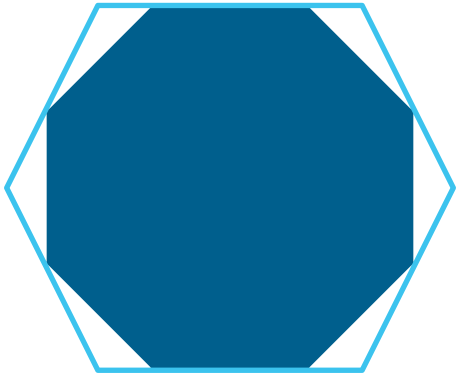
Unifying Local Consistency and MAX SAT Relaxations for Scalable Inference with Rounding Guarantees

Stephen H. Bach	Bert Huang	Lise Getoor
Maryland	Virginia Tech	UC Santa Cruz

AISTATS 2015

This Talk

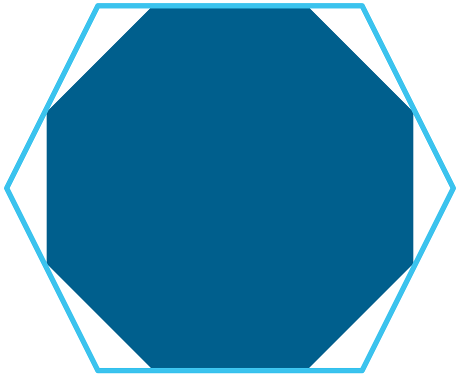
- Markov random fields capture **rich dependencies** in structured data, but **inference is NP-hard**
- Relaxed inference can help, but techniques have **tradeoffs**
- Two approaches:



Local Consistency Relaxation

MAX SAT Relaxation

Takeaways



Local Consistency Relaxation

$\supset$

MAX SAT Relaxation

- We can **combine** their advantages: **quality guarantees** and **highly scalable message-passing algorithms**
- New inference algorithm for broad class of **structured, relational models**

Modeling Relational Data with Markov Random Fields

Markov Random Fields

- Probabilistic model for high-dimensional data:

$$P(\boldsymbol{x}) \propto \exp\left(\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x})\right)$$

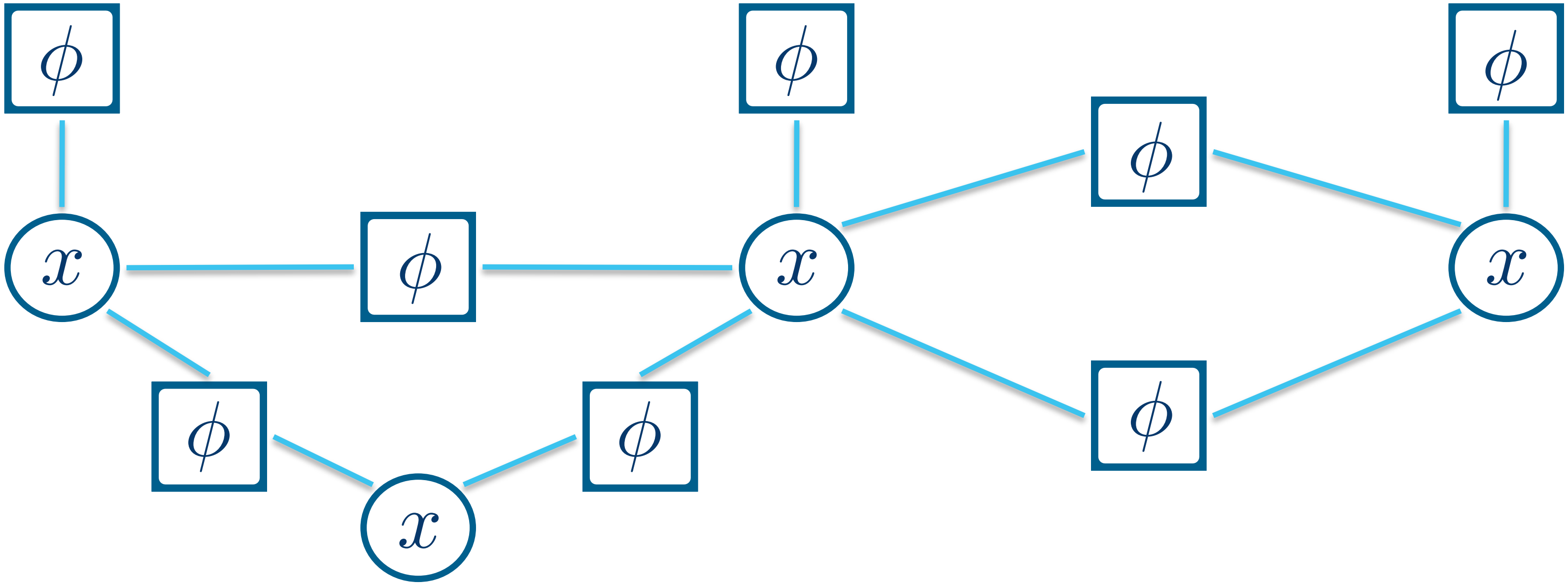
- The random variables $\boldsymbol{x}$ represent the data, such as whether a person has an attribute or whether a link exists

