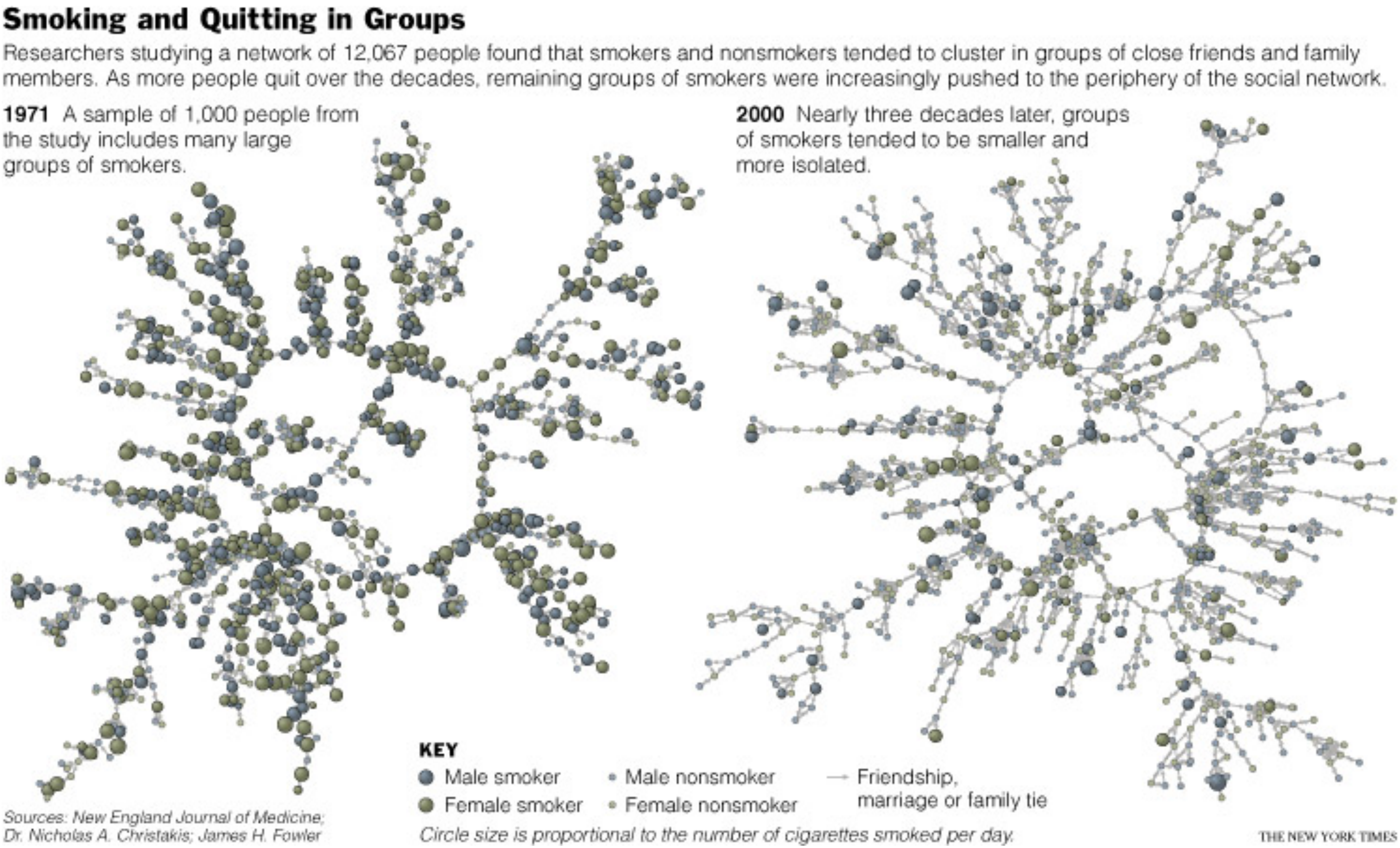
Fully Parallel Inference in Markov Logic Networks

Kaustubh Beedkar, Luciano Del Corro, Rainer Gemulla

Max-Planck-Institut für Informatik
Saarbrücken


Smoking and Quitting in Groups
Researchers studying a network of 12,067 people found that smokers and nonsmokers tended to cluster in groups of close friends and family members. As more people quit over the decades, remaining groups of smokers were increasingly pushed to the periphery of the social network.
1971 A sample of 1,000 people from the study includes many large groups of smokers.
2000 Nearly three decades later, groups of smokers tended to be smaller and more isolated.
KEY
Male smoker
Female smoker
Male nonsmoker
Female nonsmoker
Friendship, marriage or family tie
Circle size is proportional to the number of cigarettes smoked per day.
Sources: New England Journal of Medicine; Dr. Nicholas A. Christakis; James H. Fowler
THE NEW YORK TIMES

Friends

Name1	Name2	Value
Anna	Bob	yes
Bob	Anna	yes
Anna	Anna	yes
Bob	Bob	yes

Smokes

Name	Value
Anna	yes

Cancer

Name	Value
Anna	no

Friends

Name1	Name2	Value
Anna	Bob	yes
Bob	Anna	yes
Anna	Anna	yes
Bob	Bob	yes

Smokes

Name	Value
Anna	yes

Cancer

Name	Value
Anna	no

F_1: 1.5 $\forall x.\text{Smokes}(x) \Rightarrow \text{Cancer}(x)$

F_2: 1.1 $\forall x.\forall y.\text{Friends}(x, y) \Rightarrow (\text{Smokes}(x) \Leftrightarrow \text{Smokes}(y))$

Friends

Name1	Name2	Value
Anna	Bob	yes
Bob	Anna	yes
Anna	Anna	yes
Bob	Bob	yes

Smokes

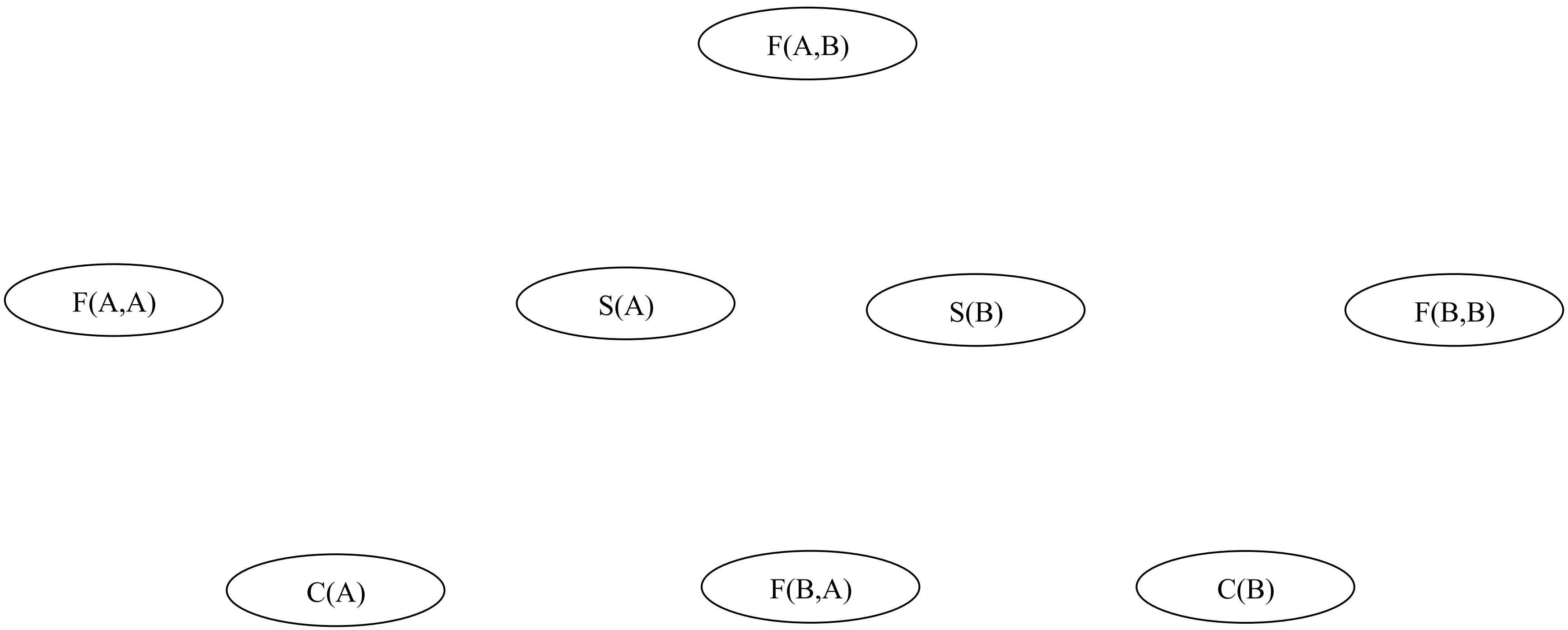
Name	Value
Anna	yes

Cancer

Name	Value
Anna	no

F_1: 1.5 $\forall x.\text{Smokes}(x) \Rightarrow \text{Cancer}(x)$

F_2: 1.1 $\forall x.\forall y.\text{Friends}(x,y) \Rightarrow (\text{Smokes}(x) \Leftrightarrow \text{Smokes}(y))$



Friends

Name1	Name2	Value
Anna	Bob	yes
Bob	Anna	yes
Anna	Anna	yes
Bob	Bob	yes

Smokes

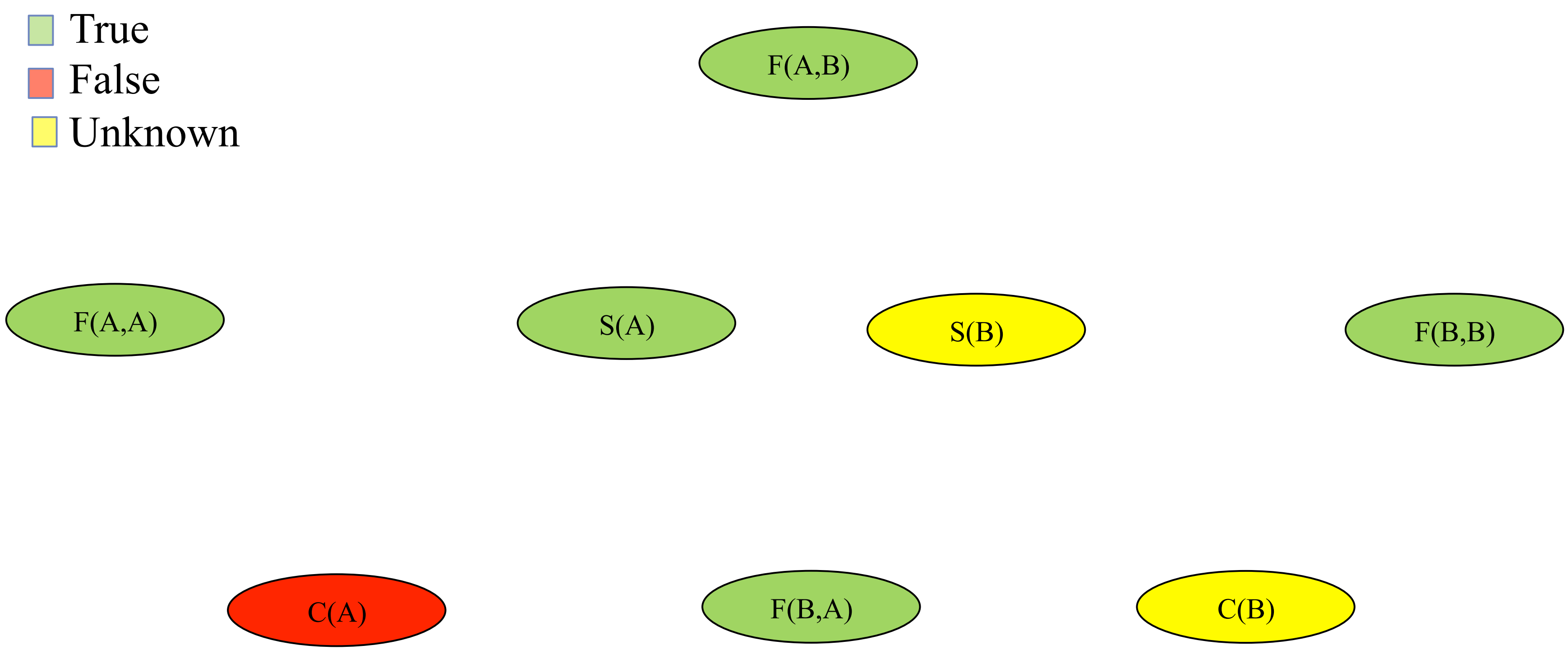
Name	Value
Anna	yes

Cancer

Name	Value
Anna	no

F_1: 1.5 $\forall x.\text{Smokes}(x) \Rightarrow \text{Cancer}(x)$

F_2: 1.1 $\forall x.\forall y.\text{Friends}(x,y) \Rightarrow (\text{Smokes}(x) \Leftrightarrow \text{Smokes}(y))$



