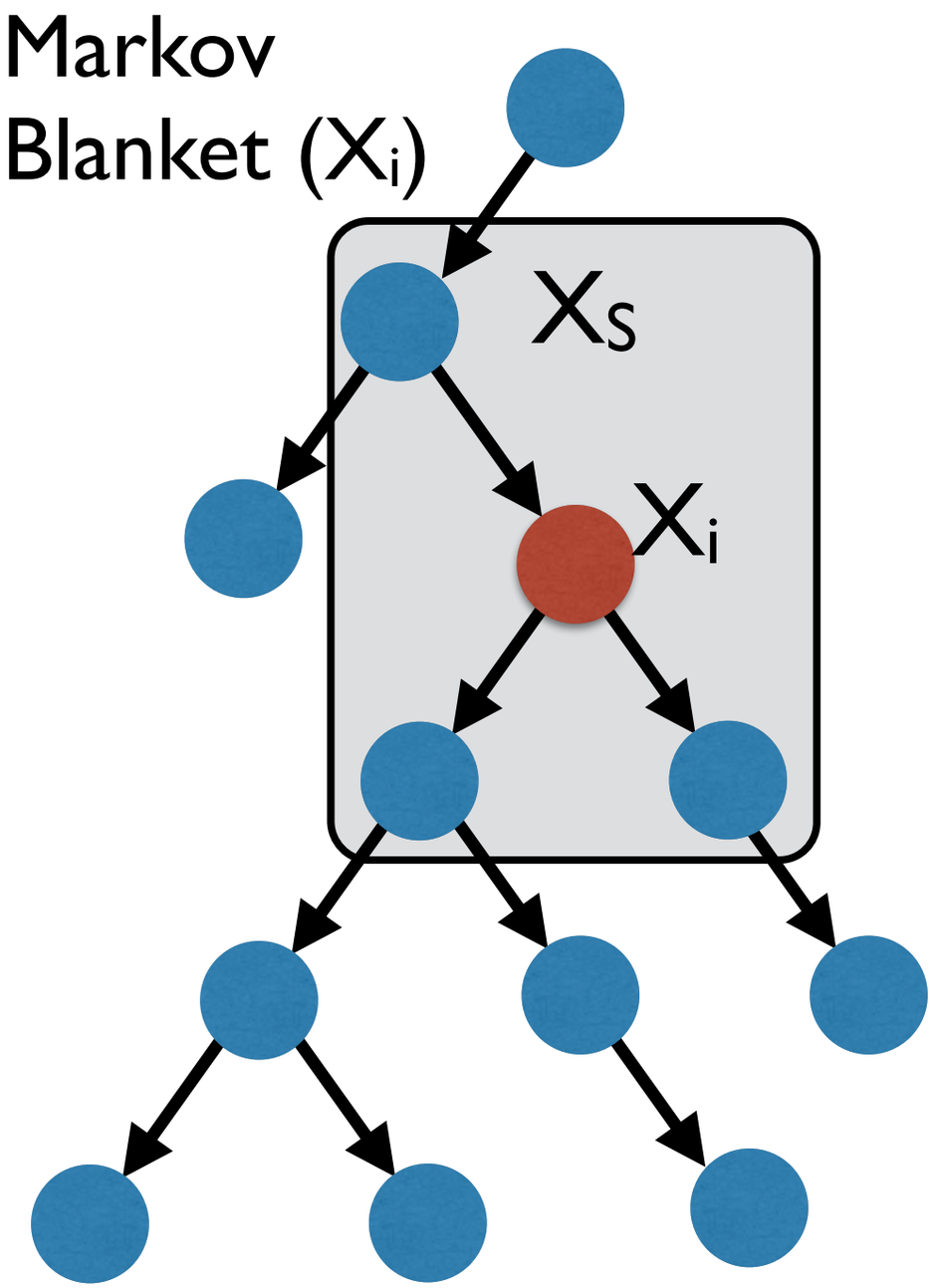
To protect X_i, correction term needed for X_R is $\exp(e(X_R|X_i))$

