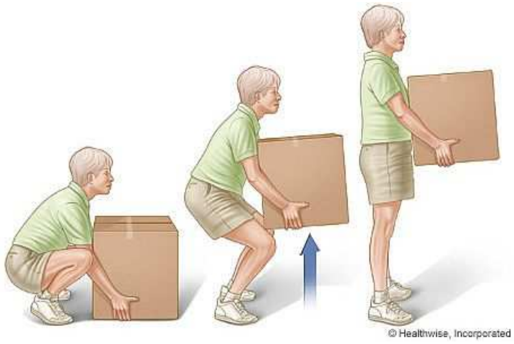
Sriraam
Natarajan

Babak
Ahmadi

Fabian
Hadiji

Scott
Sanner

Youssef El
Massaoudi



Lifted Message Passing

Statistical Relational
Activity Mining

NeSy 2010 @ AAAI 2010, Atlanta, USA July 11, 2010

Rorschach Test

Etzioni's Rorschach Test for Computer Scientists

Moore's Law?

Storage Capacity?

Number of Facebook Users?

Number of Scientific Publications?

Number of Web Pages?

Number of Actions?

Computing 2020: Science in an Exponential World

"The amount of scientific data is doubling every year"
[Szalay,Gray; *Nature* 440, 413-414 (23 March 2006)]

How to deal with millions of images ?

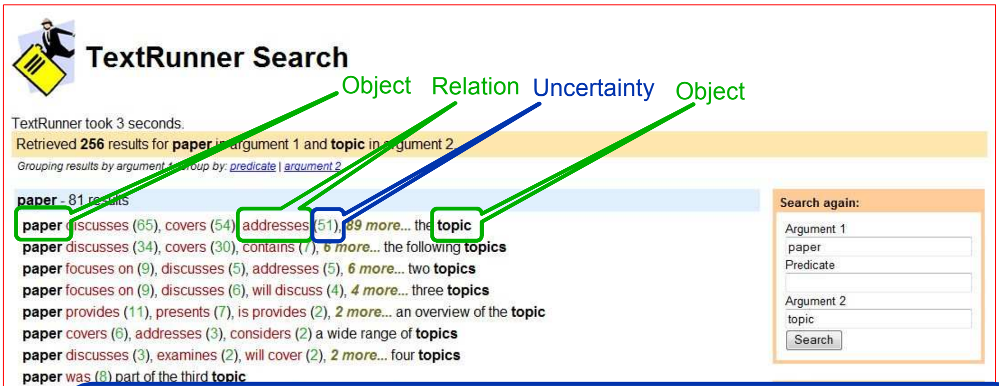
How to deal with millions of inter-related research papers ?

How to accumulate general knowledge automatically from the Web ?

How to deal with billions of shared users' perceptions stored at massive scale ?

How do realize the vision of social search?

[Etzioni et al. ACL08] http://www.cs.washington.edu/research/textrunner/



"Programs will consume, combine, and correlate everything in the universe of structured information and help users reason over it." [S. Parastatidis et al., Communications of the ACM Vol. 52(12):33-37]

Artificial Intelligence in an Exponential World

AI = Structured (Data + Model + Reasoning)

- Real world is structured in terms of objects and relations
  - Relational knowledge can reveal additional correlations between variables of interest . Abstraction allows one to compactly model general knowledge and to move to complex inference

[Fergus et al. 30(11) 2008; Halevy et al., IEEE Intelligent Systems, 24 2009]

- Most effort has gone into the modeling part
- How much can the data itself help us to solve a problem?

(First-order) Logic handles Complexity

E.g., rules of chess (which is a tiny problem):

1 page in first-order logic,

~100000 pages in propositional logic,

~100000000000000000000000000000000000000 pages as atomic-state model

- Many types of entities
- Relations between them
- Arbitrary knowledge

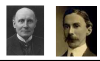
Explicit enumeration

Logic

true/false

5th C B.C.

19th C



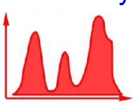
atomic | propositional | first-order/relational

Probability handles Uncertainty

17th C

20th C

Probability



- Sensor noise
- Human error
- Inconsistencies
- Unpredictability

Explicit enumeration

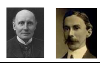
- Many types of entities
- Relations between them
- Arbitrary knowledge

Logic

true/false

5th C B.C.

19th C



atomic | propositional | first-order/relational

Will Traditional AI Scale ?

"Scaling up the environment will inevitably overtax the resources of the current AI architecture."