Friends

Name1	Name2	Value
Anna	Bob	yes
Bob	Anna	yes
Anna	Anna	yes
Bob	Bob	yes

Smokes

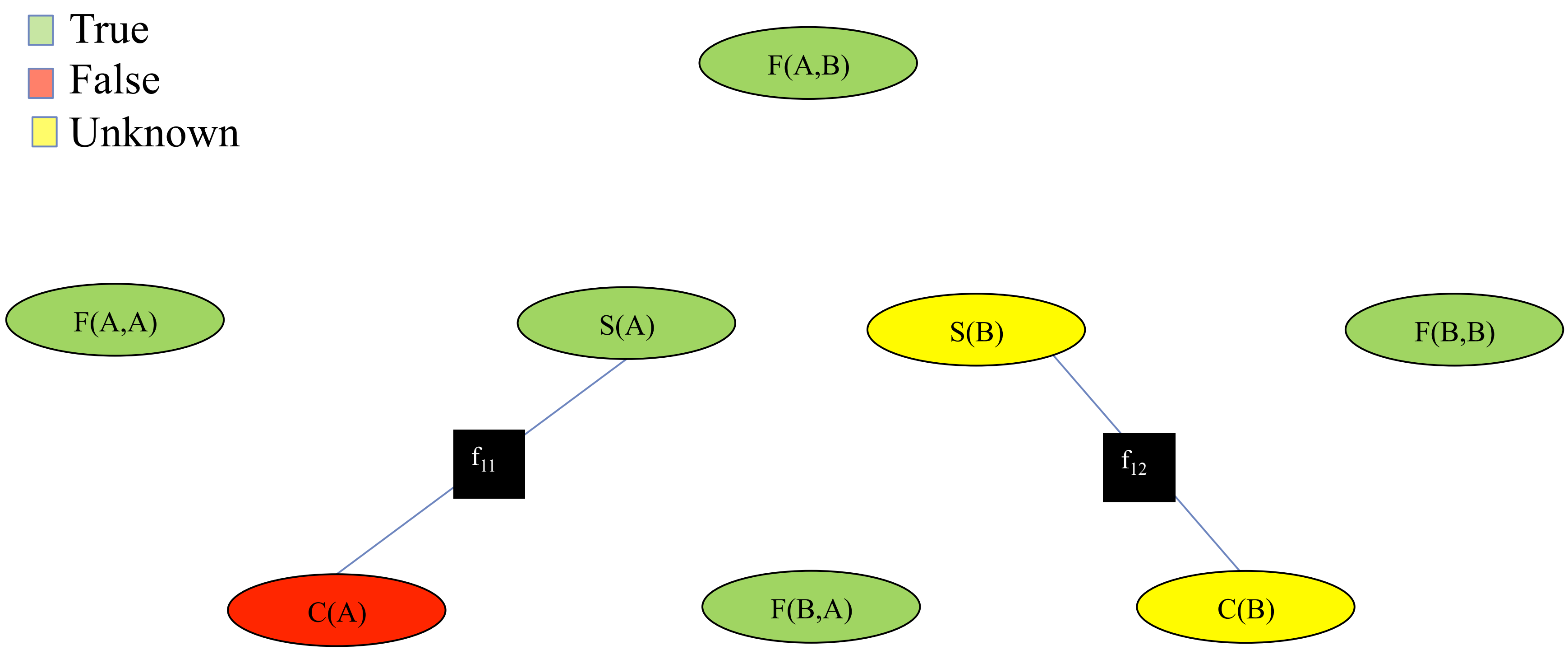
Name	Value
Anna	yes

Cancer

Name	Value
Anna	no

F_1: 1.5 $\forall x.\text{Smokes}(x) \Rightarrow \text{Cancer}(x)$

F_2: 1.1 $\forall x.\forall y.\text{Friends}(x,y) \Rightarrow (\text{Smokes}(x) \Leftrightarrow \text{Smokes}(y))$



Friends

Name1	Name2	Value
Anna	Bob	yes
Bob	Anna	yes
Anna	Anna	yes
Bob	Bob	yes

Smokes

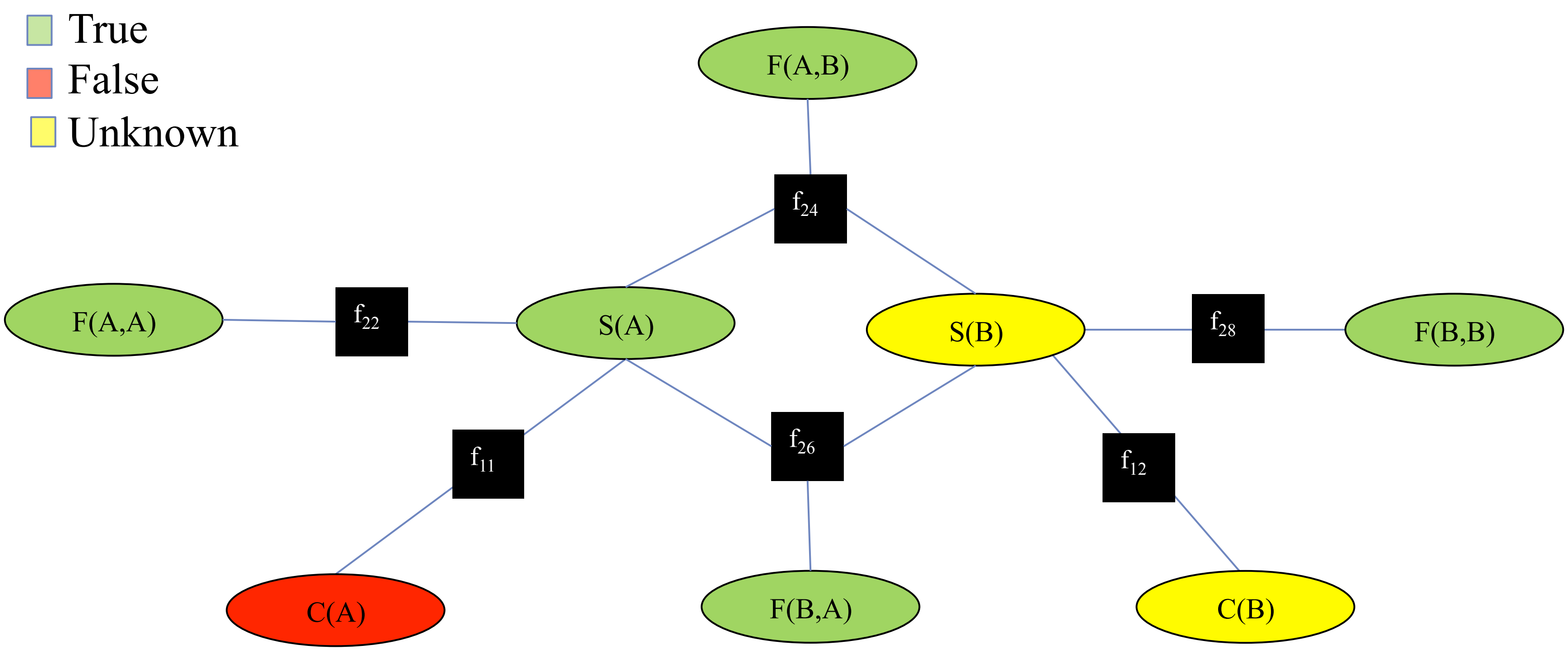
Name	Value
Anna	yes

Cancer

Name	Value
Anna	no

F_1: 1.5 $\forall x.\text{Smokes}(x) \Rightarrow \text{Cancer}(x)$

F_2: 1.1 $\forall x.\forall y.\text{Friends}(x,y) \Rightarrow (\text{Smokes}(x) \Leftrightarrow \text{Smokes}(y))$



Friends

Name1	Name2	Value
Anna	Bob	yes
Bob	Anna	yes
Anna	Anna	yes
Bob	Bob	yes

Smokes

Name	Value
Anna	yes

Cancer

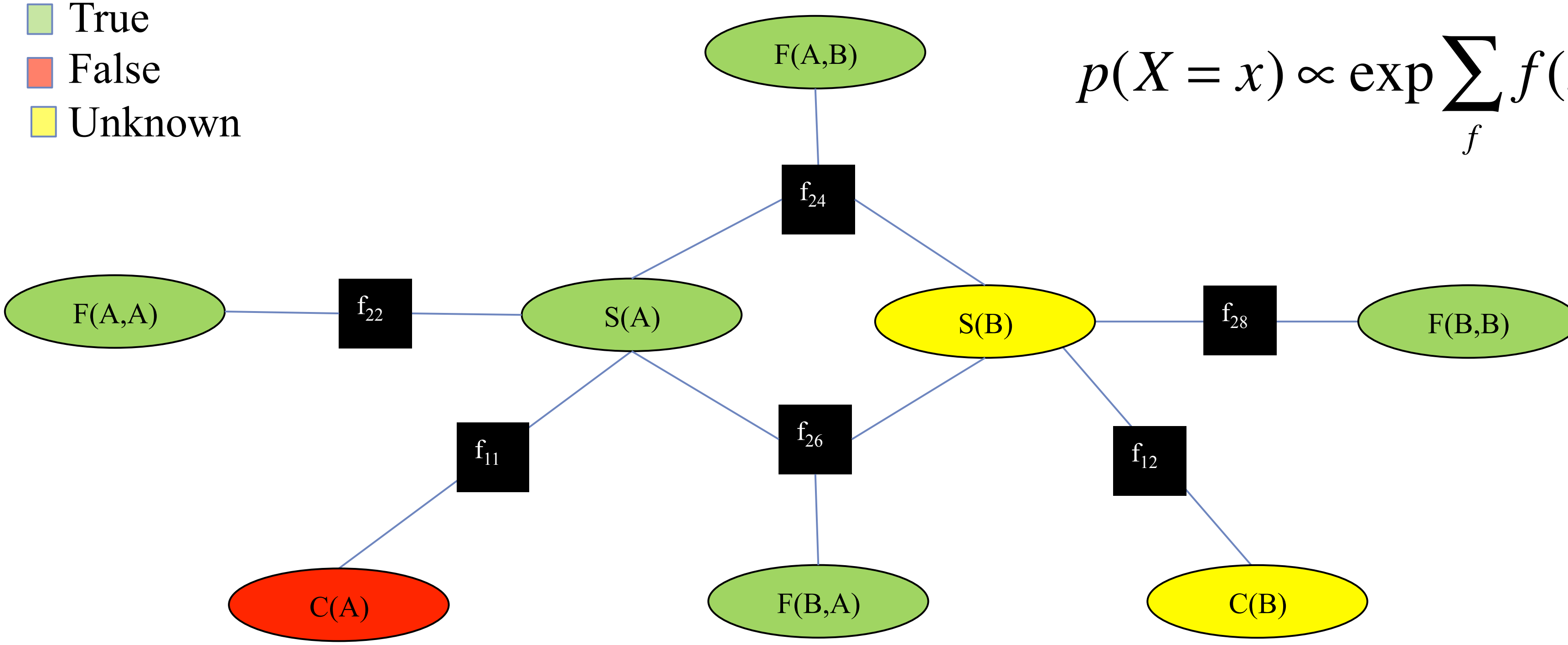
Name	Value
Anna	no

F_1: 1.5 $\forall x.\text{Smokes}(x) \Rightarrow \text{Cancer}(x)$

F_2: 1.1 $\forall x.\forall y.\text{Friends}(x,y) \Rightarrow (\text{Smokes}(x) \Leftrightarrow \text{Smokes}(y))$

- True
- False
- Unknown

$$p(X = x) \propto \exp \sum_f f(x_f)$$



Friends

Name1	Name2	Value
Anna	Bob	yes
Bob	Anna	yes
Anna	Anna	yes
Bob	Bob	yes

Smokes

Name	Value
Anna	yes

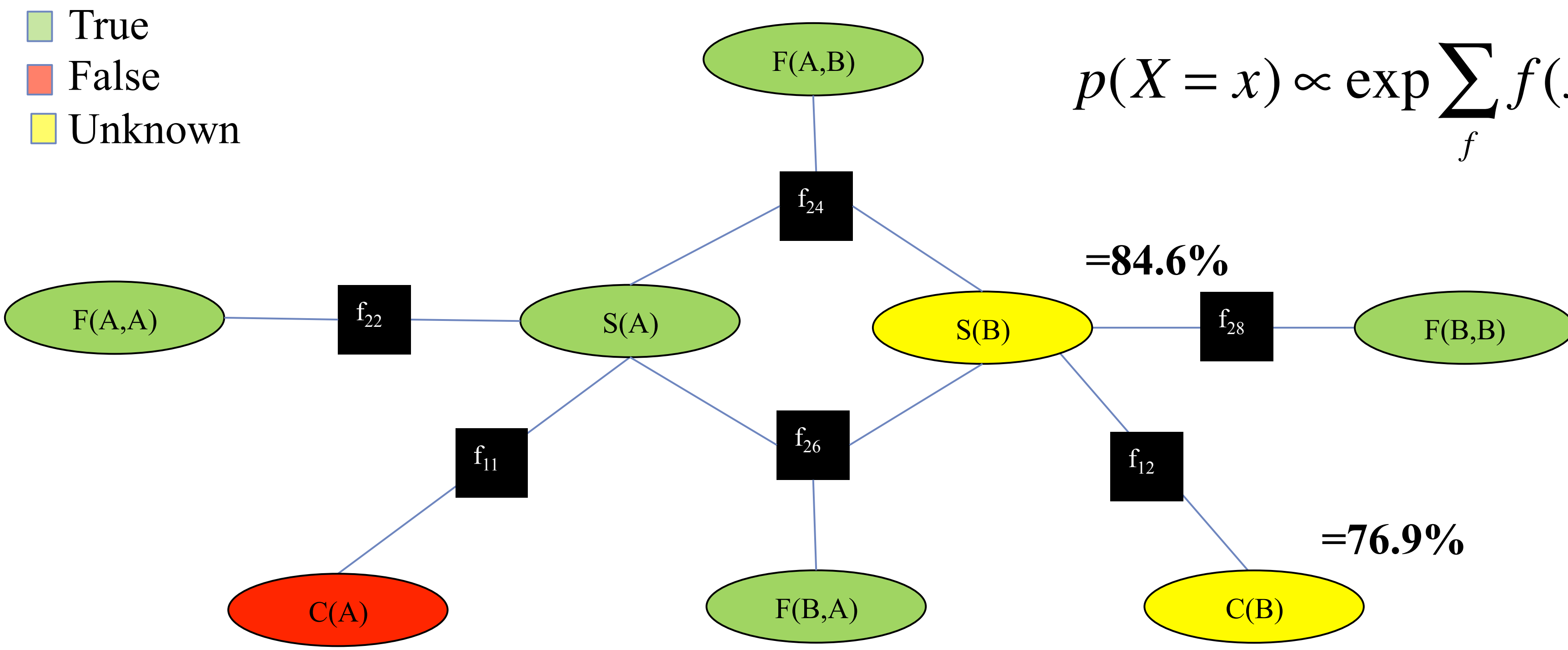
Cancer

Name	Value
Anna	no

F_1: 1.5 $\forall x.\text{Smokes}(x) \Rightarrow \text{Cancer}(x)$

F_2: 1.1 $\forall x.\forall y.\text{Friends}(x,y) \Rightarrow (\text{Smokes}(x) \Leftrightarrow \text{Smokes}(y))$

$$p(X = x) \propto \exp \sum_f f(x_f)$$



Inference in Markov Logic Networks (I)

Sampling in MNL

- Approximation unavoidable
- Generic technique to approximate expectations
- Simple, versatile, well understood

Inference in Markov Logic Networks (I)