- The potentials ϕ score different configurations of the data

- The weights $\boldsymbol{w}$ scale the influence of different potentials

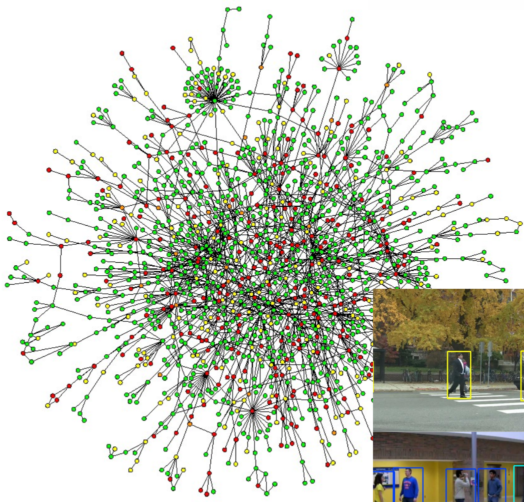
Markov Random Fields

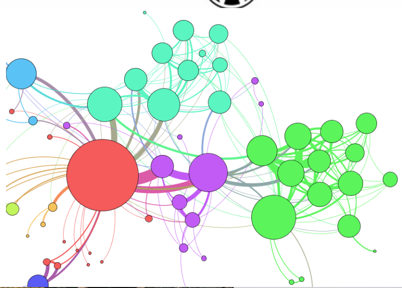
- Variables and potentials form graphical structure:



Modeling Relational Data

- Many important problems have relational structure
- Common to use logic to describe probabilistic dependencies
- Relations in data map to logical predicates





Logical Potentials

- One way to compactly define MRFs is with first-order logic, e.g., Markov logic networks
  [Richardson and Domingos, 2006]

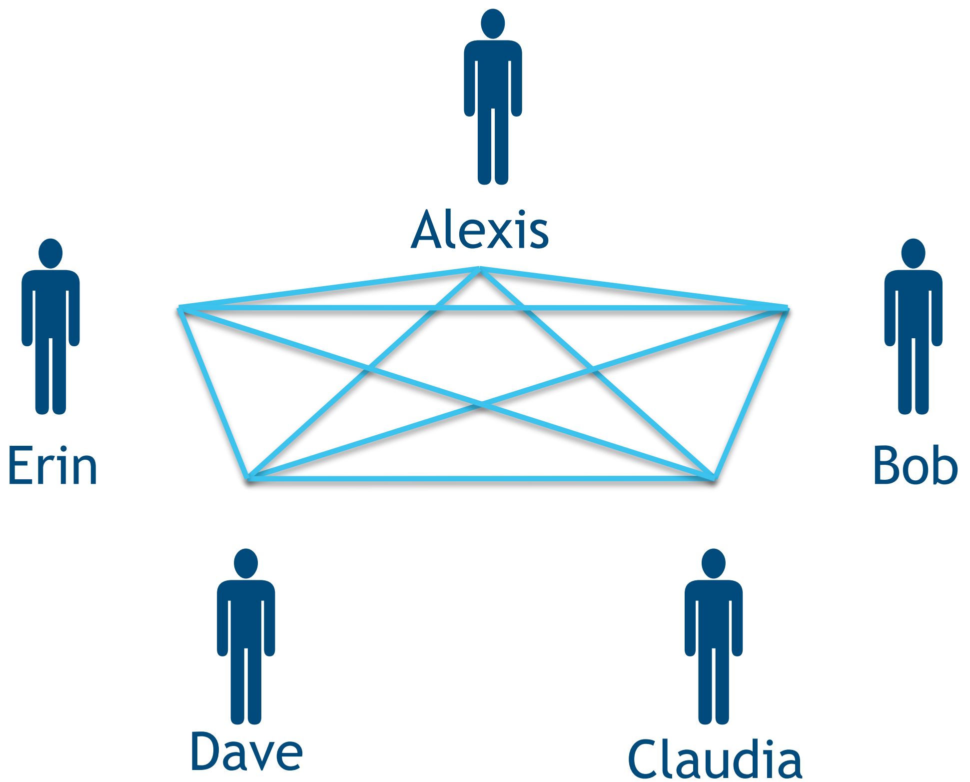
$$5.0 \ : \ \textsc{Friends}(X, Y) \wedge \textsc{Smokes}(X) \implies \textsc{Smokes}(Y)$$