MIND THE GAP

Probability

17th C | 20th C

- Sensor noise
- Human error
- Inconsistencies
- Unpredictability

Explicit enumeration

- Many types of entities
- Relations between them
- Arbitrary knowledge

Logic

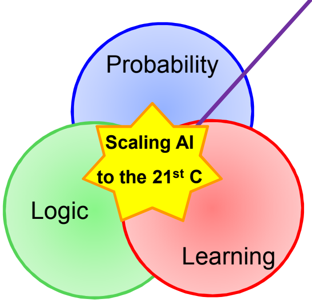
true/false

5th C B.C. | 19th C

atomic | propositional | first-order/relational

The real world is complex and uncertain

Let's deal with uncertainty, objects, and relations jointly



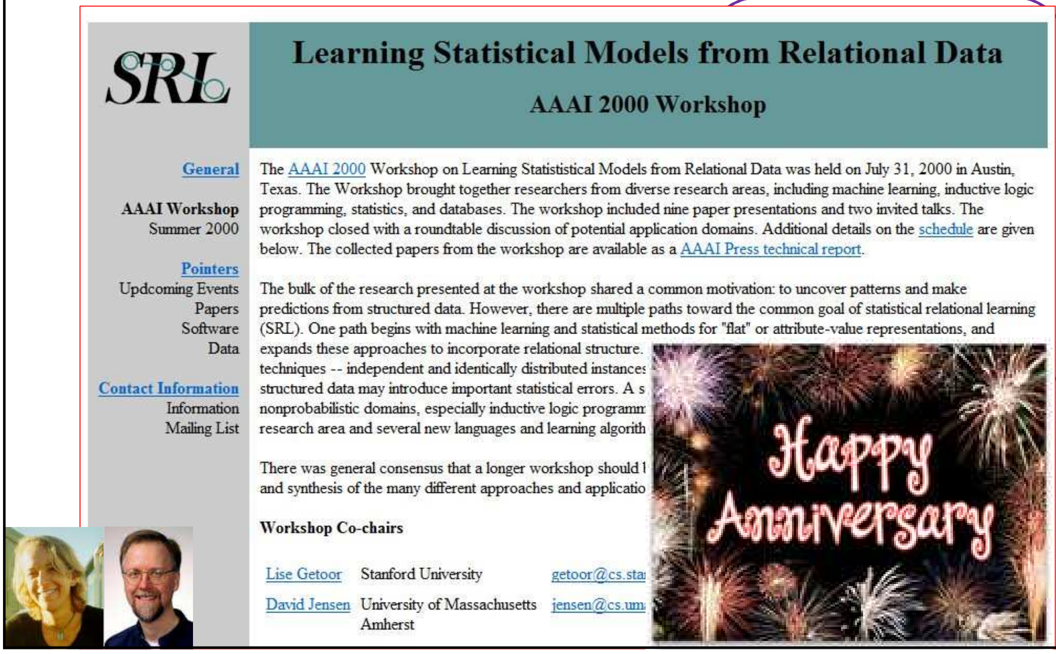
- Natural domain modeling: objects, properties, relations
- Compact, natural probability models
- Properties of entities can depend on properties of related entities
- Generalization over a variety of situations
- Pragmatic view on probabilistic logics / inconsistency

… unifies logical and statistical AI,

… solid formal foundations,

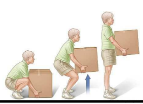
… is of interest to many communities.

The real world is complex and uncertain



Today, we can …

- … learn probabilistic relational models automatically from millions of inter-related objects
- … generate optimal plans and learn to act optimally in uncertain environments involving millions of objects and relations among them
- … perform lifted probabilistic inference avoiding explicit state enumeration by manipulating first-order state representations directly
- … exploit shared factors to speed up message-passing algorithms for relational inference but also for classical propositional inference such as solving SAT problems

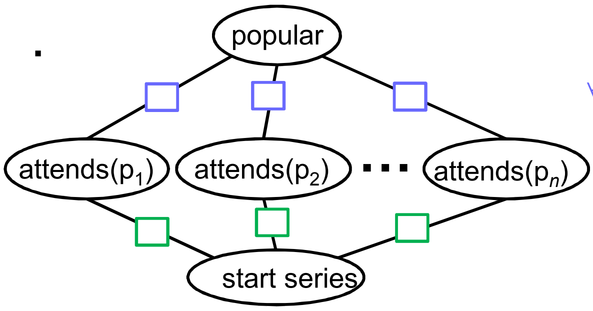


Lifted Inference

[Pfeffer et al. 1999; Poole 2003; de Salvo Braz et al. 2005]