How to find large “almost independent” sets

Brute force search is expensive

Use structural properties of the Bayesian network

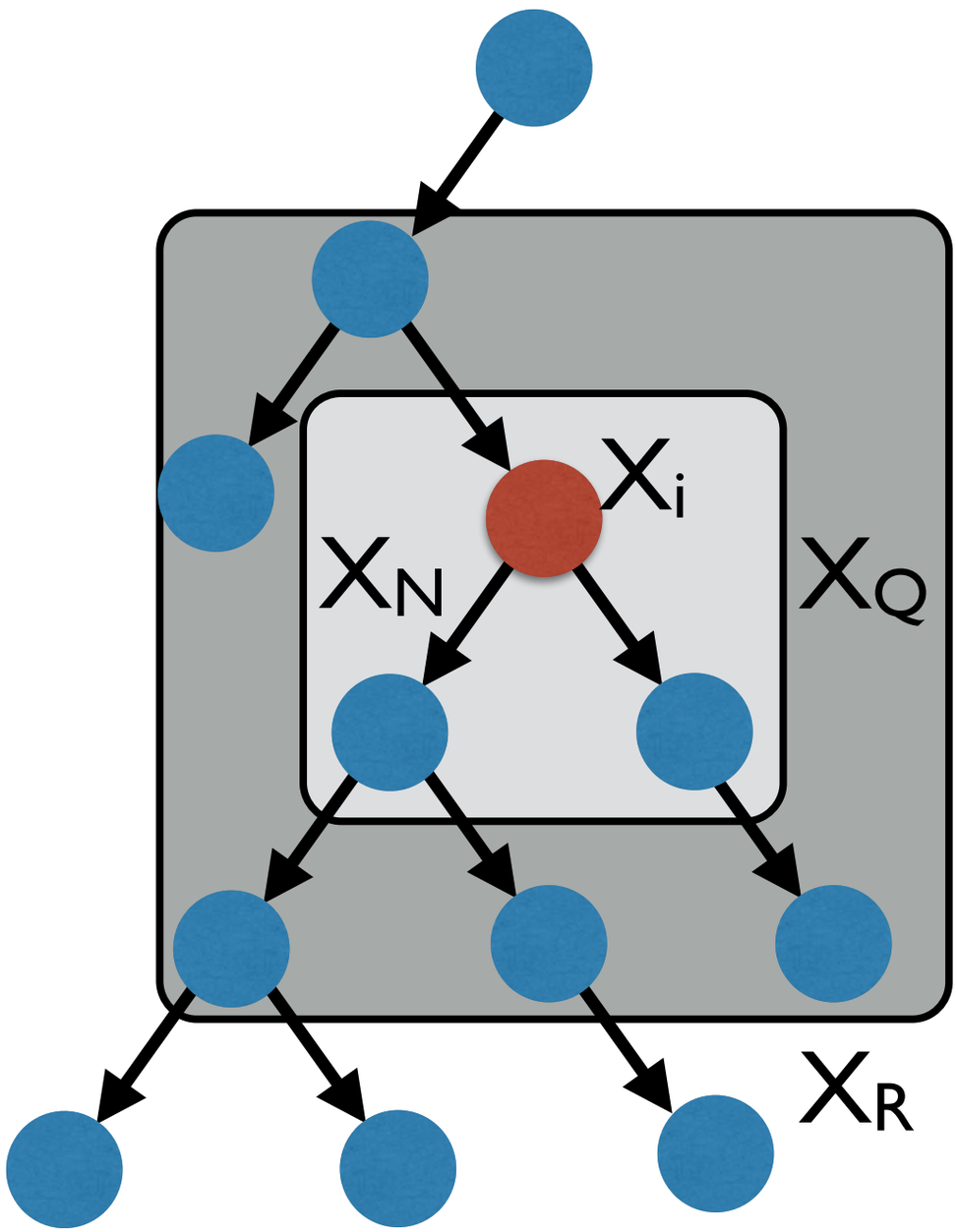
Markov Blanket



Markov Blanket(X_i) =
Set of nodes X_S s.t Xi is
independent of $X \setminus (X_i \cup X_S)$
given X_S

(usually, parents, children,
other parents of children)

Define: Markov Quilt

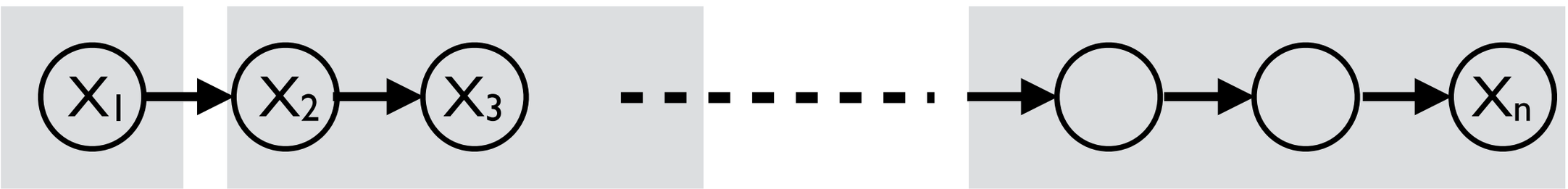


X_Q is a Markov Quilt of X_i if:

1. Deleting X_Q breaks graph into X_N and X_R
2. X_i lies in X_N
3. X_R is independent of X_i given X_Q

(For Markov Blanket $X_N = X_i$)

Recall: Algorithm



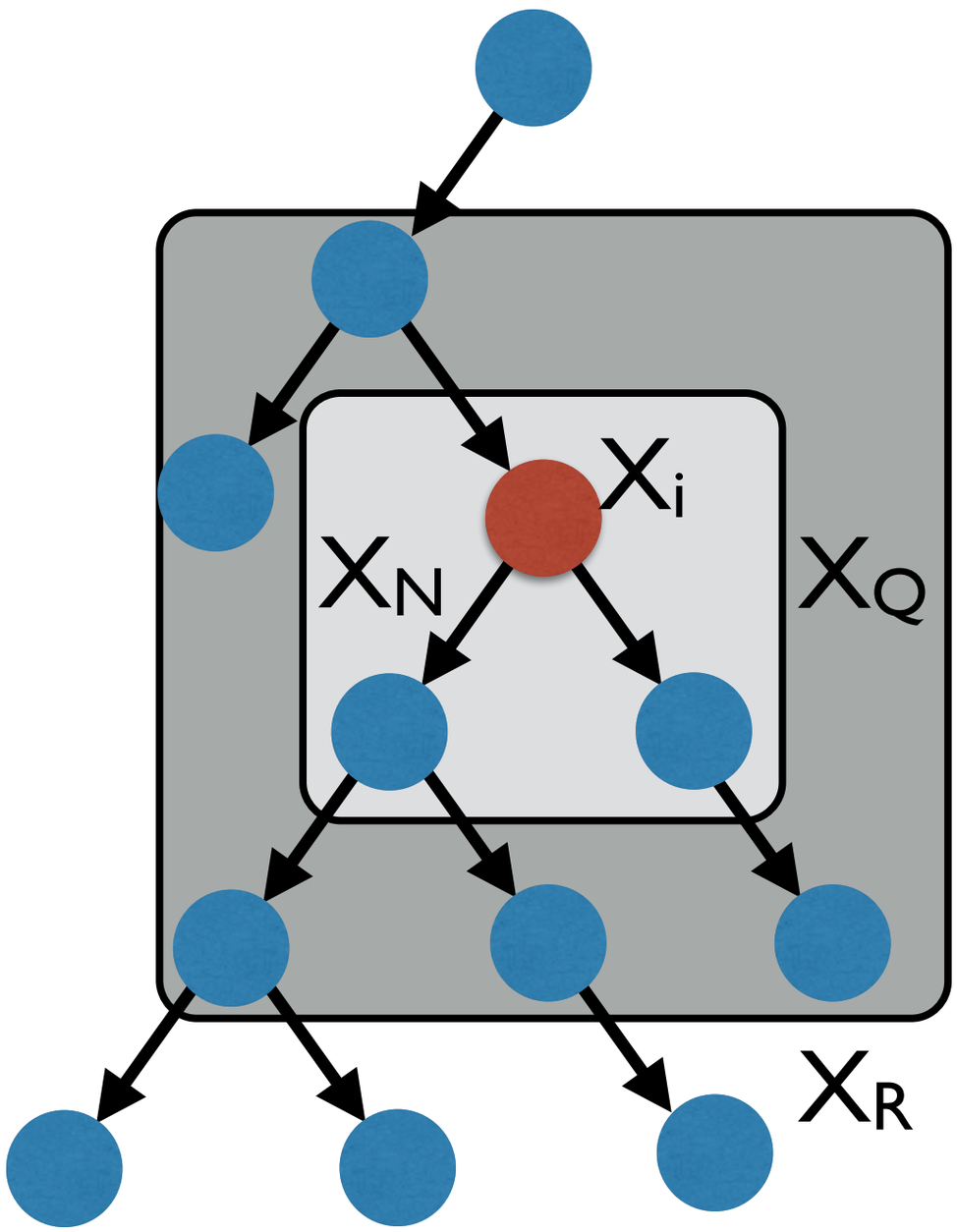
Local nodes
(high correlation)

Rest
(almost independent)

Goal: Protect X_1

Add noise to hide local nodes **+** Small correction for rest

Why do we need Markov Quilts?