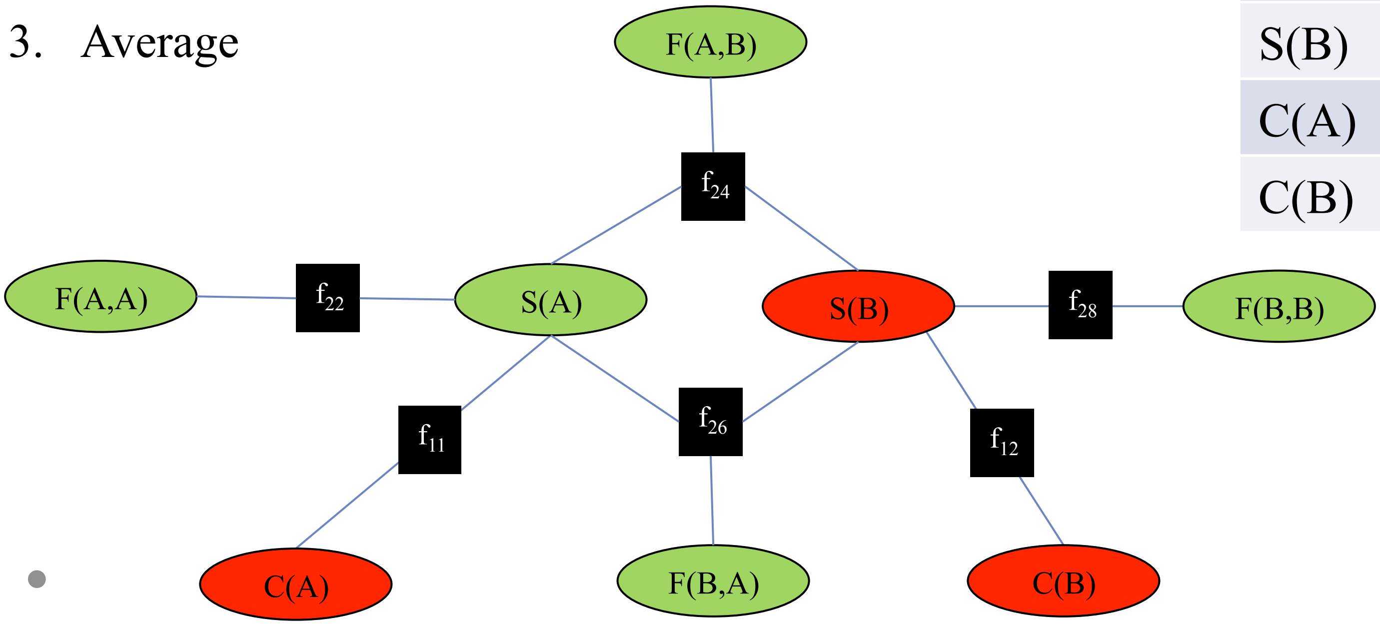
Sampling in MNL

- Approximation unavoidable
- Generic technique to approximate expectations
- Simple, versatile, well understood

Sampling process

1. Assign a value to each variable
2. Count
3. Average

Var	#true	#false
F(A,A)	xx	xx
F(A,B)	xx	xx
F(B,A)	xx	xx
F(B,B)	xx	xx
S(A)	xx	xx
S(B)	xx	xx
C(A)	xx	xx
C(B)	xx	xx



Inference in Markov Logic Networks (I)

Sampling in MNL

- Approximation unavoidable
- Generic technique to approximate expectations
- Simple, versatile, well understood

Sampling process

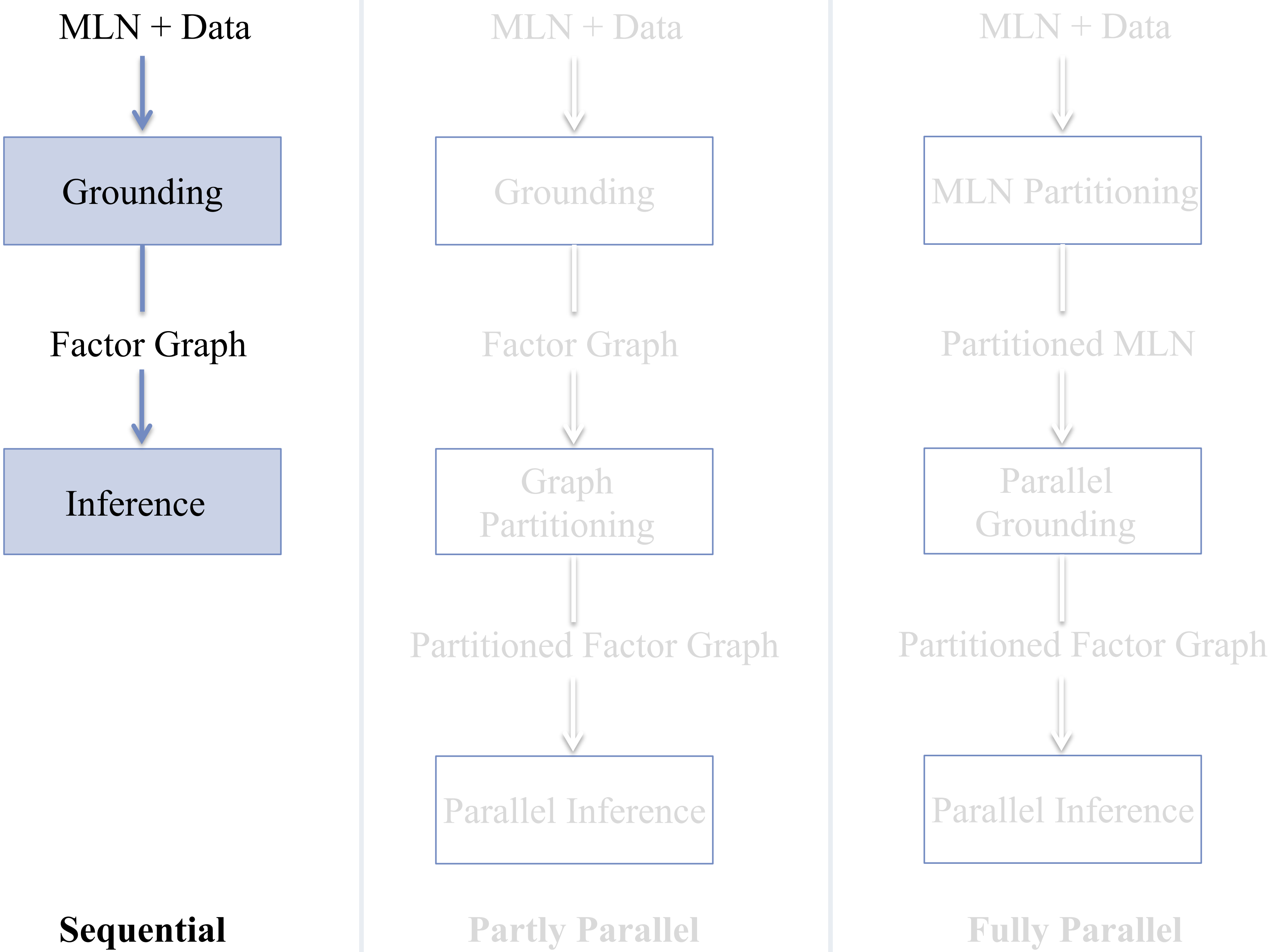
1. Assign a value to each variable
2. Count
3. Average

$$\frac{1}{n}\sum_{i=1}^{n} h(x^{(i)}) = \hat{\mu}$$

$$Var_p[\hat{\mu}] = \frac{Var_p[h(x)]}{n}$$

More samples more efficiency

Sequential approach



Networks can be very large

Lots of applications

- Link prediction
- Information Extraction
- Entity Resolution
- Ontology Learning

How to gain scalability?

- Grounding is expensive
- Inference is expensive

Networks can be very large

Lots of applications

- Link prediction
- Information Extraction
- Entity Resolution
- Ontology Learning

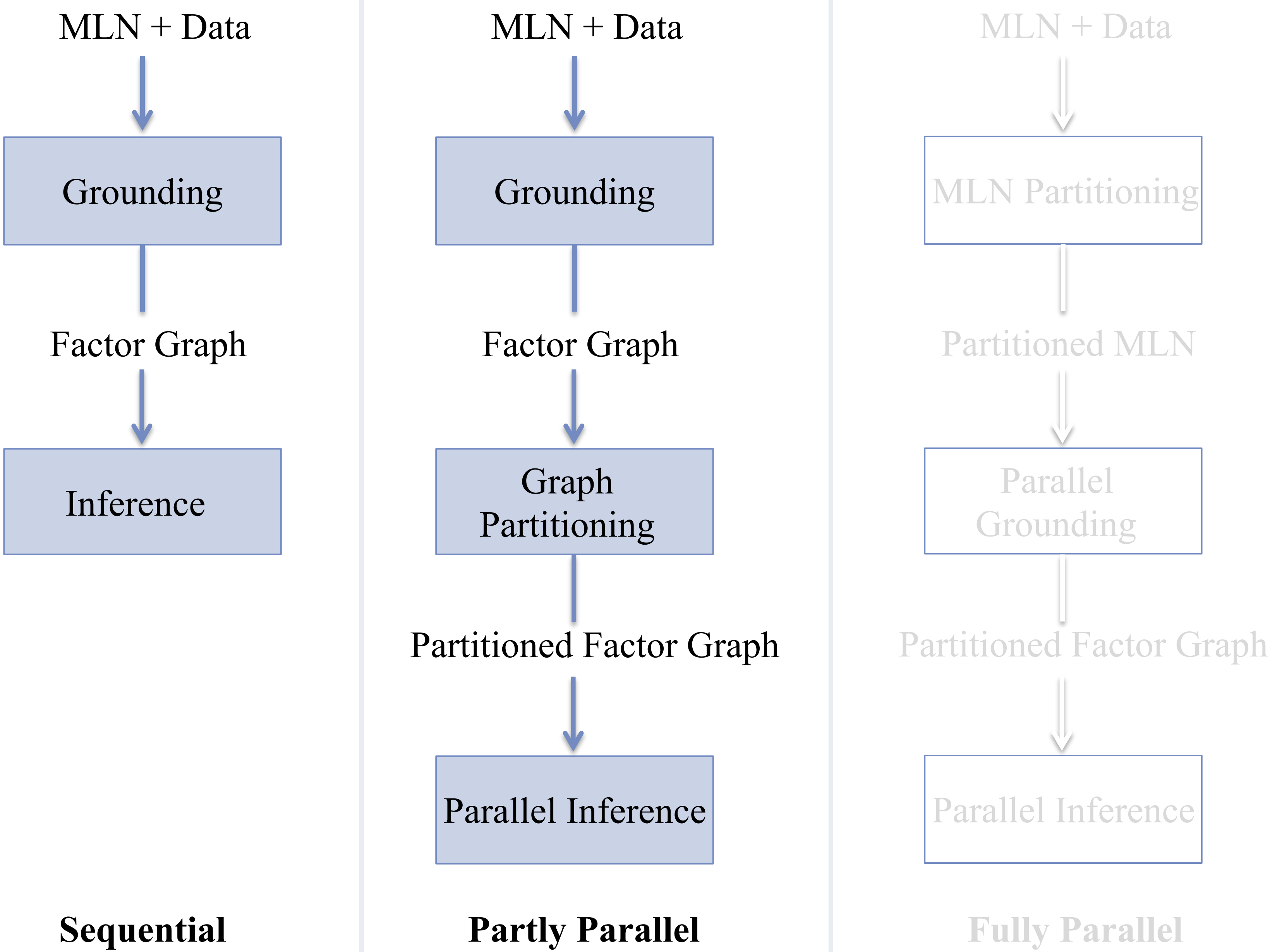
How to gain scalability?

- Grounding is expensive
- Inference is expensive

Why speed up sampling?

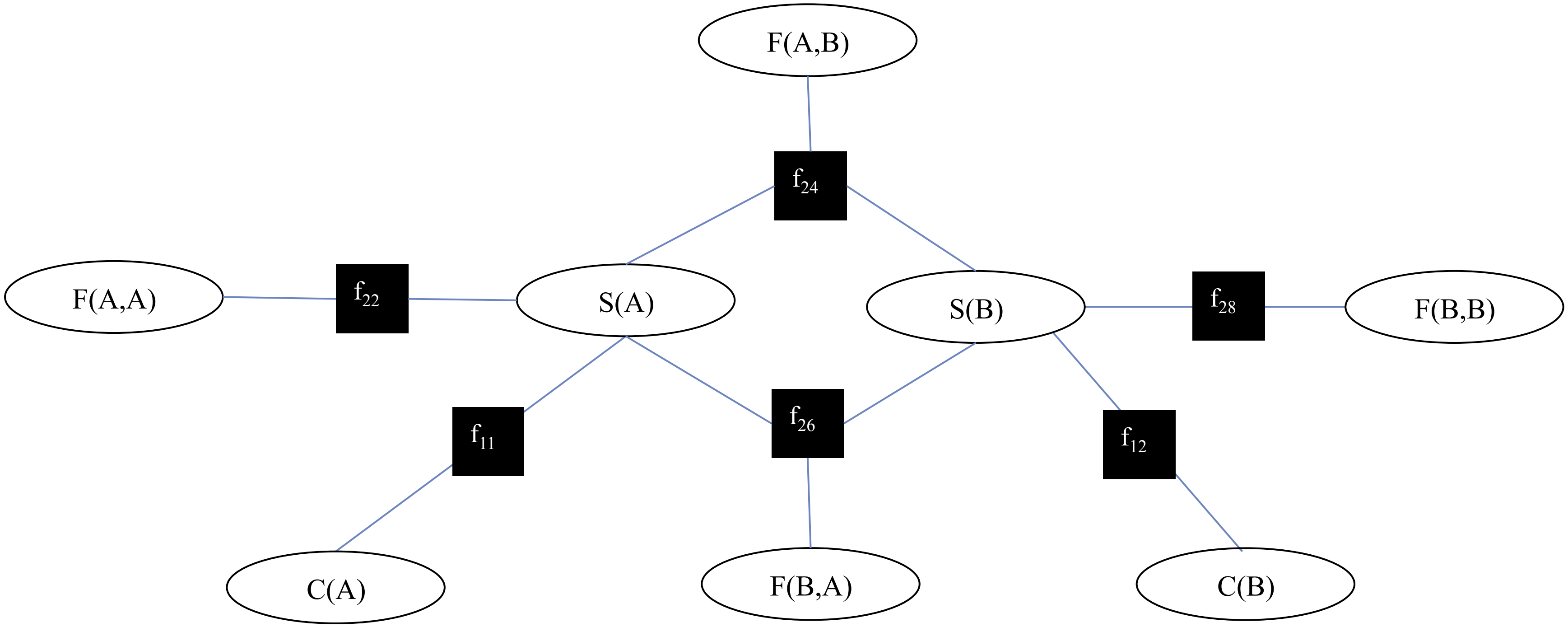
- Expensive
- Datasets can be big
- Dataset 72k variables each sample between 2-5 seconds
- 1 million samples ≈ 50 days