- Example: Inviting ***n*** people to a workshop

Factor graph



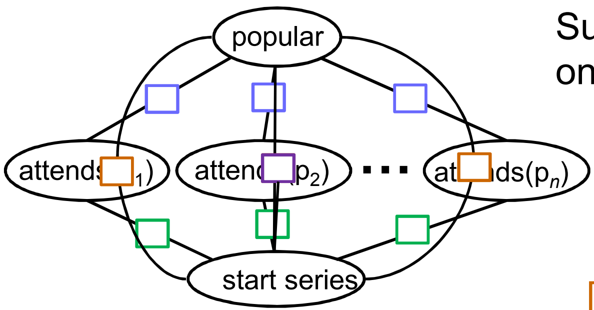
Parfactors

$\forall X.\ \phi_1(\text{popular}, \text{attends}(X))$

$\forall X.\ \phi_2(\text{attends}(X), \text{series})$

- Lifted inference exploits symmetries revealed by relational model

Variable Elimination



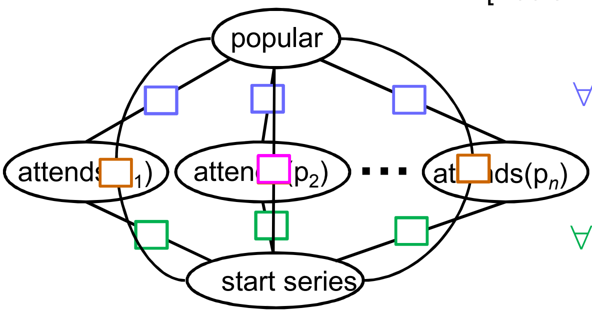
Sum out non-query variables one by one

Time is linear in number of invitees ***n***

$$\underbrace{\sum_{\text{att}(p_1)} \phi_1(\text{pop}, \text{att}(p_1))\ \phi_2(\text{att}(p_1), \text{ser})}_{\phi'(\text{pop, ser})}$$

First-Order Variable Elimination

[Poole 2003; de Salvo Braz *et al.* 2005]



$\forall X.\ \phi_1$(popular, attends(X))

$\forall X.\ \phi_2$(attends(X), series)

⇩

$\forall X.\ \phi'$(popular, series)

⇩

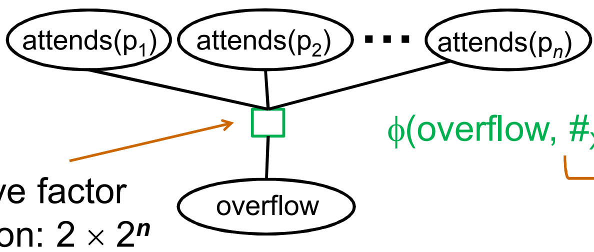
$\phi'(\text{popular, series})^{n}$

Sum out all attends(X) variables at once

Time is constant in ***n***

Symmetry Within Factors

[cf Zhang & Poole 1996; Gupta et al. 2007, Milch et al. 2008]



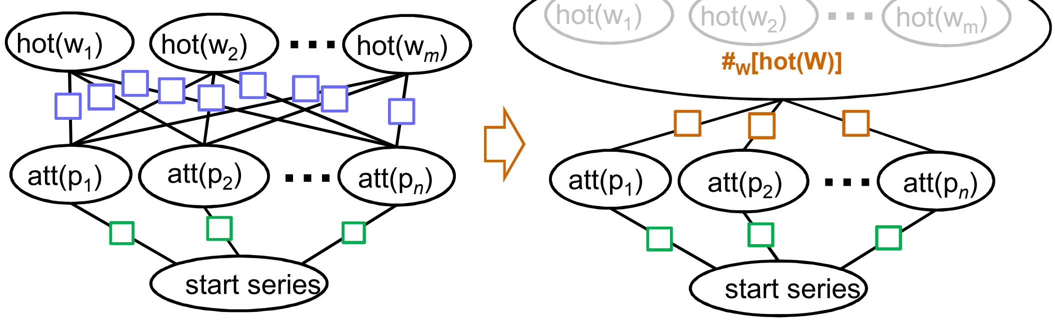
Size of naïve factor representation: 2×2^{n}

$\underbrace{\phi(\text{overflow}, \#_X[\text{attends}(X)])}_{\text{counting formula}}$