- Each first-order rule is a template for potentials
  - Ground out rule over relational data
  - The truth table of each ground rule is a potential
  - Each potential’s weight comes from the rule that templated it

Logical Potentials: Grounding

$$5.0 : \textsc{Friends}(X, Y) \wedge \textsc{Smokes}(X) \implies \textsc{Smokes}(Y)$$

$$5.0 : \textsc{Friends}(\texttt{Alexis}, \texttt{Bob}) \wedge \textsc{Smokes}(\texttt{Alexis}) \implies \textsc{Smokes}(\texttt{Bob})$$



Logical Potentials

- Let $\boldsymbol{R}$ be a set of rules, where each rule R_j has the general form

$$w_j \quad : \quad \left(\bigvee_{i \in I_j^+} x_i \right) \bigvee \left(\bigvee_{i \in I_j^-} \neg x_i \right)$$

  - Weights $w_j \geq 0$ and sets I_j^- and I_j^+ index variables

MAP Inference

- MAP (*maximum a posteriori*) inference seeks a most-probable assignment to the unobserved variables

- MAP inference is

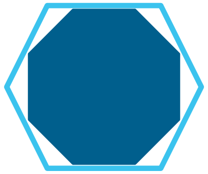
$$\arg\max_{\boldsymbol{x}} P(\boldsymbol{x}) \quad \equiv \quad \arg\max_{\boldsymbol{x}\in\{0,1\}^n} \sum_{R_j \in \boldsymbol{R}} w_j \left(\left(\bigvee_{i \in I_j^+} x_i \right) \bigvee \left(\bigvee_{i \in I_j^-} \neg x_i \right) \right)$$

- This MAX SAT problem is combinatorial and NP-hard!

Relaxed MAP Inference

Approaches to Relaxed Inference

- Local consistency relaxation
  - Developed in probabilistic graphical models community
  - ADVANTAGE: Many highly scalable algorithms available
  - DISADVANTAGE: No known quality guarantees for logical MRFs



- MAX SAT relaxation
  - Developed in randomized algorithms community
  - ADVANTAGE: Provides strong quality guarantees
  - DISADVANTAGE: No algorithms designed for large-scale models

- How can we combine these advantages?