Partly parallel approach



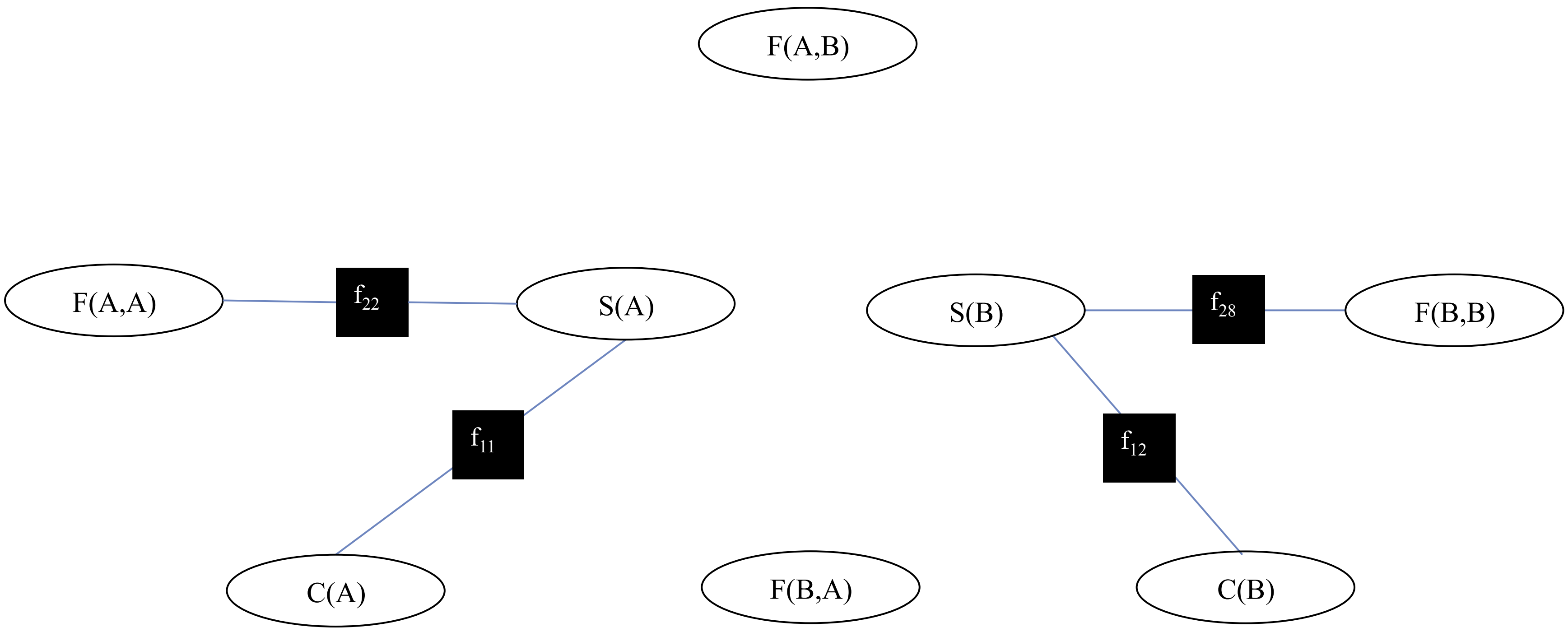
Distributing a network via graph cuts

Cut the network to sample each partition in parallel



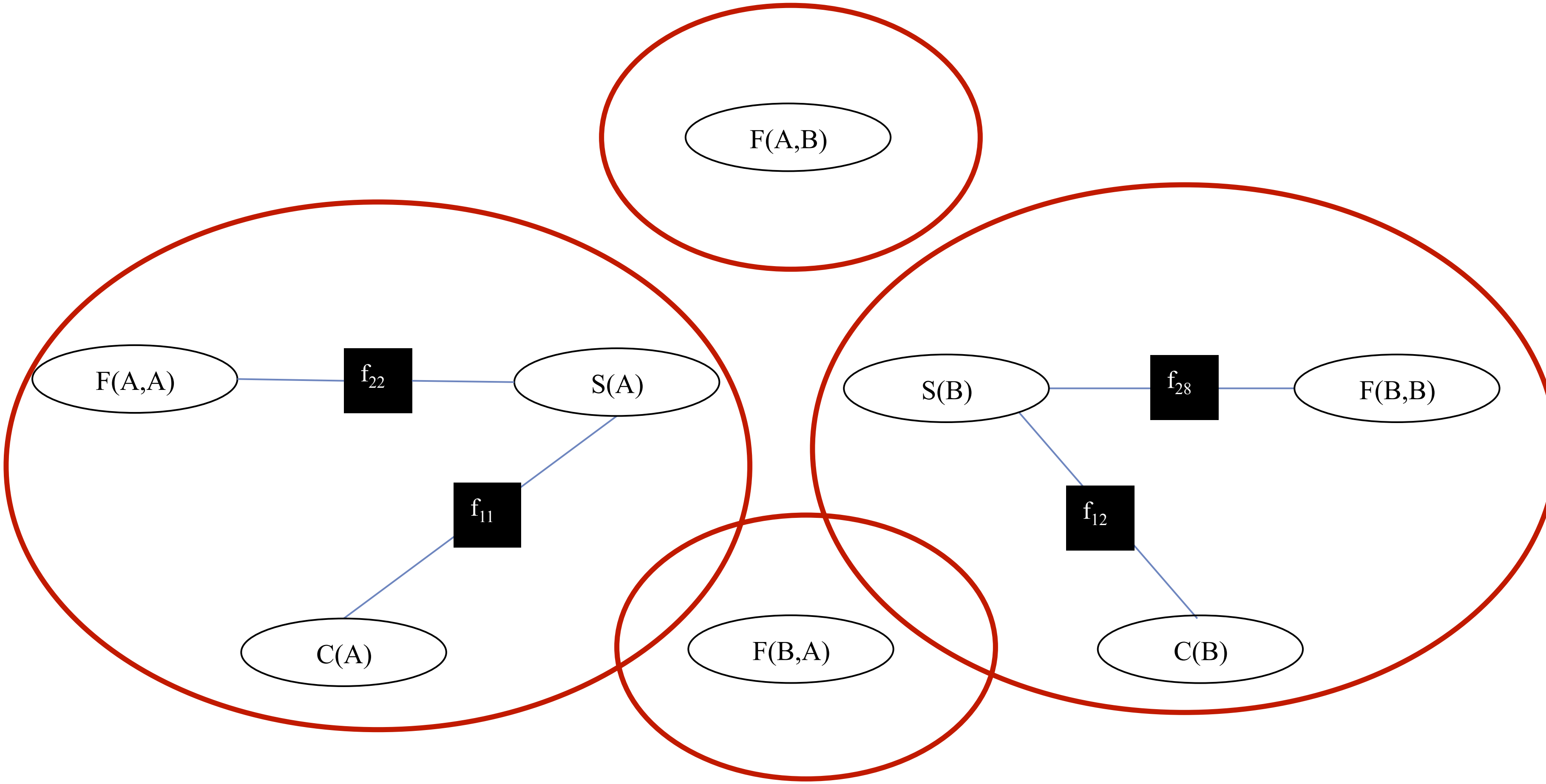
Distributing a network via graph cuts

Cut is performed by removing factors to generate independent components



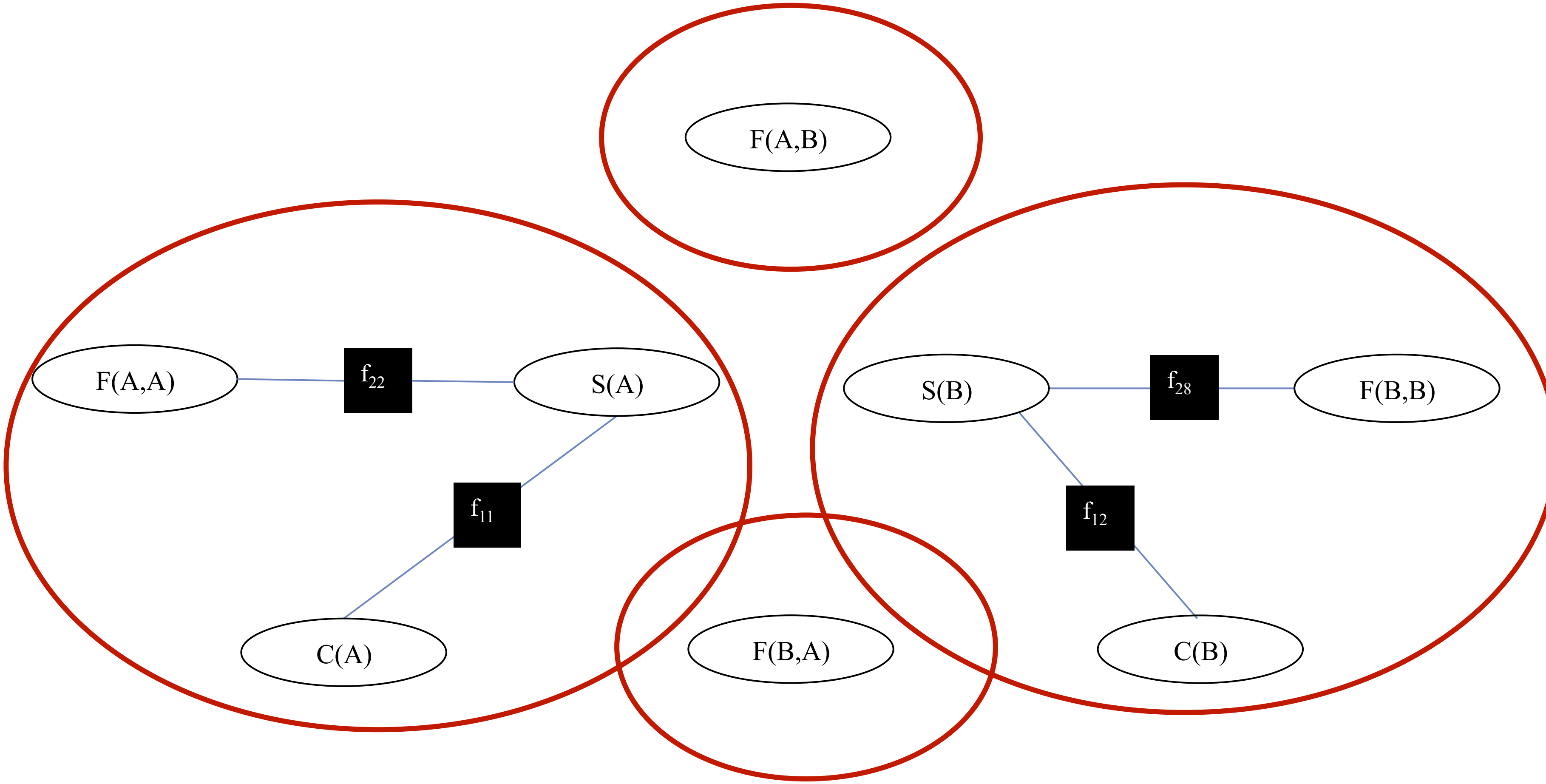
Distributing a network via graph cuts

Each component can be sampled in parallel



Distributing a network via graph cuts

Each component can be sampled in parallel

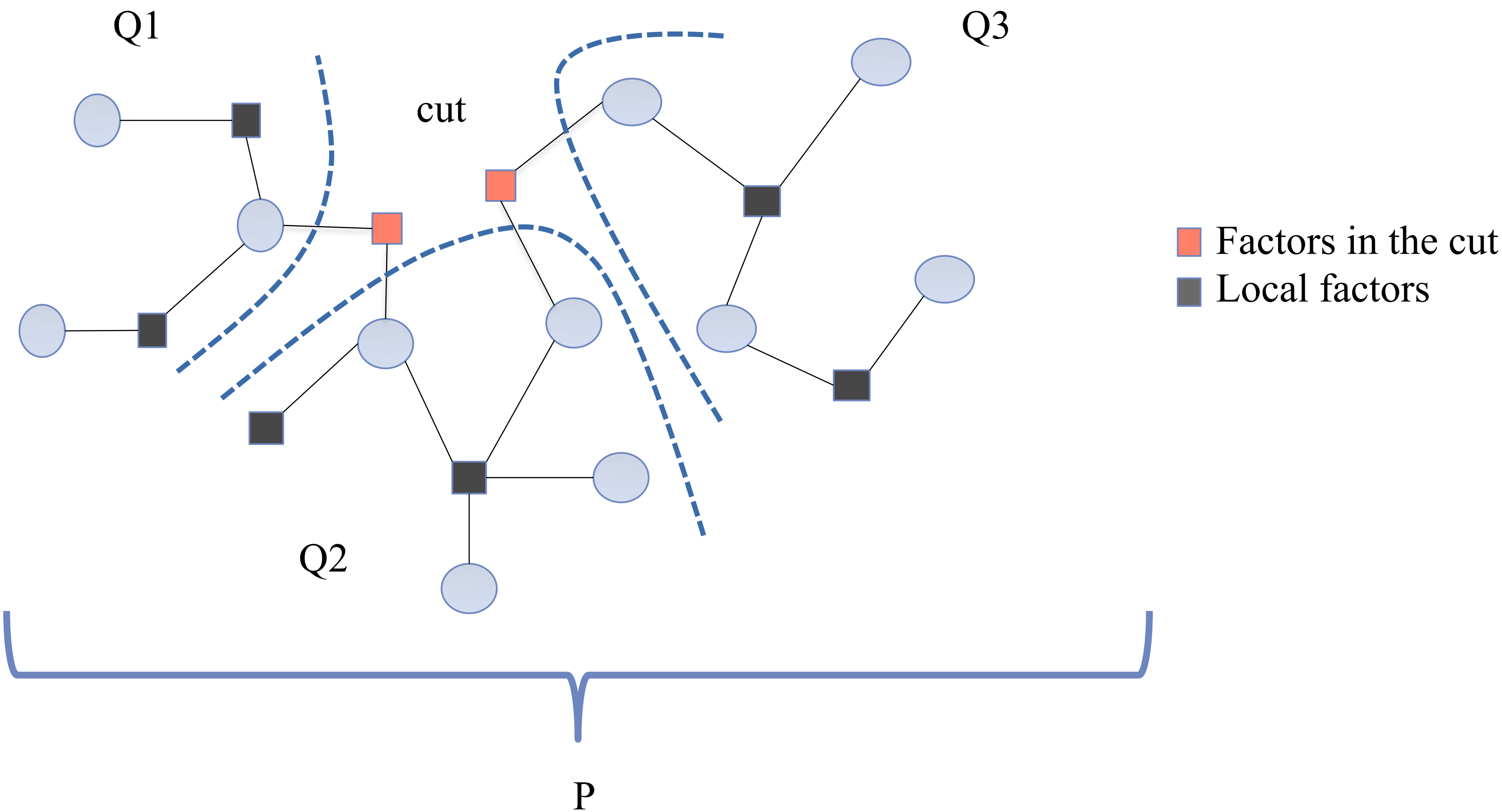


Information loss equivalent to lost connections. How big is the information loss?