- Values of counting formula are histograms counting how many objects X yield each possible value of attends(X)
  - Only ***n***+1 histograms, e.g., [50, 0], [49, 1], …, [0, 50]
  - Factor size now $2 \times (n+1)$: linear in ***n***

Example: Competing Workshops

$\forall W\ \forall X.\ \phi(hot(W), att(X))$

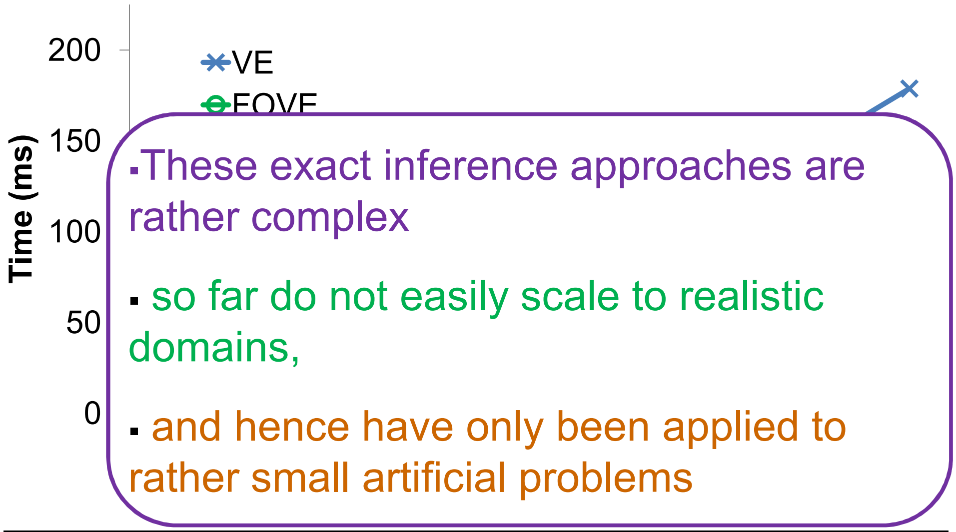


Can't sum out attends(X) without joining all the hot(W) variables

Create counting formula on hot(W), then sum out attends(X) at lifted level

Conversion to counting formulas creates new opportunities for lifted elimination

Results: Competing Workshops



- These exact inference approaches are rather complex
- so far do not easily scale to realistic domains,
- and hence have only been applied to rather small artificial problems

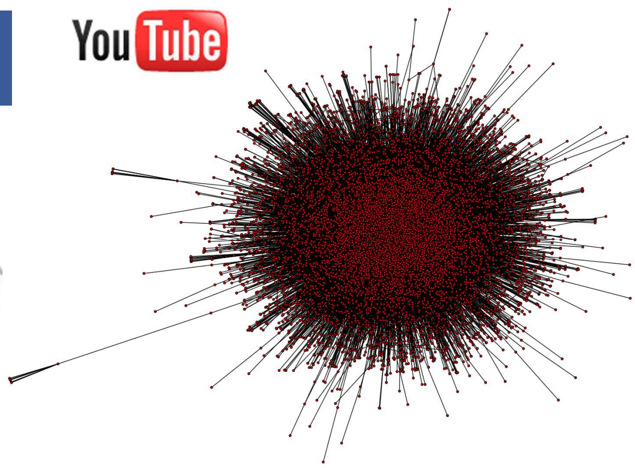
How do you spend your spare time?

YouTube like media portals have changed the way users access media content in the Internet

Every day, millions of people visit social media sites such as Flickr, YouTube, and Jumpcut, among others, to share their photos and videos, …

while others enjoy themselves by searching, watching, commenting, and rating the photos and videos; what your friends like will bear great significance for you.

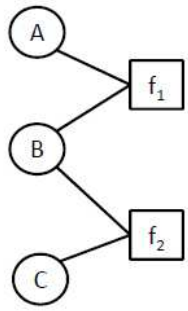
How do you efficiently broadcast information?