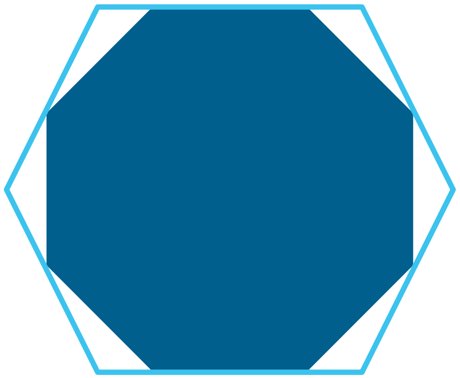
Local Consistency Relaxation

Local Consistency Relaxation

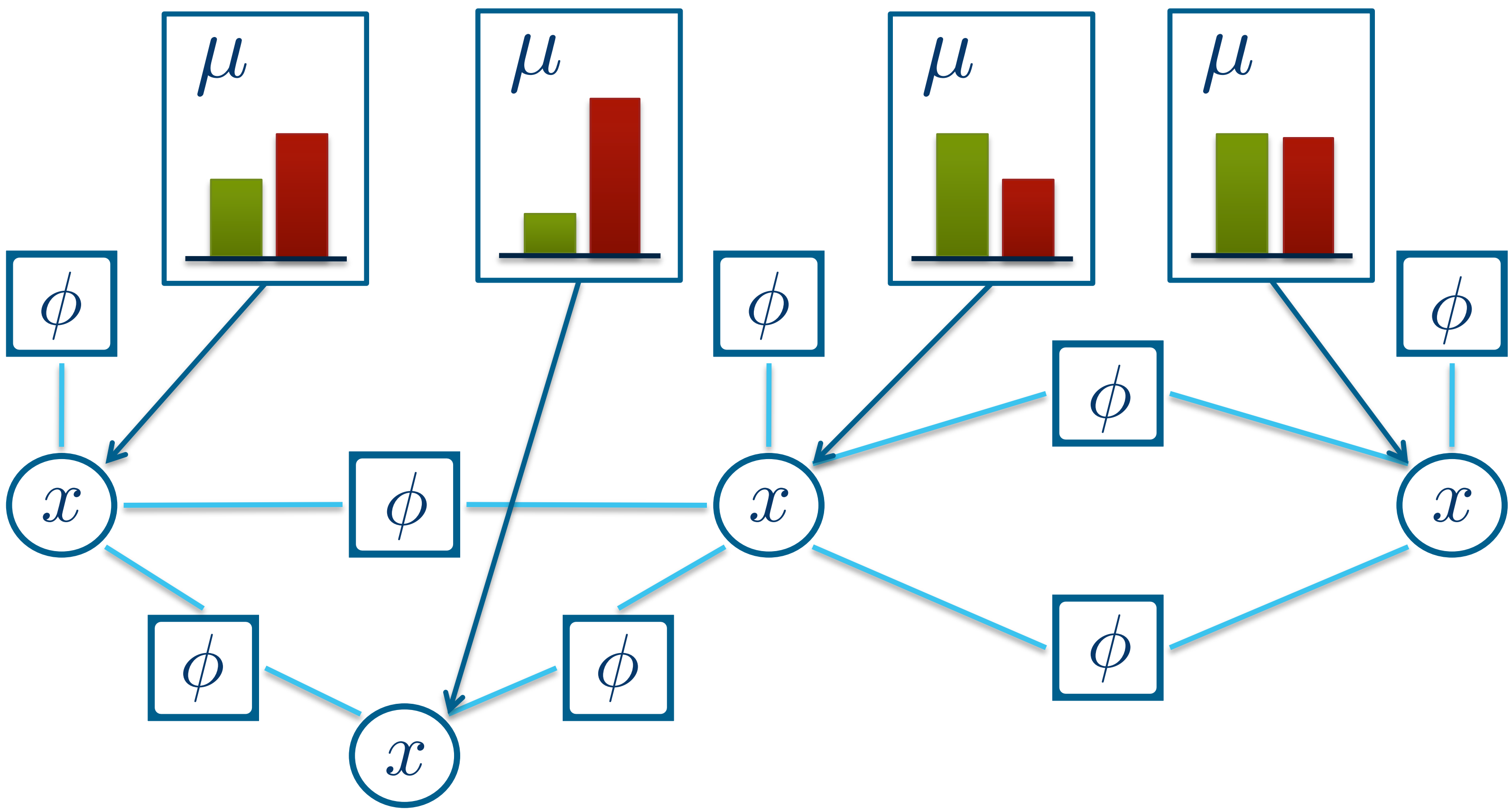
- LCR is a popular technique for approximating MAP in MRFs
  - Often simply called linear programming (LP) relaxation
  - Dual decomposition solves dual to LCR objective

- Lots of work in PGM community, e.g.,
  - Globerson and Jaakkola, 2007
  - Wainwright and Jordan, 2008
  - Sontag et al. 2008, 2012