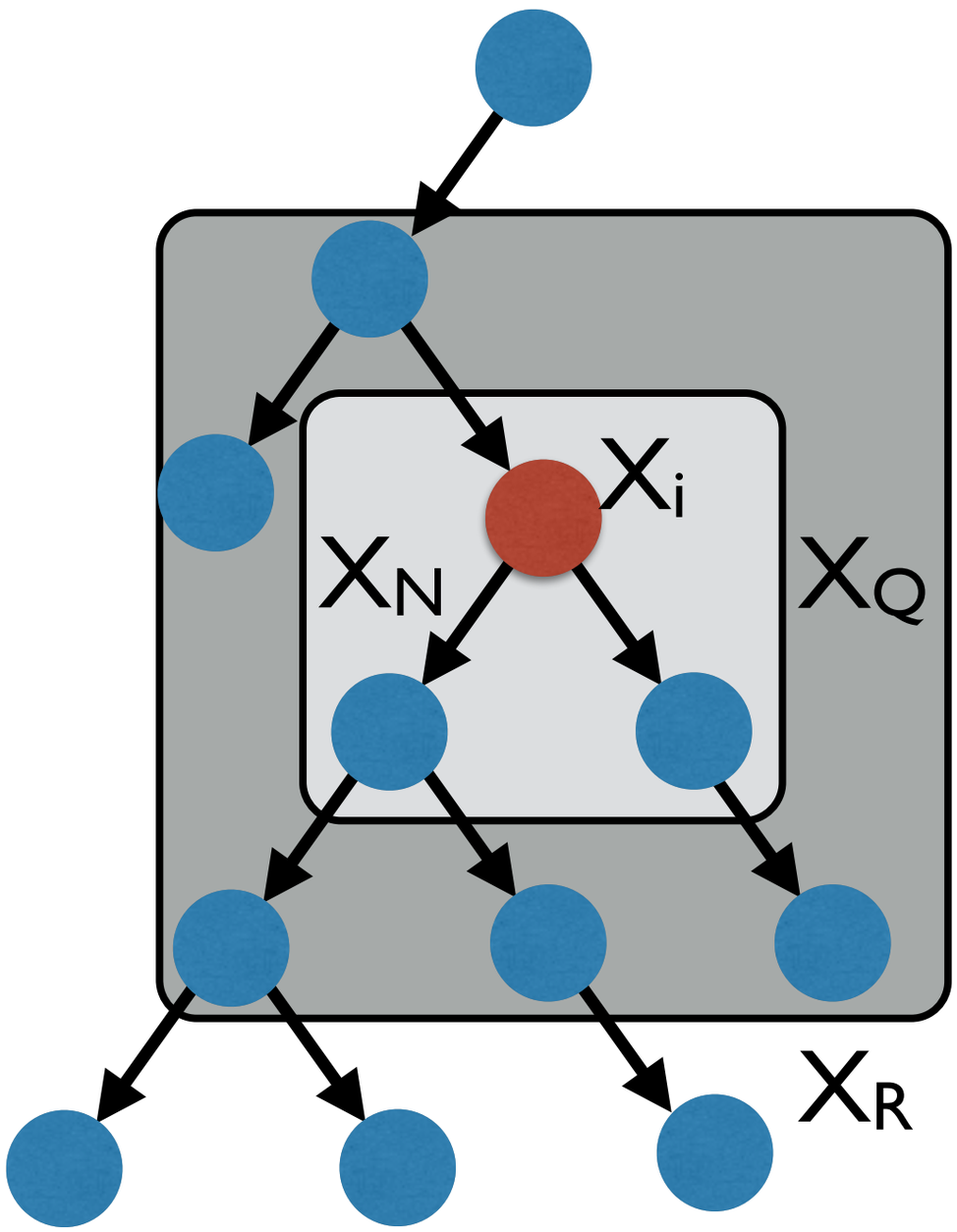


Given a Markov Quilt,

X_N = local nodes for X_i

$X_Q \cup X_R$ = rest

Why do we need Markov Quilts?