Content Distribution using Belief Propagation

[Bickson et al. 04]

- Approximate inference
- Compute the marginal P(xi|X) for each xi with local computations only
- Computer vision, combinatorial problems, SAT, NLP, …



A	B	f_1
True	True	1.2
True	False	1.4
False	True	2.0
False	False	0.4

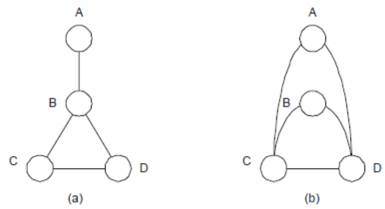
C	B	f_2
True	True	1.2
True	False	1.4
False	True	2.0
False	False	0.4

$$\mu_{X\to f}(x) = \prod_{h\in \mathrm{nb}(X)\setminus\{f\}} \mu_{h\to X}(x)$$

$$\mu_{f\to X}(x) = \sum_{\neg\{x\}} \left(f(\mathbf{x}) \prod_{y\in \mathrm{nb}(f)\setminus\{X\}} \mu_{y\to f}(y) \right)$$

Content Distribution using Belief Propagation

[Bickson et al. WDAS04]



$\psi_{BB}(X_B)$	3 from A	1 from C	3 from C	3 from D
Uniform policy	1/4	1/4	1/4	1/4
Rarest part first	1/6	1/2	1/6	1/6

Table 2. Possible actions for the node B in the example graph shown in Figure 1

$\psi_{BD}(X_B, X_D)$	3 from A	**2 from C**	**3 from C**	3 from D
1 from B	1/8	1/8	1/8	1/8
2 from C	1/2	ϵ	ϵ	1/2
3 from C	1/2	ϵ	ϵ	1/2

Table 3. Example of the edge potentials for the edge BD for the graph shown in Figure 2. The matrix rows are then normalized.

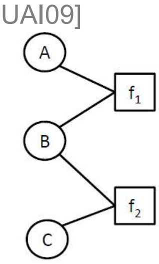
A lot of shared factors, so use lifted belief propagation

http://www-kd.iai.uni-bonn.de/index.php?page=software_details&id=16

Lifted Belief Propagation

[Singla, Domingos AAAI08, K, Ahmadi, Natarajan UAI09]

Counting shared factors can result in great efficiency gains for (loopy) belief propagation



A	B	f_1
True	True	1.2
True	False	1.4
False	True	2.0
False	False	0.4

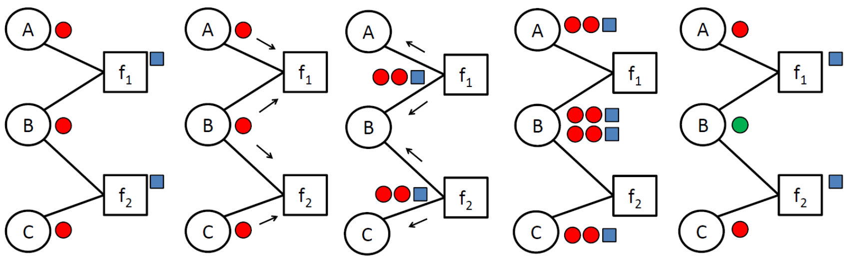
C	B	f_2
True	True	1.2
True	False	1.4
False	True	2.0
False	False	0.4

identical

Shared factors appear more often than you think in relevant real world problems

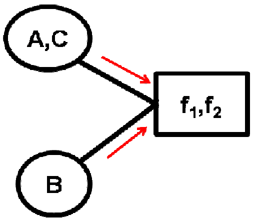
http://www-kd.iai.uni-bonn.de/index.php?page=software_details&id=16

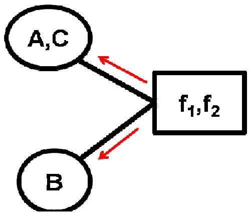
Step 1: Compression



http://www-kd.iai.uni-bonn.de/index.php?page=software_details&id=16

Step 2: Modified Belief Propagation



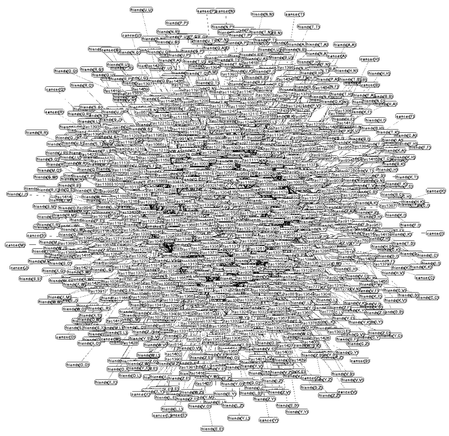


$$\mu_{\mathfrak{X}\to\mathfrak{f}}(x) = \mu_{\mathfrak{f}\to\mathfrak{X}}(x)^{\boxed{c(\mathfrak{f},\mathfrak{X})-1}} \cdot \prod_{\mathfrak{h}\in \mathrm{nb}(\mathfrak{X})\setminus\{\mathfrak{f}\}} \mu_{\mathfrak{h}\to\mathfrak{X}}(x)^{\boxed{c(\mathfrak{h},\mathfrak{X})}}$$