Outline

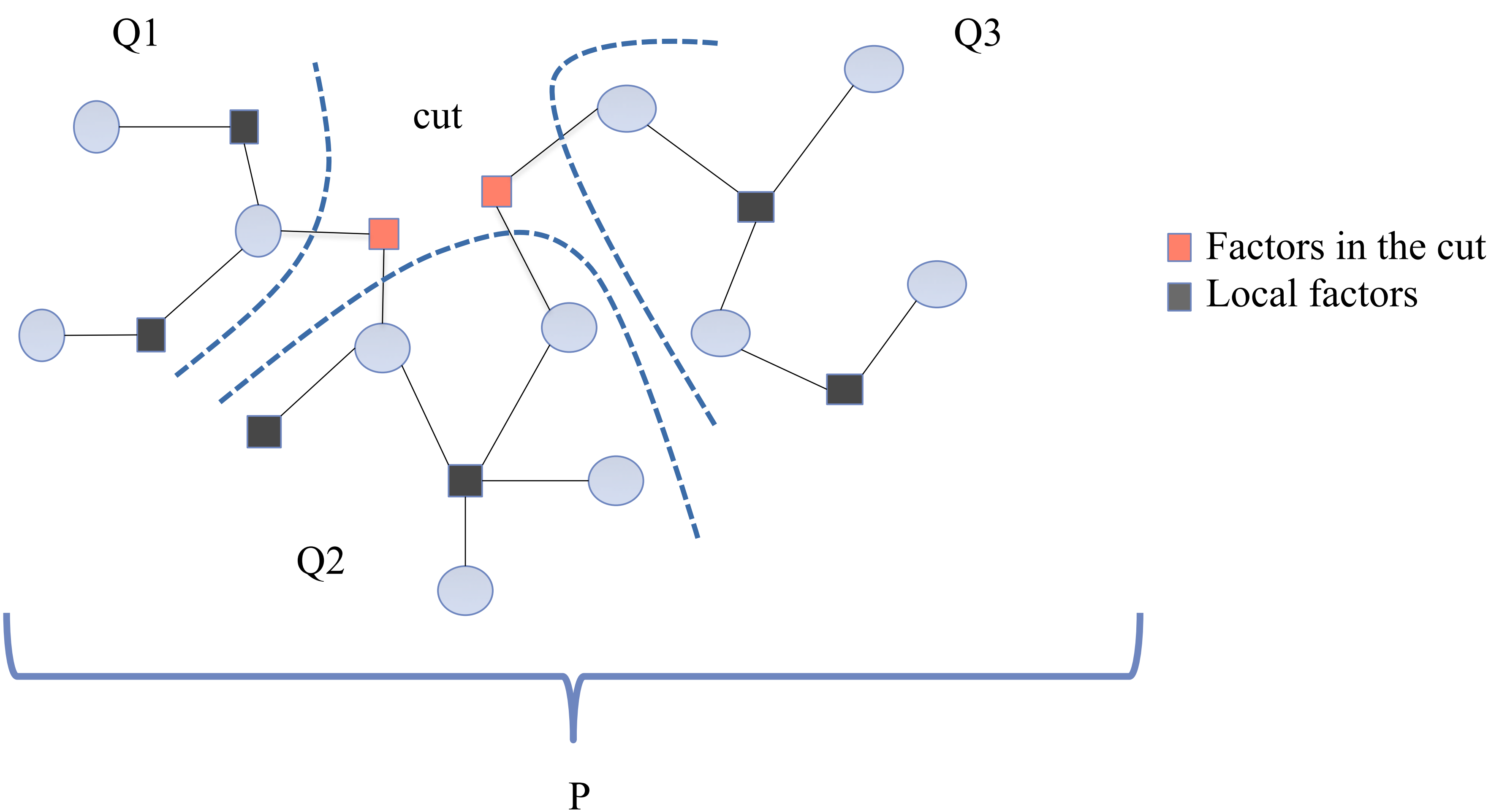
- Background and Motivation
- Parallel Inference
- Parallel Grounding
- Conclusion

What is the best partitioning?



If factors in "cut"are weak ⟶ Q1 * Q2 * Q3 ≈ P

What is the best partitioning?

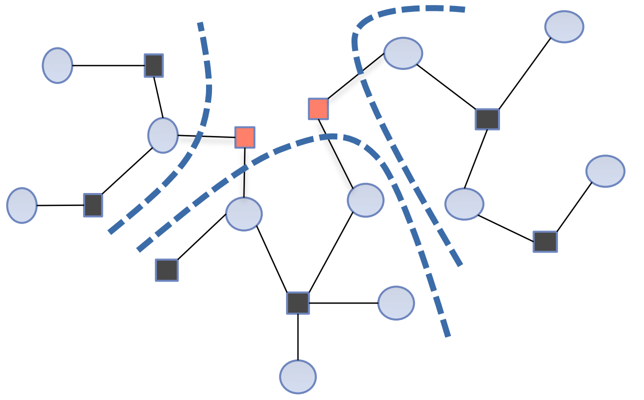


If factors in “cut”are weak → Q1 * Q2 * Q3 ≈ P

How to find a cut with weak factors?

What is the best partitioning?

Importance Sampling



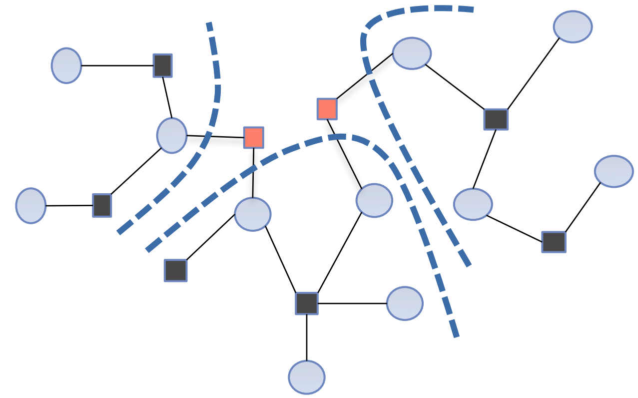
1. Cut the graph to get independent components
2. Get a sample from each component independently
3. Correct the sample to match the original distribution
4. Correction determined by factors in cut

Factors in the cut
Local factors

Efficiency of the estimation depends on the information loss (factors in the cut)

What is the best partitioning?

Importance Sampling



1. Cut the graph to get independent components
2. Get a sample from each component independently
3. Correct the sample to match the original distribution
4. Correction determined by factors in cut

Factors in the cut
Local factors

Efficiency of the estimation depends on the information loss (factors in the cut)

Standard Monte-Carlo

$$Var_p[\hat{\mu}] = \frac{Var_p[h(x)]}{n}$$

Importance Sampling

$$Var_q(\hat{\mu}_{is}) \approx \frac{(1 + Var_q[w(x)])Var_p[h(x)]}{n}$$

w(x): sum of the instantiated factors in the cut for a sample

Quality of the cut: a bound

- Calculating the dispersion of the weights in the cut is intractable
- What is the worst possible quality for each cut?
- What is the best of the worst?

Quality of the cut: a bound

- Calculating the dispersion of the weights in the cut is intractable
- What is the worst possible quality for each cut?

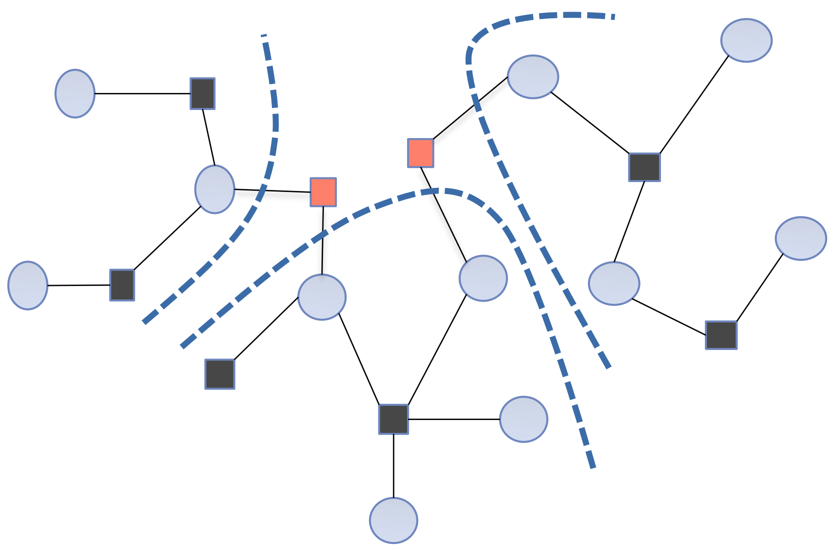
What is the best worst cut?

↓

Minimize the sum of the factors in the cut

↓

Can be easily casted into a standard min-cut algorithm



■ Factors in the cut
■ Local factors

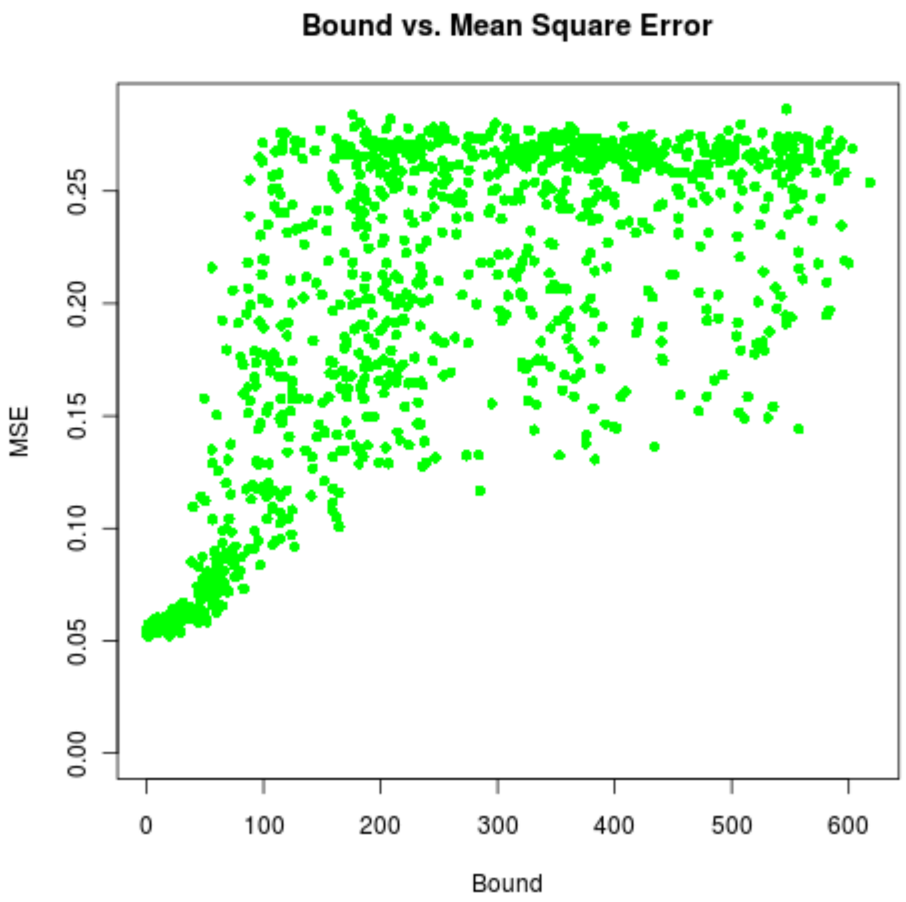
Quality of the cut: a bound

- Calculating the dispersion of the weights in the cut is intractable
- What is the worst possible quality for each cut?

What is the best worst cut?