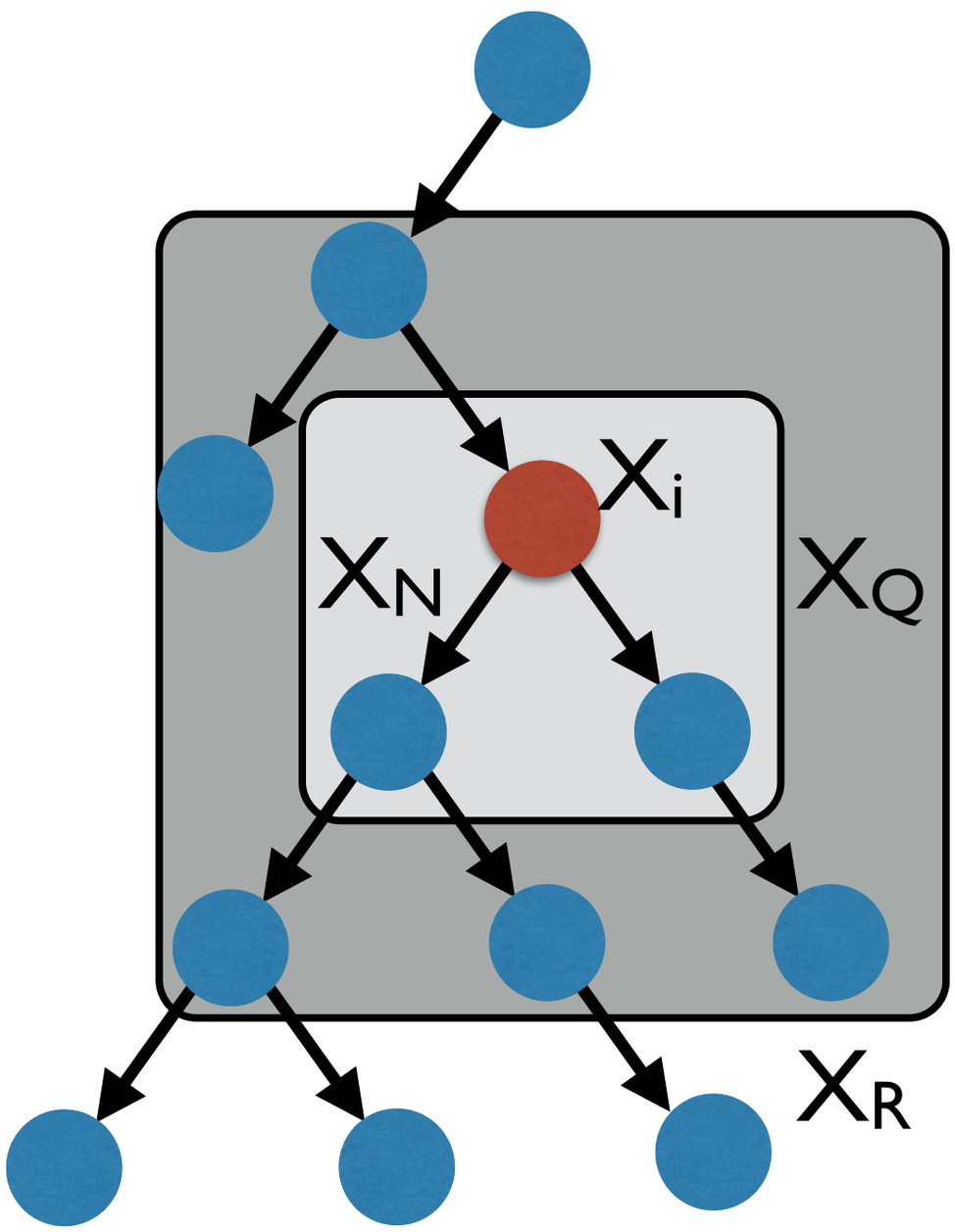
Given a Markov Quilt,

X_N = local nodes for X_i

$X_Q \cup X_R$ = rest

Need to search over Markov Quilts X_Q to find the one which needs optimal amount of noise

From Markov Quilts to Amount of Noise



Let X_Q = Markov Quilt for X_i

Stdev of noise to protect X_i:

Noise due to X_N

$$\text{Score}(X_Q) = \frac{card(X_N)}{\epsilon - e(X_Q|X_i)}$$

Correction for $X_Q \cup X_R$

The Markov Quilt Mechanism

For each X_i

Find the Markov Quilt X_Q for X_i with minimum score s_i

Output F(D) + ($\max_i s_i$) Z where $Z \sim Lap(1)$

The Markov Quilt Mechanism

For each X_i

Find the Markov Quilt X_Q for X_i with minimum score s_i

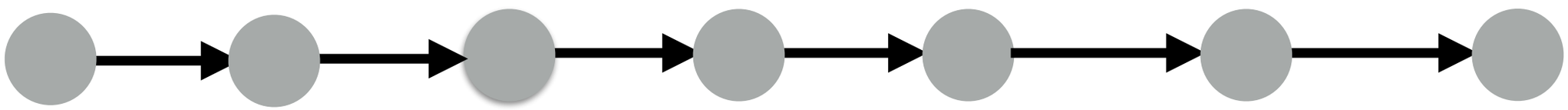
Output F(D) + ($\max_i s_i$) Z where $Z \sim Lap(1)$

Theorem: This preserves ϵ-Pufferfish privacy

Advantage: Poly-time in special cases.

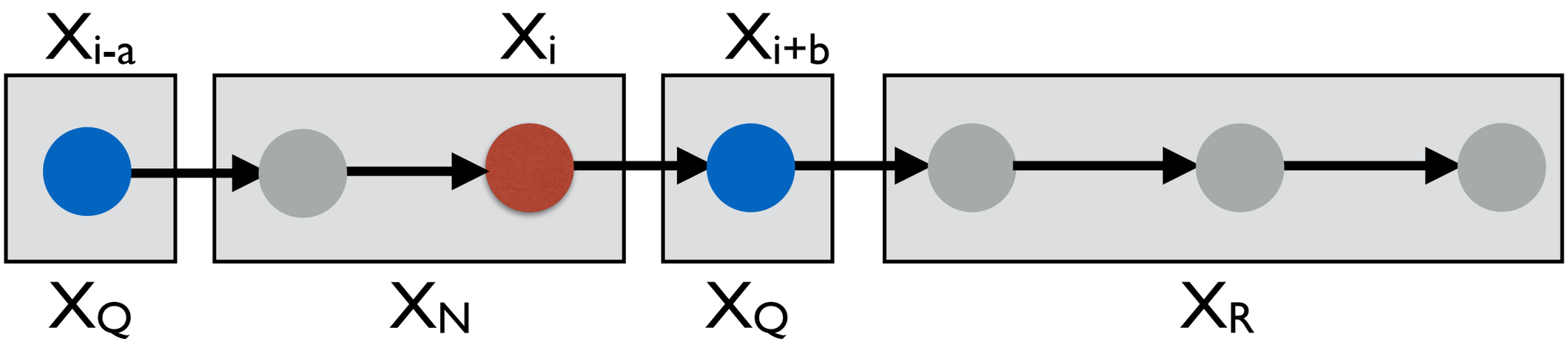
Example: Activity Monitoring

$D = (x_1, .., x_T)$, x_t = activity at time t



Example: Activity Monitoring

$D = (x_1, .., x_T)$, x_t = activity at time t



(Minimal) Markov Quilts for X_i have form $\{X_{i-a}, X_{i+b}\}$

Efficiently searchable

Example: Activity Monitoring

$\mathcal{X}$: set of states

P_θ : transition matrix describing each $\theta \in \Theta$

Example: Activity Monitoring

$\mathcal{X}$: set of states

P_θ : transition matrix describing each $\theta \in \Theta$

Under some assumptions, relevant parameters are:

$\pi_\Theta = \min_{x \in \mathcal{X}, \theta \in \Theta} \pi_\theta(x)$ (min prob of x under stationary distr.)

$g_\Theta = \min_{\theta \in \Theta} \min\{1 - |\lambda| : P_\theta x = \lambda x, \lambda < 1\}$ (min eigengap of any P_θ)

Example: Activity Monitoring

$\mathcal{X}$: set of states

P_θ : transition matrix describing each $\theta \in \Theta$

Under some assumptions, relevant parameters are:

$\pi_\Theta = \min_{x \in \mathcal{X}, \theta \in \Theta} \pi_\theta(x)$ (min prob of x under stationary distr.)