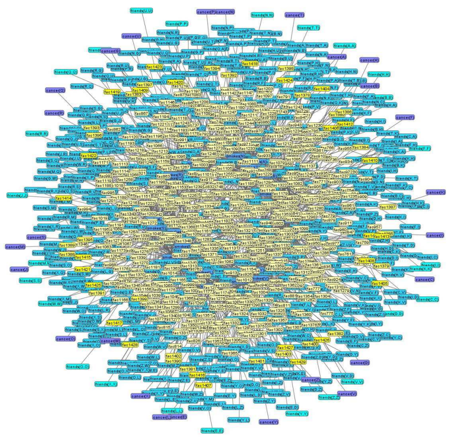
$$\mu_{\mathfrak{f}\to\mathfrak{X}}(x) = \prod_{\mathfrak{f}\in \mathrm{nb}(\mathfrak{X}_i)} \mu_{\mathfrak{f}\to x_i}(x_i)$$

$$b_i(x_i) = \prod_{\mathfrak{f}\in \mathrm{nb}(\mathfrak{X}_i)} \mu_{\mathfrak{f}\to x_i}(x_i)^{\boxed{c(\mathfrak{f},\mathfrak{X})}}$$

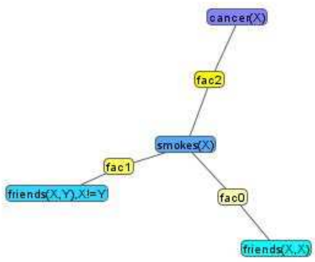
Social Network Analysis



Social Network Analysis



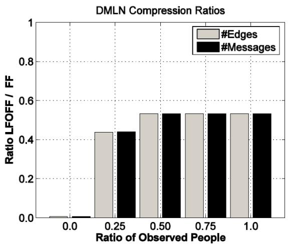
Social Network Analysis

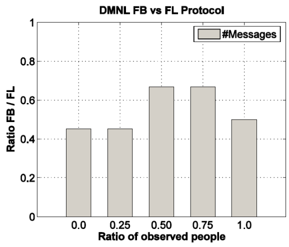


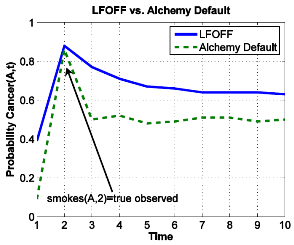
Lifted First-order Factored Frontier

Apriori most people do not smoke	$\neg\mathtt{Smokes}(x,0)$
Apriori most people do not have cancer	$\neg\mathtt{Cancer}(x,0)$
Apriori most people are not friends	$\neg\mathtt{Friends}(x,y,0)$
Smoking causes cancer	$\mathtt{Smokes}(x,t) \Rightarrow \mathtt{Cancer}(x,t)$
Friends have similar smoking habits	$\mathtt{Friends}(x,y,t) \Rightarrow (\mathtt{Smokes}(x,t) <=> \mathtt{Smokes}(y,t))$
Most friends stay friends	$\mathtt{Friends}(x,y,t) \Leftrightarrow \mathtt{Friends}(x,y,\mathtt{succ}(t))$
Most smokers stay smokers	$\mathtt{Smokes}(x,t) \Leftrightarrow \mathtt{Smokes}(x,\mathtt{succ}(t))$

20 people over 10 time steps. Max number of friends 5. Cancer never observed. Time step randomly selected.





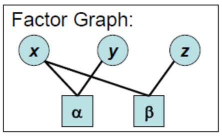


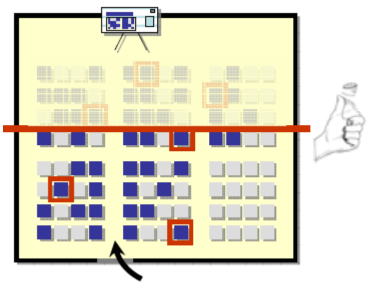
Lower Bound on Model Count of CNF

- BPCount [Kroc et al 08]
  - BP used to estimate marginals
  - Provable bound

e.g. SAT Problem:

$$\underbrace{(x \vee y)}_{\alpha} \wedge \underbrace{(\neg x \vee z)}_{\beta}$$



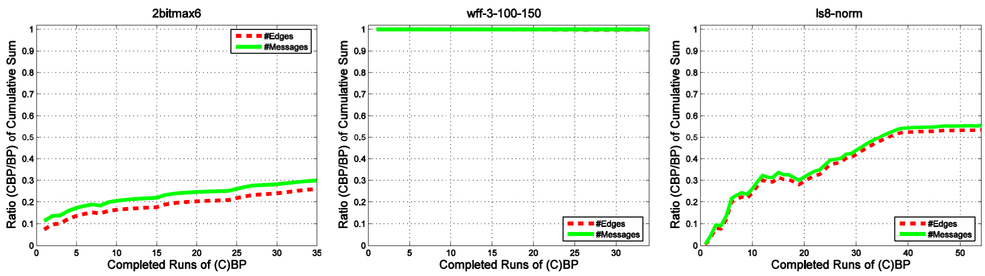


Idea:

- Identify a "balanced" row split or column split (roughly equal number of solutions on each side)
  - Use marginals for estimate
- Pick one side *at random*
- Count on that side recursively
- Multiply result by 2

[similar to decimation]

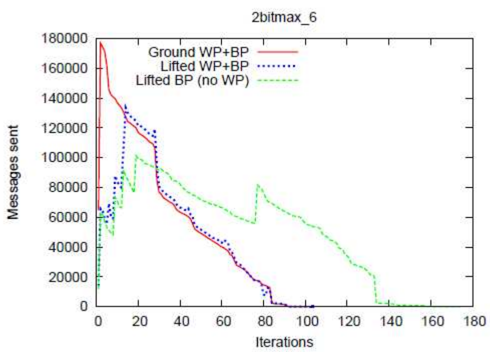
Model Counting

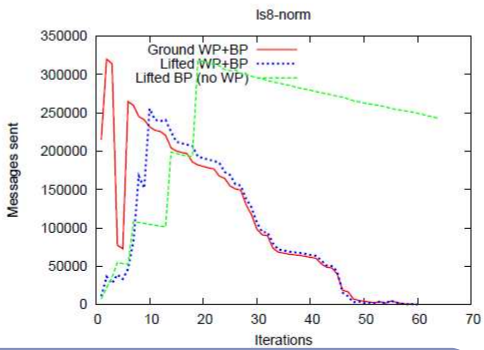


Satisfied by Lifted Message Passing?

Message Passing for Satisfiability

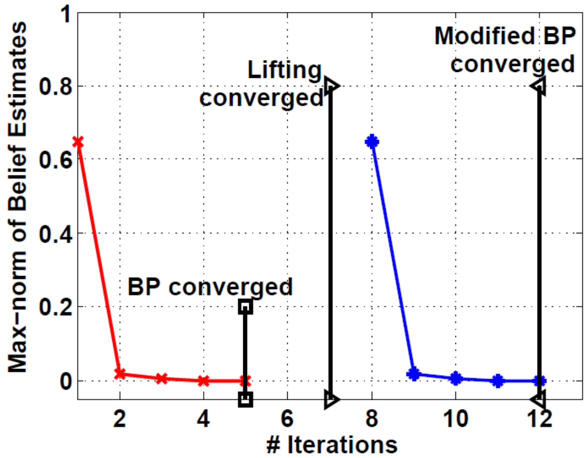
- Warning and survey propagation can also be lifted
- Enables lifted treatment of both prob. and det. knowledge





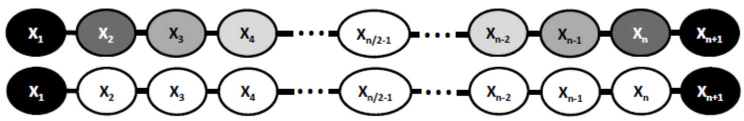
Gaussian Belief Propagation can also be lifted! Lifted Solvers for Systems of Linear Equations? Lifted Page Rank? Lifted HITS? Lifted Kalman Filter? …

Content Distribution (Gnutella): Lifted BP vs. BP



Message Errors to the Rescue!

- **Ihler et al. 05:** BP message errors decay along paths
- LBP may spuriously assume some nodes send and receive different messages and, hence, produce pessimistic lifted network



Make use of decaying message errors already at lifting time

Informed Lifted Belief Propagation [El Massaoudi et al. AAAI10]

Algorithm 1: iLBP – informed Lifted BP. We use $b_i(x_i)$ resp. $m_i(x_i)$ to denote the unnormalized beliefs resp. messages of both variable node X_i and variable nodes covered by supernodes $\mathfrak{X}_i$.