Minimize the sum of the factors in the cut

Can be easily casted into a standard min-cut algorithm



Results Parallel Inference with Importance Sampling

Dataset

- UW-CSE (22 predicates, 94 clauses)
  - Link prediction
- ~9K variables and ~1M factors (after grounding)

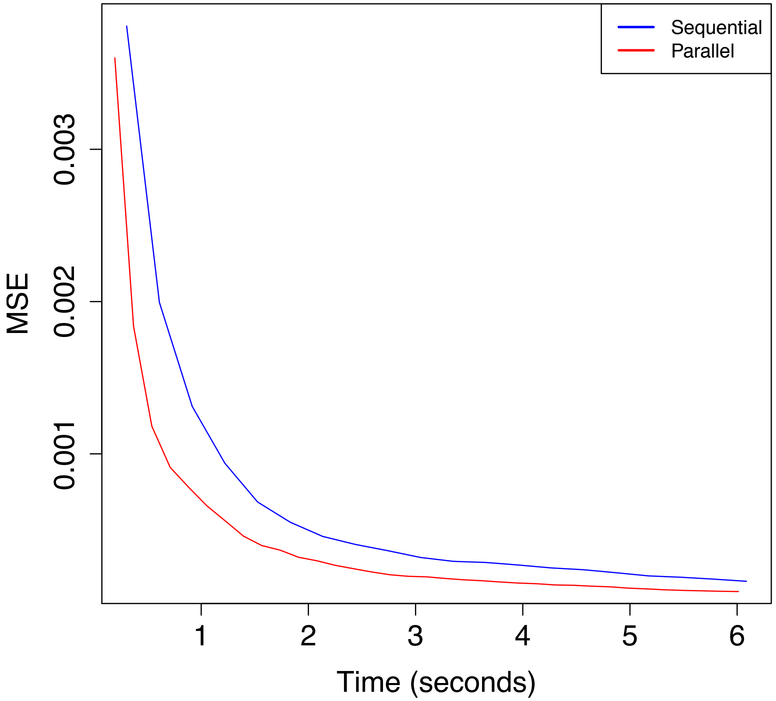
Maximum SE

Results Parallel Inference with Importance Sampling

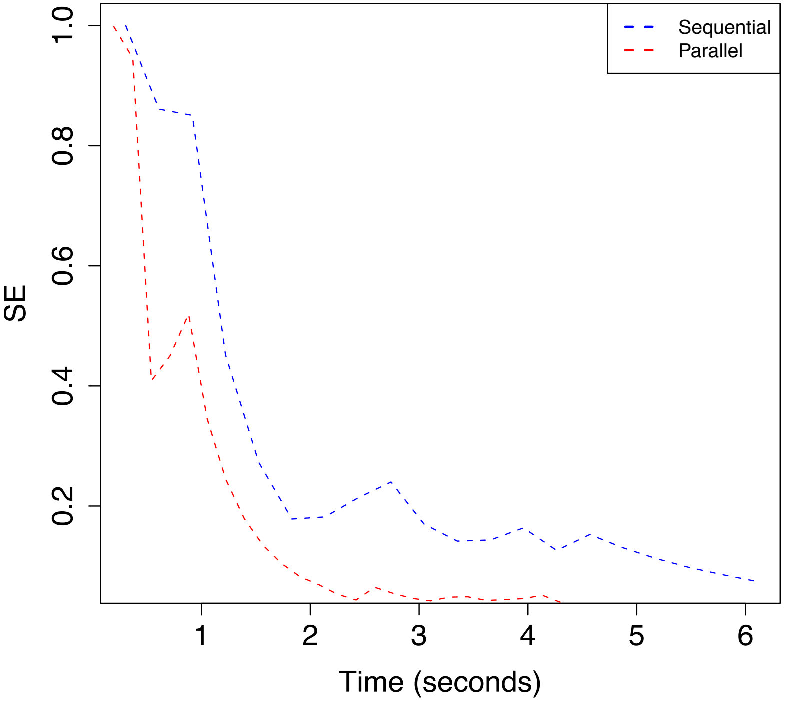
Dataset

- UW-CSE (22 predicates, 94 clauses)
  - Link prediction
- ~9K variables and ~1M factors (after grounding)

Sequential and parallel probabilistic inference (4 partitions)

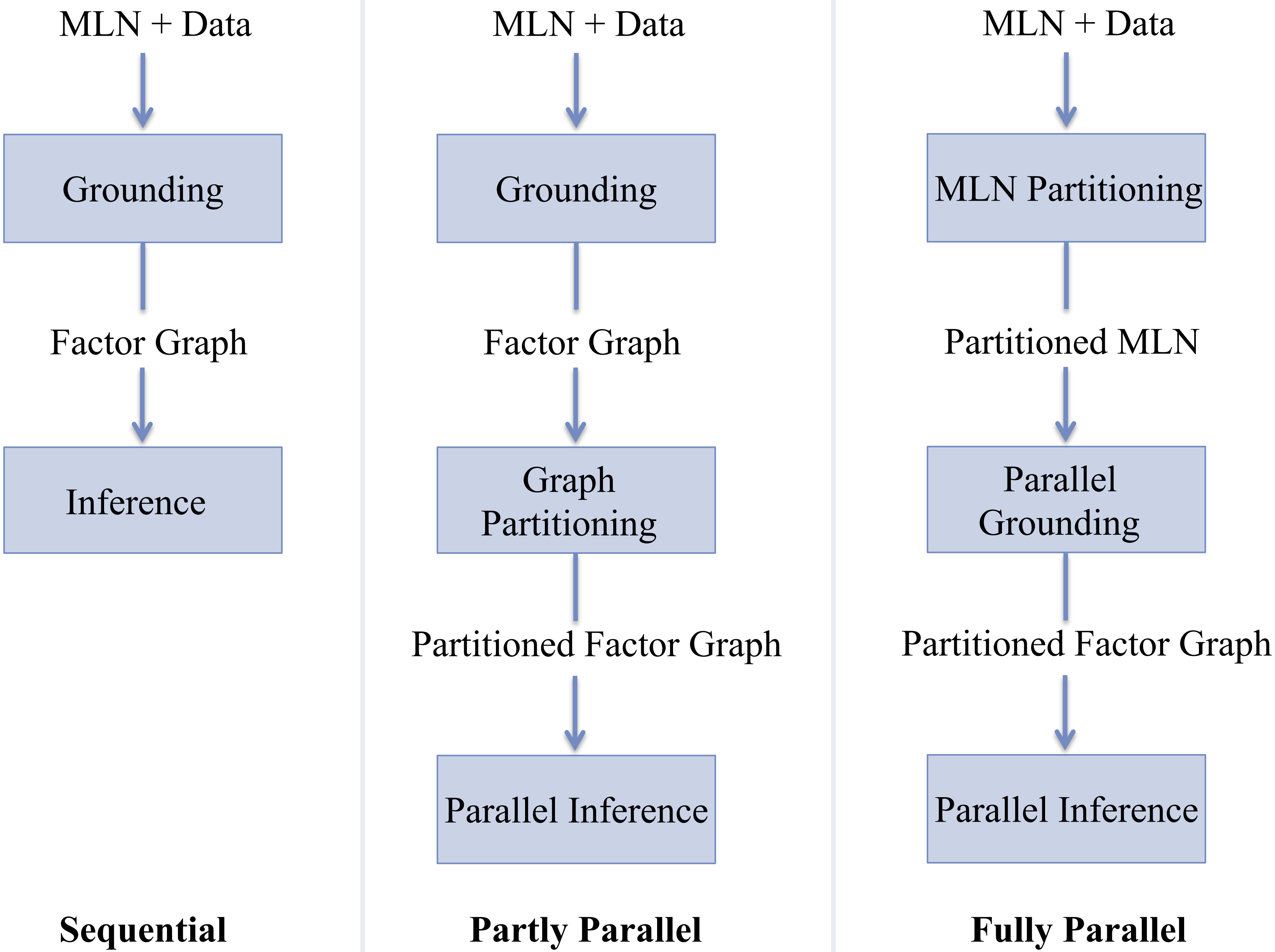


Average MSE



Maximum SE

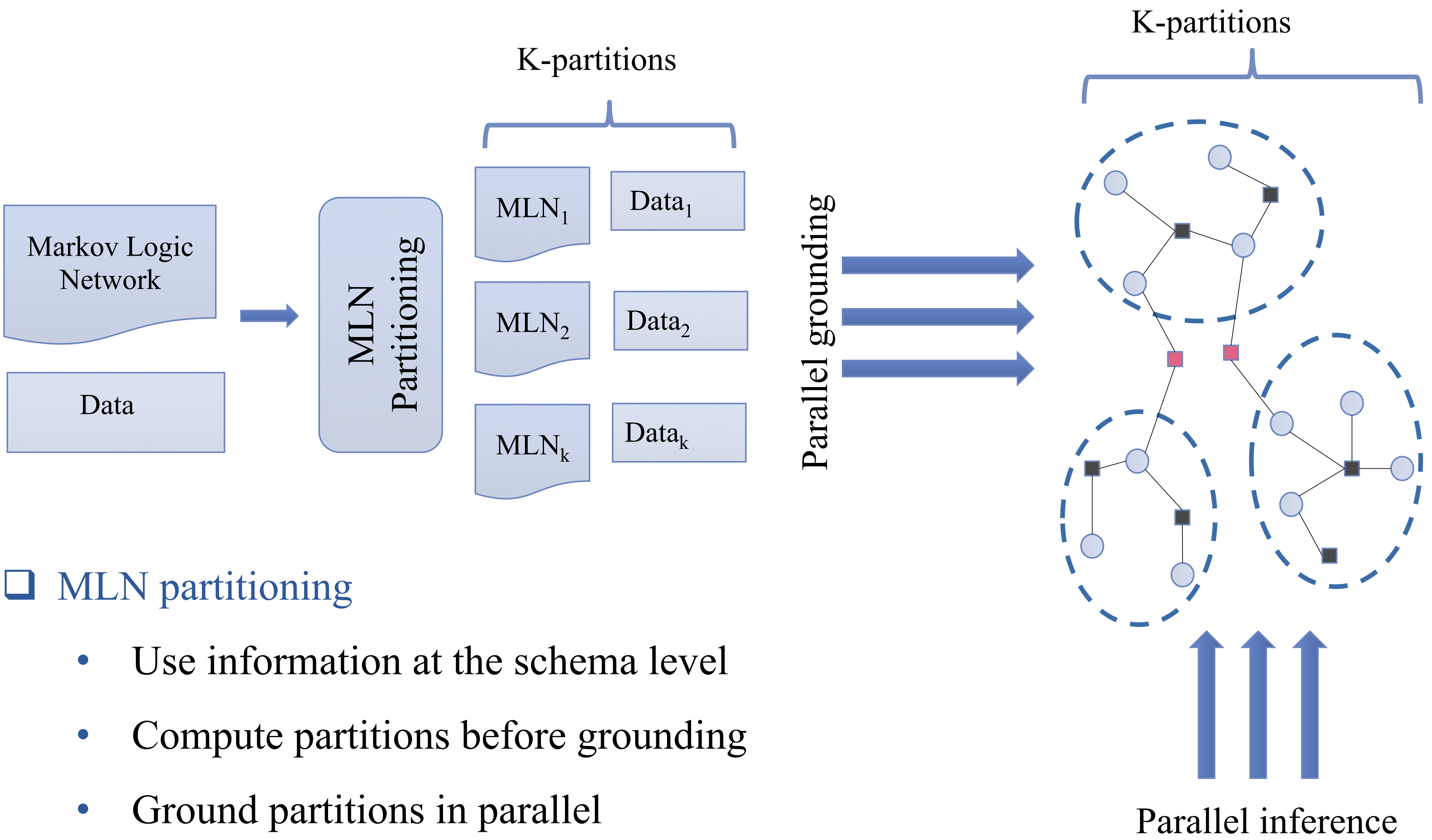
Fully Parallel Approach



Outline

- Background and Motivation
- Parallel Inference
- Parallel Grounding
- Conclusion

Parallel Grounding



- MLN partitioning
  - Use information at the schema level
  - Compute partitions before grounding
  - Ground partitions in parallel
  - Avoids expensive graph cuts

Grounding $\equiv$ Database joins

Formula $\text{Smokes}(x) \Rightarrow \text{Cancer}(x)$	CNF $\neg\text{Smokes}(x) \vee \text{Cancer}(x)$
Predicates and domain $\text{Smokes}(person)$ $\text{Cancer}(person)$ $person = \{Anna, Bob\}$	Ground Clauses $\neg\text{Smokes}(Anna) \vee \text{Cancer}(Anna)$ $\neg\text{Smokes}(Bob) \vee \text{Cancer}(Bob)$

Grounding ≡ Database joins

Formula	$\text{Smokes}(x) \Rightarrow \text{Cancer}(x)$	CNF	$\neg\text{Smokes}(x) \vee \text{Cancer}(x)$

Predicates and domain

$\text{Smokes}(person)$

$\text{Cancer}(person)$

$person = \{Anna, Bob\}$

Ground Clauses

$\neg\text{Smokes}(Anna) \vee \text{Cancer}(Anna)$

$\neg\text{Smokes}(Bob) \vee \text{Cancer}(Bob)$

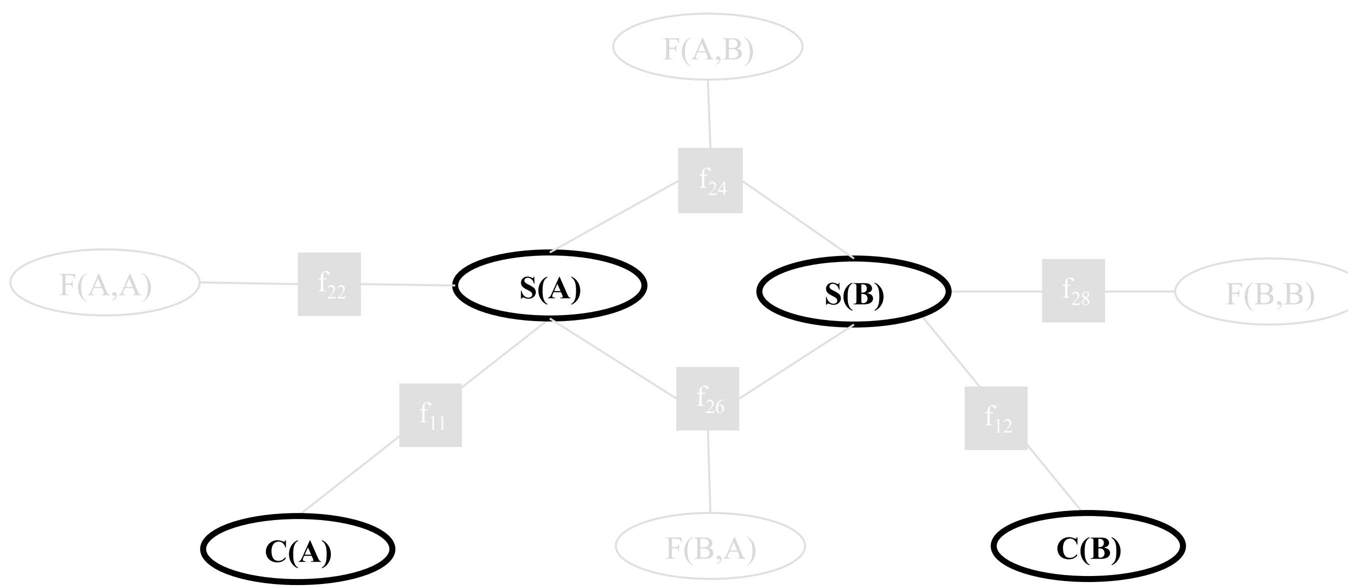
- Ground variables corresponds to Relations