$g_\Theta = \min_{\theta \in \Theta} \min\{1 - |\lambda| : P_\theta x = \lambda x, \lambda < 1\}$ (min eigengap of any P_θ)

Max-influence of $X_Q = \{X_{i-a}, X_{i+b}\}$ for X_i

$$e(X_Q|X_i) \leq \log\left(\frac{\pi_\Theta + \exp(-g_\Theta b)}{\pi_\Theta - \exp(-g_\Theta b)}\right) + 2\log\left(\frac{\pi_\Theta + \exp(-g_\Theta a)}{\pi_\Theta - \exp(-g_\Theta a)}\right)$$

Score(X_Q) = $\dfrac{a + b - 1}{\epsilon - e(X_Q|X_i)}$

Markov Quilt Mechanism for Activity Monitoring

For each X_i

Find Markov Quilt $X_Q = \{X_{i-a}, X_{i+b}\}$ with minimum score s_i

Output $F(D) + (\max_i s_i)\, Z$ where $Z \sim Lap(1)$

Running Time: $O(T^3)$ (can be made $O(T^2)$)

Advantage: Consistency

Conclusion

New mechanism for computing statistics on correlated data

Open Problems:

1. Composing multiple releases on correlated data
2. Other correlation models (beyond Bayesian nets)
3. More mechanisms (for optimization)
4. Applications - activity recognition, location privacy

Conclusion

Learning with Privacy:

Learning from iid data based on convex opt

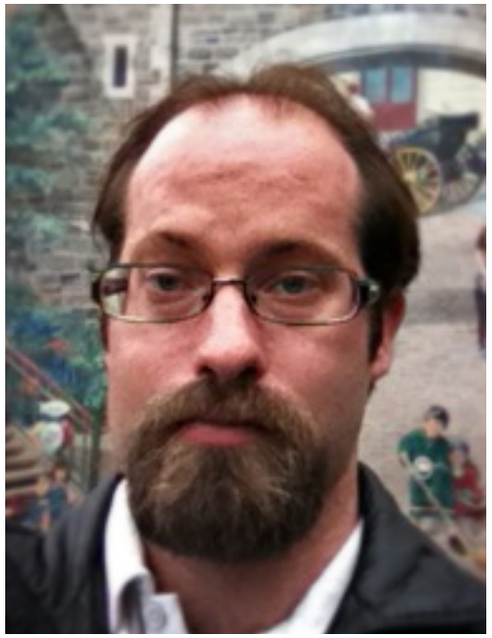
Relatively well-understood

New Directions:

Bayesian Inference

Learning from Correlated Data

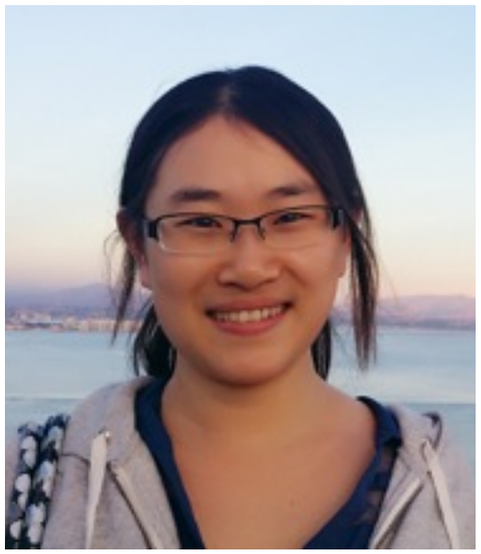
Acknowledgements



James Foulds

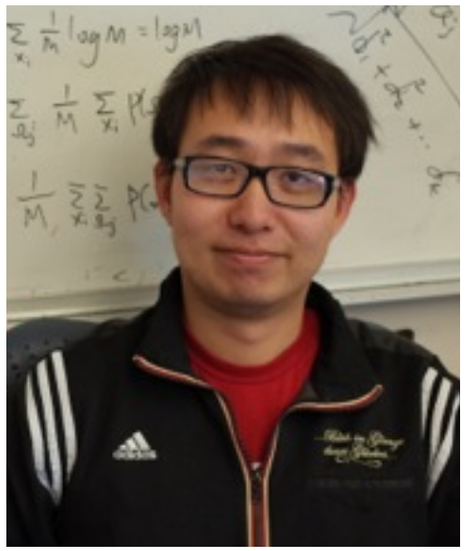
Joseph Geumlek

Max Welling



Shuang Song

Mani Srivastava



Yizhen Wang

Questions?