Data: A factor graph G with variable nodes X and factors f, Evidence E
Result: Unnormalized marginals $b_i(x_i)$ for all supernodes and, hence, for all variable nodes

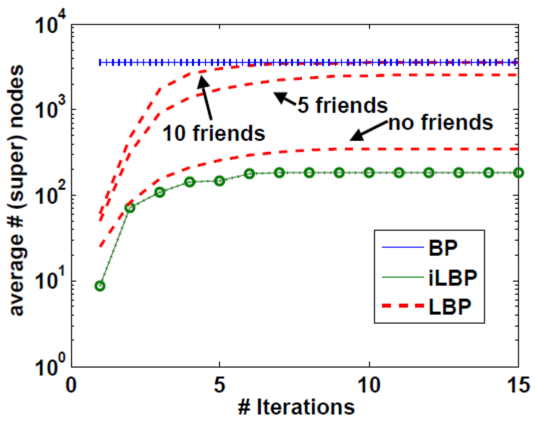
```
Colorize X and f w.r.t. E;
𝔊 ← one iteration CP;
Initialize messages for 𝔊;
(b_i(x_i), m_i(x_i)) ← one iteration MBP on 𝔊;
Colorize all X_i s according to m_i(x_i);
while b_i(x_i)s have not converged do
    𝔊′ ← one iteration CP (based on new colors);
    Initialize novel supernodes using b_i(x_i) and m_i(x_i);
    (b_i(x_i), m(x_i)) ← one iteration of MBP on 𝔊′;
    foreach supernode 𝔛 in 𝔊 do
        if the m(x_i)s of the X_i s in 𝔛 differ then
            Colorize all X_i in 𝔛 according to m(x_i)
Return b_i(x_i) for all supernodes
```

Iterate

Refine Lifting

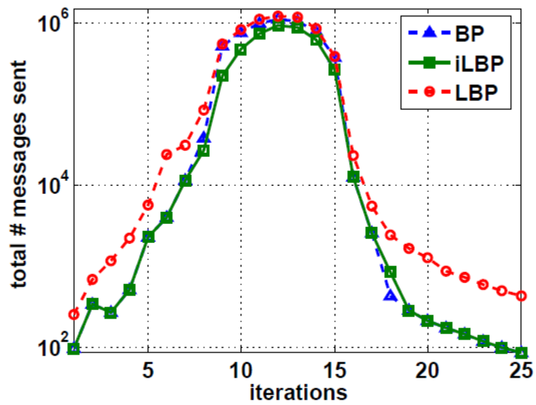
Modified BP

Social Networks



Lifted Content Distribution

- 1 file, Gnutella snapshort
  - 10876 nodes
  - 39994 edges
- iLBP 4.272.164 mess.
- < BP 5.761.952 mess.
- < LBP 6.381.516 mess.



- On a different network:
  - iLBP 1.972.662 < LBP 2.962.311 < BP 5.761.952

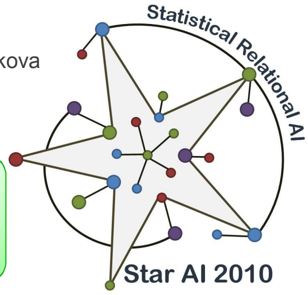
Conclusions

- SRL $\geq$ Objects&Relations + Probabilities + Machine Learning
- SRL works and covers the whole AI spectrum
  - Lifted Reasoning / SAT / Message Passing
  - Relational Machine Learning
  - Relational POMDPs [Sanner, K, AAAI10]
- Relational / Symbolic Neural Networks?
  - **Deep Relational Networks?**
  - **Relational LPs?**
  - **Lifted Boltzmann Machines?**
  - **…**

http://www.biostat.wisc.edu/~natarasr/starai.html

StarAI@AAAI-10

Together with S. Russell, L. Kaelbling, A.Halevy, S. Natarajan, and L. Milhalkova



Let's explore the minimal perturbations required for each of the AI areas to start using SR techniques

- Planning: from PDDL to SRPDDL?
- MDPs: from 2-TBN models to DSRL models?
- Vision: from graphical models and scene grammars to SR generative models?
- NLP: from unification grammars to SR-unification grammars?
- KR: ontologies and event calculus to SR models?
- …

Thanks for your attention