R_1: Smokes

Attr	Person
	Anna
	Bob

R_2: Cancer

Attr	Person
	Anna
	Bob

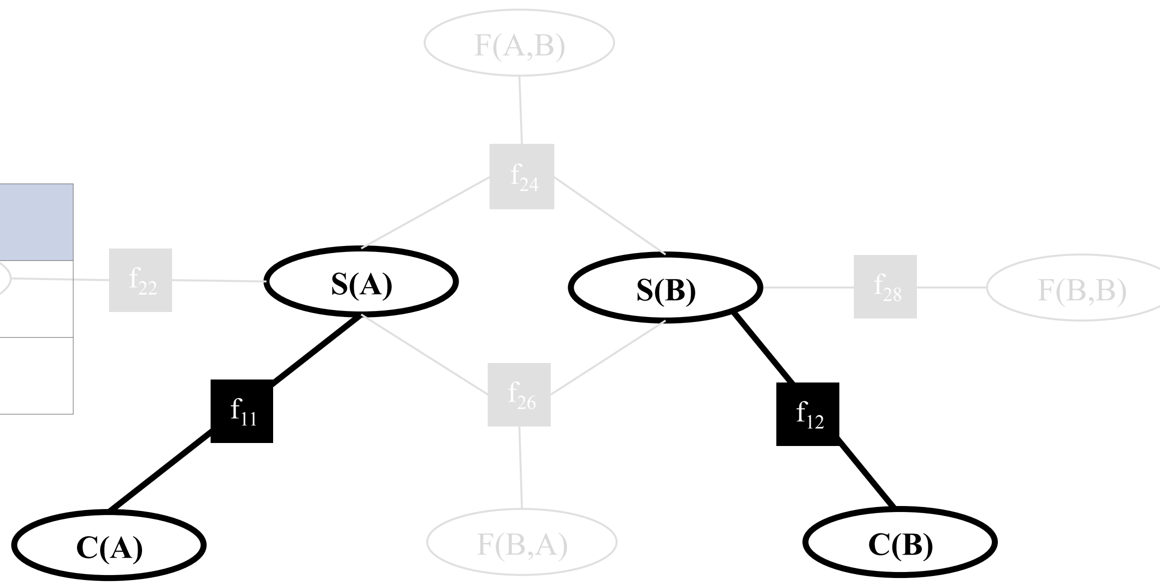


Grounding ≡ Database joins

Formula $Smokes(x) \Rightarrow Cancer(x)$	CNF $\neg Smokes(x) \vee Cancer(x)$
Predicates and domain $Smokes(person)$ $Cancer(person)$ $person = \{Anna, Bob\}$	Ground Clauses $\mathbf{f_{11}}$ $\neg Smokes(Anna) \vee Cancer(Anna)$ $\mathbf{f_{12}}$ $\neg Smokes(Bob) \vee Cancer(Bob)$

- Ground clauses corresponds to natural join: **Smokes ⋈ Cancer**

	Smokes	Cancer
	Person	Person
f_{11}	Anna	Anna
f_{12}	Bob	Bob



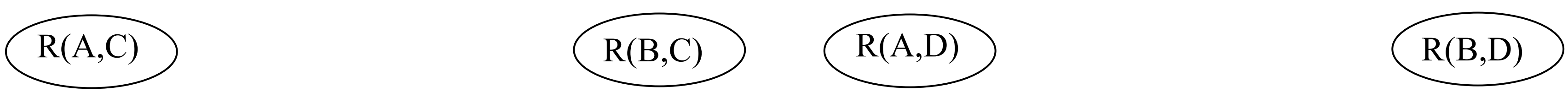
MLN Partitioning (partitioning a relation)

R(x,y)
Dom(x) = {Anna, Bob}
Dom(y) = {Charles, Debbie}

R

Attributes

x	y
Anna	Charles
Bob	Charles
Anna	Debbie
Bob	Debbie



MLN Partitioning (partitioning a relation)

R(x,y)
Dom(x) = {Anna, Bob}
Dom(y) = {Charles, Debbie}

$$R_1 = \sigma_{y=Charles}(R)$$

$$R_2 = \sigma_{y=Debbie}(R)$$

R_1

Attributes

x	y
Anna	Charles
Bob	Charles

R_2

Attributes

x	y
Anna	Debbie
Bob	Debbie

R(A,**C**) R(B,**C**) R(A,**D**) R(B,**D**)

MLN Partitioning (partitioning a relation)

R(x,y)
Dom(x) = {Anna, Bob}
Dom(y) = {Charles, Debbie}

$$R_1 = \sigma_{y=Charles}(R)$$

$$R_2 = \sigma_{y=Debbie}(R)$$

R_1

Attributes	x	y
	Anna	Charles
	Bob	Charles

R_2

Attributes	x	y
	Anna	Debbie
	Bob	Debbie

R(A,**C**) R(B,**C**) R(A,**D**) R(B,**D**)

Node1

Ground
R(x,y)
Dom(x) = {Anna, Bob}
Dom(y) = {**Charles**}

Node 2

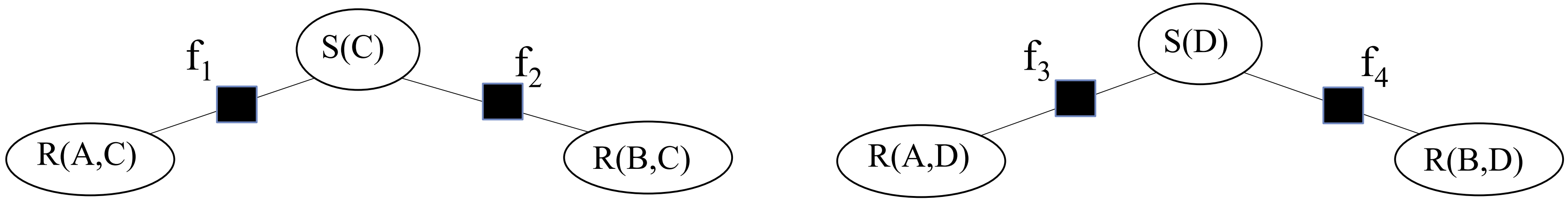
Ground
R(x,y)
Dom(x) = {Anna, Bob}
Dom(y) = {**Debbie**}

MLN Partitioning (co-partitioning relations)

$\mathbf{f:}\ \mathrm{R}(x,y) \vee \mathrm{S}(y)$

$\mathrm{Dom}(x) = \{\mathrm{Anna, Bob}\}$

$\mathrm{Dom}(y) = \{\mathrm{Charles, Debbie}\}$

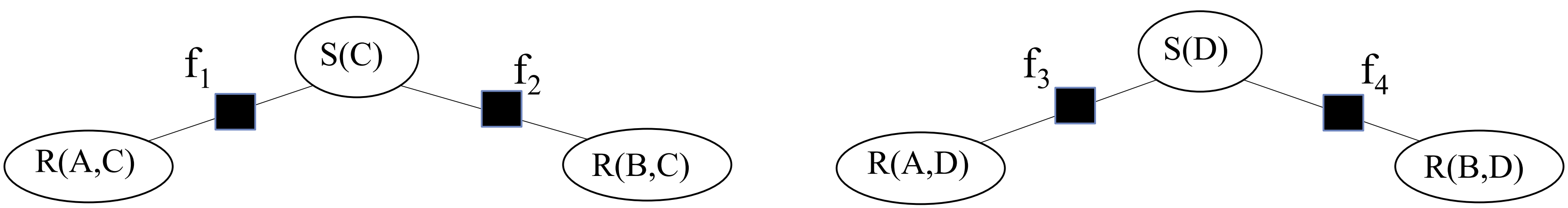


MLN Partitioning (co-partitioning relations)

$\mathbf{f}: R(x,y) \vee S(y)$
Dom(x) = {Anna, Bob}
Dom(y) = {Charles, Debbie}

	R		S
	x	y	y
$\mathbf{f_1}$	Anna	Charles	Charles
$\mathbf{f_2}$	Bob	Charles	Charles
$\mathbf{f_3}$	Anna	Debbie	Debbie
$\mathbf{f_4}$	Bob	Debbie	Debbie

$\mathbf{R} \bowtie_{\mathbf{R}.y=\mathbf{S}.y} \mathbf{S}$



MLN Partitioning (co-partitioning relations)

$\mathbf{f}: \mathrm{R}(x,y) \vee \mathrm{S}(y)$
Dom(x) = {Anna, Bob}
Dom(y) = {Charles, Debbie}

$$\mathrm{R}_1 = \sigma_{y=Charles}(\mathrm{R})$$
$$\mathrm{S}_1 = \sigma_{y=Charles}(\mathrm{S})$$

$$\mathrm{R}_2 = \sigma_{y=Debbie}(\mathrm{R})$$
$$\mathrm{S}_2 = \sigma_{y=Debbie}(\mathrm{S})$$

	R_1: x	R_1: y	S_1: y
$\mathbf{f_1}$	Anna	Charles	Charles
$\mathbf{f_2}$	Bob	Charles	Charles

$$\mathrm{R}_1 \bowtie_{\mathbf{R_1}.y=\mathbf{S_1}.y} \mathrm{S}_1$$

	R_2: x	R_2: y	S_2: y
$\mathbf{f_3}$	Anna	Debbie	Debbie
$\mathbf{f_4}$	Bob	Debbie	Debbie

$$\mathrm{R}_2 \bowtie_{\mathbf{R_2}.y=\mathbf{S_2}.y} \mathrm{S}_2$$

Local join

MLN Partitioning (co-partitioning relations)

f : $R(x,y) \vee S(y)$
$Dom(x)$ = {Anna, Bob}
$Dom(y)$ = {Charles, Debbie}

$$R_1 = \sigma_{y=Charles}(R)$$
$$S_1 = \sigma_{y=Charles}(S)$$

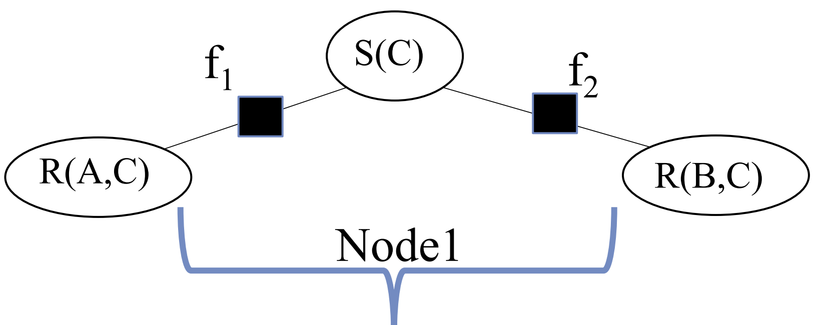
$$R_2 = \sigma_{y=Debbie}(R)$$
$$S_2 = \sigma_{y=Debbie}(S)$$

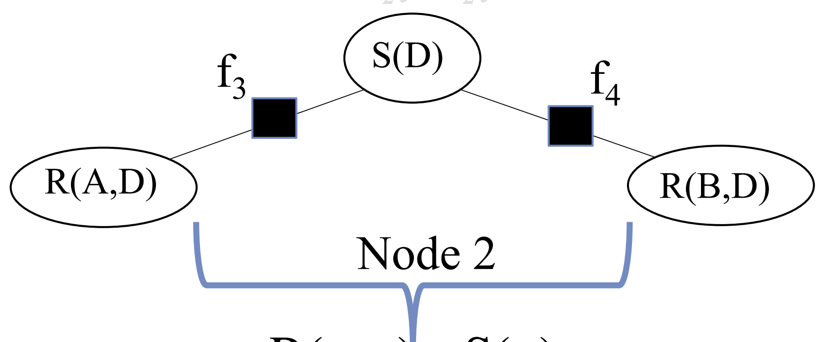
	R_1		S_1
	x	y	y
f_1	Anna	Charles	Charles
f_2	Bob	Charles	Charles

$$R_1 \bowtie_{R_1.y=S_1.y} S_1$$

	R_2		S_2
	x	y	y
f_3	Anna	Debbie	Debbie
f_4	Bob	Debbie	Debbie

$$R_2 \bowtie_{R_2.y=S_2.y} S_2$$





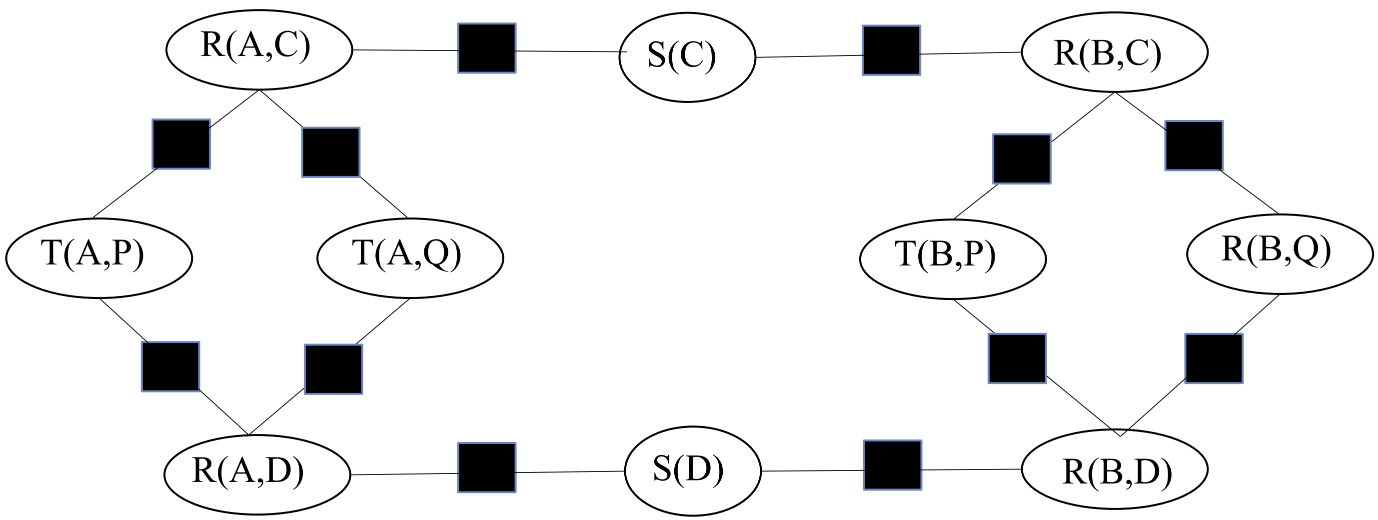
Ground $R(x,y) \vee S(y)$
$Dom(x)$ = {Anna, Bob}
$Dom(y)$ = {**Charles**}

Ground $R(x,y) \vee S(y)$
$Dom(x)$ = {Anna, Bob}
$Dom(y)$ = {**Debbie**}

MLN Partitioning

$f_1 : R(x,y) \vee S(y)$
$f_2 : R(x,y) \vee T(x,z)$

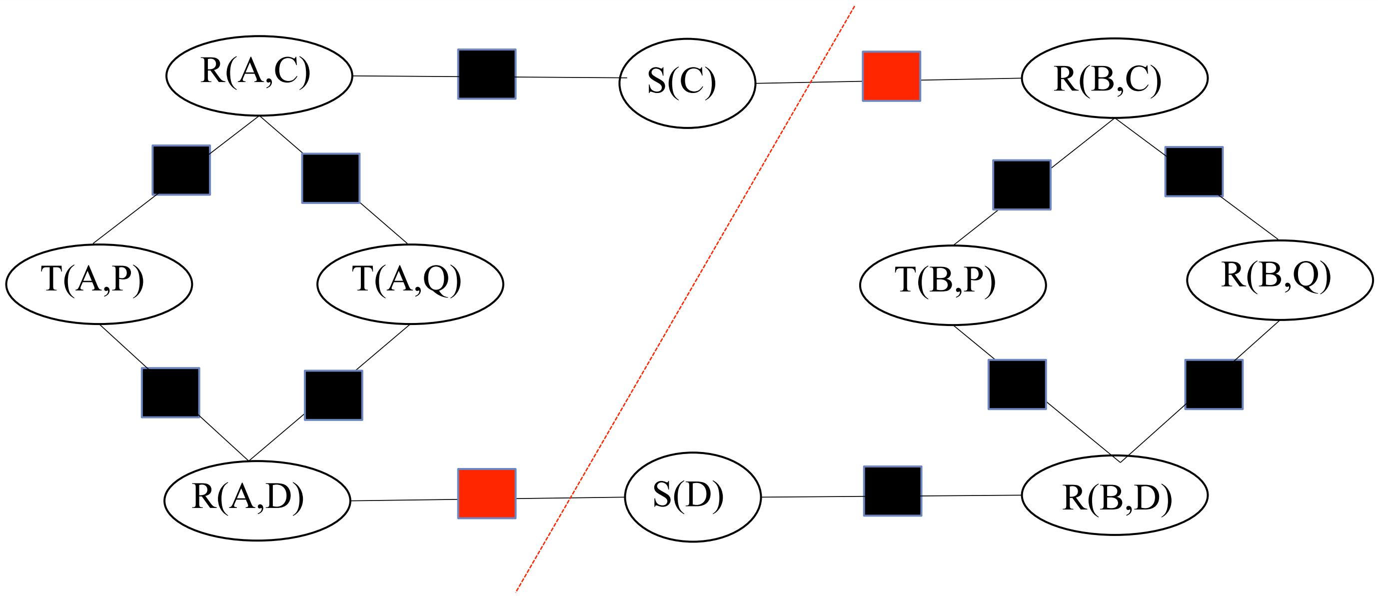
Dom(x) = {A, B}
Dom(y) = {C, D}
Dom(z) = {P, Q}



MLN Partitioning

$f_1 : R(x,y) \vee S(y)$
$f_2 : R(x,y) \vee T(x,z)$

Dom(x) = {A, B}
Dom(y) = {C, D}
Dom(z) = {P, Q}

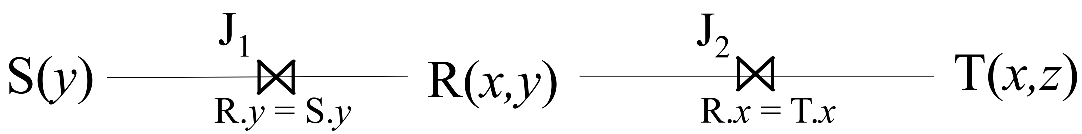


How to compute partitions at the Markov logic level ?

MLN Partitioning

$f_1 : R(x,y) \vee S(y)$
$f_2 : R(x,y) \vee T(x,z)$

$Dom(x) = \{A, B\}$
$Dom(y) = \{C, D\}$
$Dom(z) = \{P, Q\}$



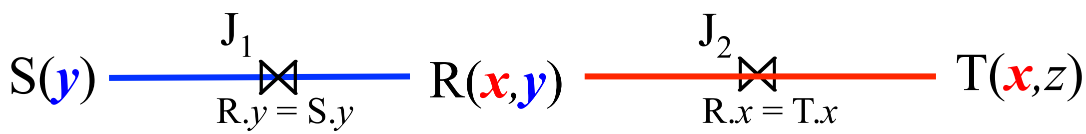
Partitioning at Rule level

- Model MLN as a join graph

MLN Partitioning

$f_1 : R(x,y) \vee S(y)$
$f_2 : R(x,y) \vee T(x,z)$

$Dom(x) = \{A, B\}$
$Dom(y) = \{C, D\}$
$Dom(z) = \{P, Q\}$



Partitioning at Rule level

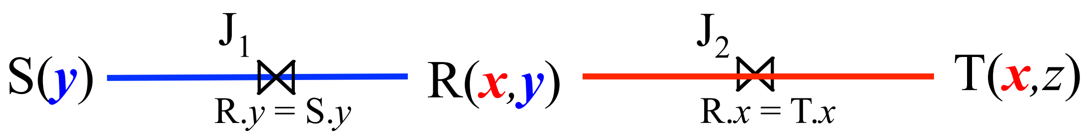
Co-partitioning strategy?

- Model MLN as a join graph

MLN Partitioning

$f_1 : R(x,y) \vee S(y)$
$f_2 : R(x,y) \vee T(x,z)$

$Dom(x) = \{A, B\}$
$Dom(y) = \{C, D\}$
$Dom(z) = \{P, Q\}$



$|J_1| = 4$ $|J_2| = 8$

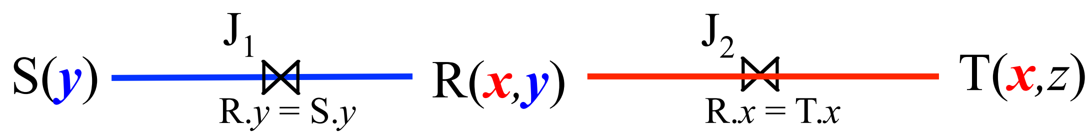
Partitioning at Rule level

- Model MLN as a join graph
- Estimate join sizes

MLN Partitioning

$f_1 : R(x,y) \vee S(y)$
$f_2 : R(x,y) \vee T(x,z)$

$\text{Dom}(x) = \{A, B\}$
$\text{Dom}(y) = \{C, D\}$
$\text{Dom}(z) = \{P, Q\}$



$|J_1| = 4$ $|J_2| = 8$

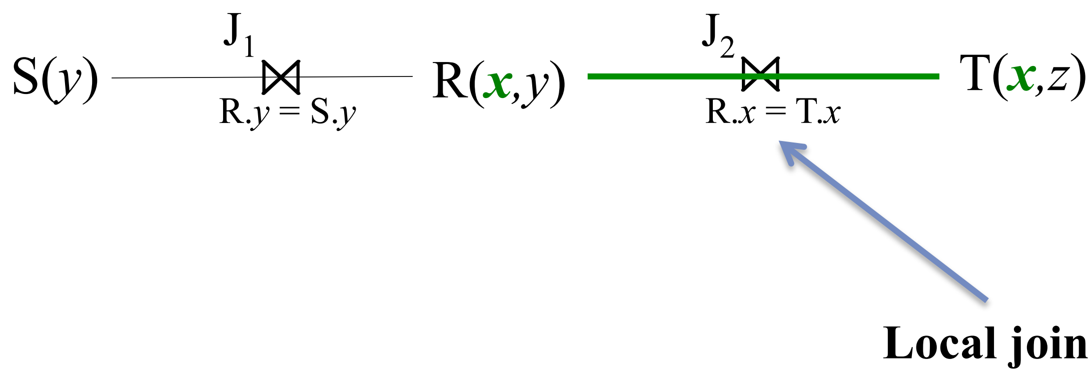
Partitioning at Rule level

- Model MLN as a join graph
- Estimate join sizes
- Co-partition to maximize size of local joins – optimization problem
- Encode as an ILP

MLN Partitioning

$f_1 : R(x,y) \vee S(y)$
$f_2 : R(x,y) \vee T(x,z)$

$Dom(x) = \{A, B\}$
$Dom(y) = \{C, D\}$
$Dom(z) = \{P, Q\}$



- $JS(J_1) = 4$
- $JS(J_2) = 8$
- Co-partition R and T on Dom(x)
- S on Dom(y)

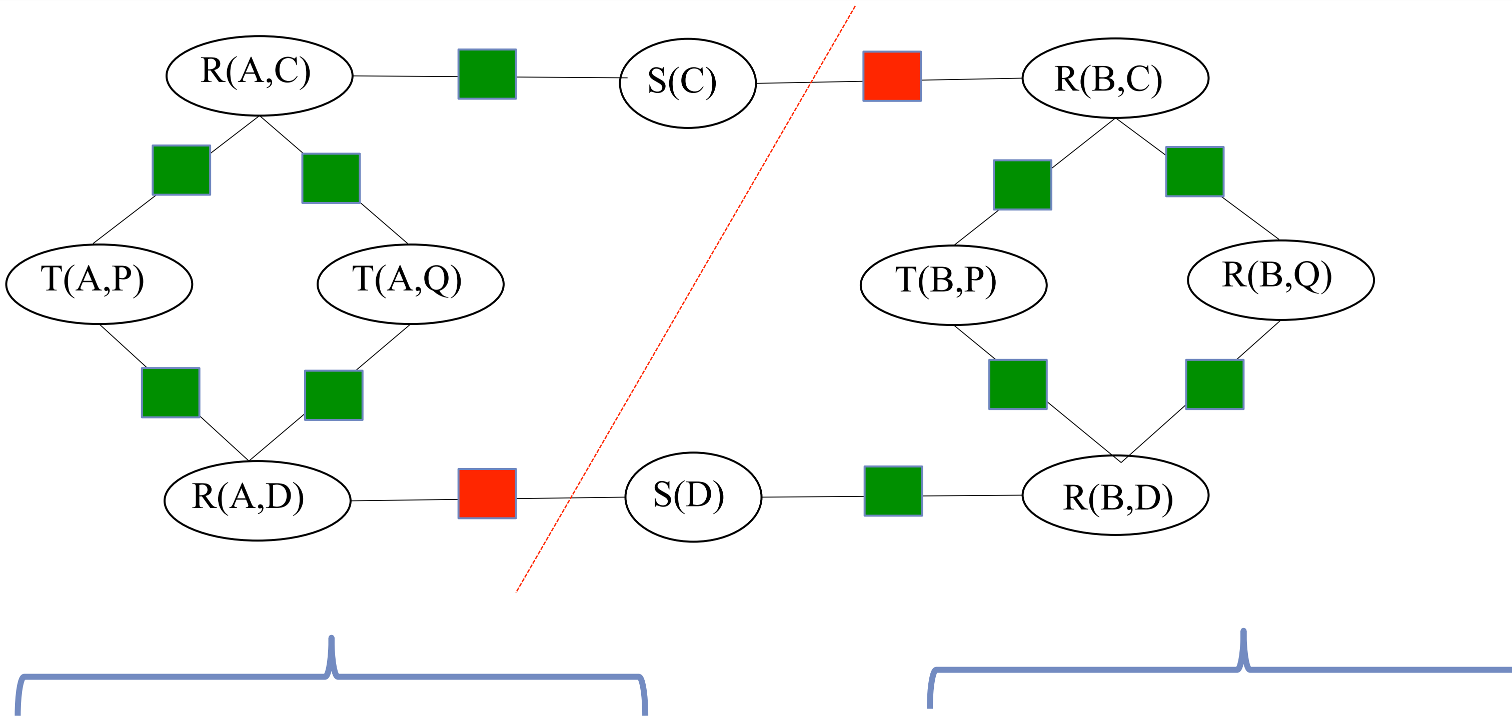
Ground
$R(x,y)$ $Dom(x) = \{$**A**$\}$, $Dom(y) = \{C,D\}$
$T(x,z)$ $Dom(x) = \{$**A**$\}$, $Dom(z) = \{P,Q\}$
$S(y)$ $Dom(y) = \{C\}$

Ground
$R(x,y)$ $Dom(x) = \{$**B**$\}$, $Dom(y) = \{C,D\}$
$T(x,z)$ $Dom(x) = \{$**B**$\}$, $Dom(z) = \{P,Q\}$
$S(y)$ $Dom(y) = \{D\}$

MLN Partitioning

$f_1 : R(x,y) \vee S(y)$
$f_2 : R(x,y) \vee T(x,z)$

$Dom(x) = \{A, B\}$
$Dom(y) = \{C, D\}$
$Dom(z) = \{P, Q\}$



Ground
$R(x,y)$ $Dom(x) = \{\mathbf{A}\}$, $Dom(y) = \{C,D\}$
$T(x,z)$ $Dom(x) = \{\mathbf{A}\}$, $Dom(z) = \{P,Q\}$
$S(y)$ $Dom(y) = \{C\}$

Ground
$R(x,y)$ $Dom(x) = \{\mathbf{B}\}$, $Dom(y) = \{C,D\}$
$T(x,z)$ $Dom(x) = \{\mathbf{B}\}$, $Dom(z) = \{P,Q\}$
$S(y)$ $Dom(y) = \{D\}$

MLN Partitioning (evaluation)

Comparison of various graph partitioning approaches for *k* partitions

k	Approach	Factors in cut	Weight of cut	Balancing	Runtime
$k = 2$	PaToH	4678	1109.04	0.000	948.288s
	Tuffy	4686	1108.66	0.000	1.092s
	MLN part.	**4690**	**1109.47**	**0.000**	**0.003s**
$k = 4$	PaToH	63001	64500.40	0.012	952.254s
	Tuffy	7040	1662.46	0.000	1.288s
	MLN part.	**7023**	**1662.84**	**0.000**	**0.003s**

Outline

- Background and Motivation
- Parallel Inference
- Parallel Grounding
- Conclusion

Conclusions

Markov logic networks

- Incomplete database + first order rules
- Scalability challenges

First fully parallel approach to MLN inference

- Partition the MLN before grounding
- Ground partitions in parallel
- Run parallel inference

Preliminary experimental results

- Orders of magnitude faster partitioning at similar quality
- Parallel inference effective

Conclusions

Markov logic networks

Questions?

- Incomplete database + first order rules
- Scalability challenges

First fully parallel approach to MLN inference

- Partition the MLN before grounding
- Ground partitions in parallel
- Run parallel inference

Thank you!

Preliminary experimental results

- Orders of magnitude faster partitioning at similar quality
- Parallel inference effective