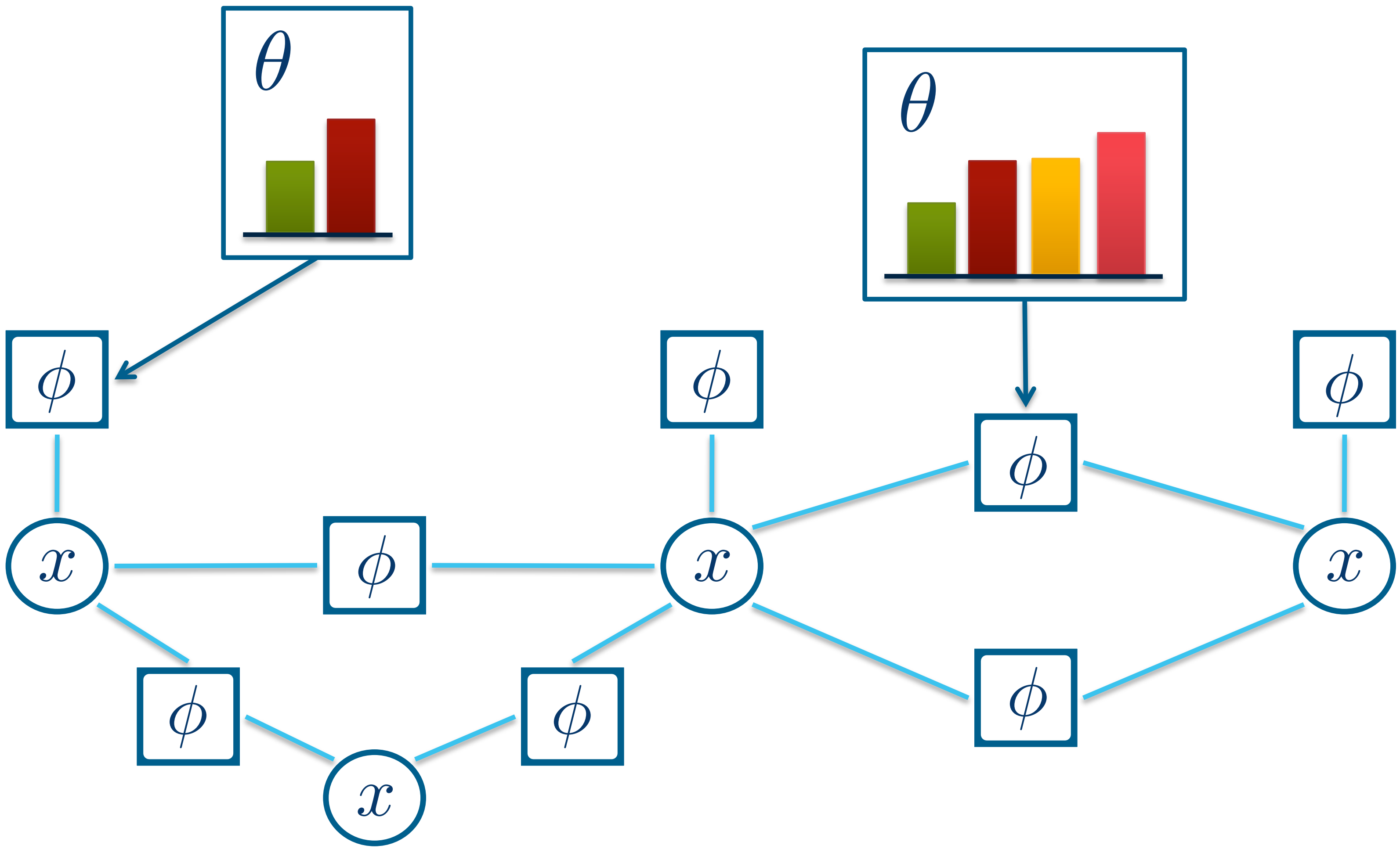
- Idea: relax search over consistent marginals to simpler set



Local Consistency Relaxation



Local Consistency Relaxation



Local Consistency Relaxation

$$\arg\max_{(\boldsymbol{\theta},\boldsymbol{\mu})\in\mathbb{L}} \sum_{R_j\in\boldsymbol{R}} w_j \sum_{\boldsymbol{x}_j} \theta_j(\boldsymbol{x}_j)\ \phi_j(\boldsymbol{x}_j)$$

$\boldsymbol{\mu}$: pseudomarginals over variable states $\boldsymbol{x}$

$\boldsymbol{\theta}$: pseudomarginals over joint potential states $\phi(\boldsymbol{x}_j)$

MAX SAT Relaxation

Approximate Inference

- View MAP inference as optimizing rounding probabilities
- Expected score of a clause is a weighted noisy-or function:

$$w_j \left(1 - \prod_{i \in I_j^+} (1 - p_i) \prod_{i \in I_j^-} p_i \right)$$

- Then expected total score is

$$\hat{W} = \sum_{R_j \in \boldsymbol{R}} w_j \left(1 - \prod_{i \in I_j^+} (1 - p_i) \prod_{i \in I_j^-} p_i \right)$$

- But, $\arg\max_{\boldsymbol{p}} \hat{W}$ is highly non-convex!

Approximate Inference

- It is the products in the objective that make it non-convex
- The expected score can be lower bounded using the relationship between arithmetic and harmonic means:

$$\frac{p_1 + p_2 + \cdots + p_k}{k} \geq \sqrt[k]{p_1 p_2 \cdots p_k}$$

- This leads to the lower bound

$$\sum_{R_j \in \boldsymbol{R}} w_j \left(1 - \prod_{i \in I_j^+} (1 - p_i) \prod_{i \in I_j^-} p_i \right) \geq \left(1 - \frac{1}{e}\right) \sum_{R_j \in \boldsymbol{R}} w_j \min \left\{ \sum_{i \in I_j^+} p_i + \sum_{i \in I_j^-} (1 - p_i), 1 \right\}$$

Goemans and Williamson, 1994

Approximate Inference

- So, we solve the linear program

$$\arg\max_{\boldsymbol{y}\in[0,1]^n} \sum_{R_j \in \boldsymbol{R}} w_j \min\left\{ \sum_{i \in I_j^+} y_i + \sum_{i \in I_j^-} (1 - y_i), 1 \right\}$$

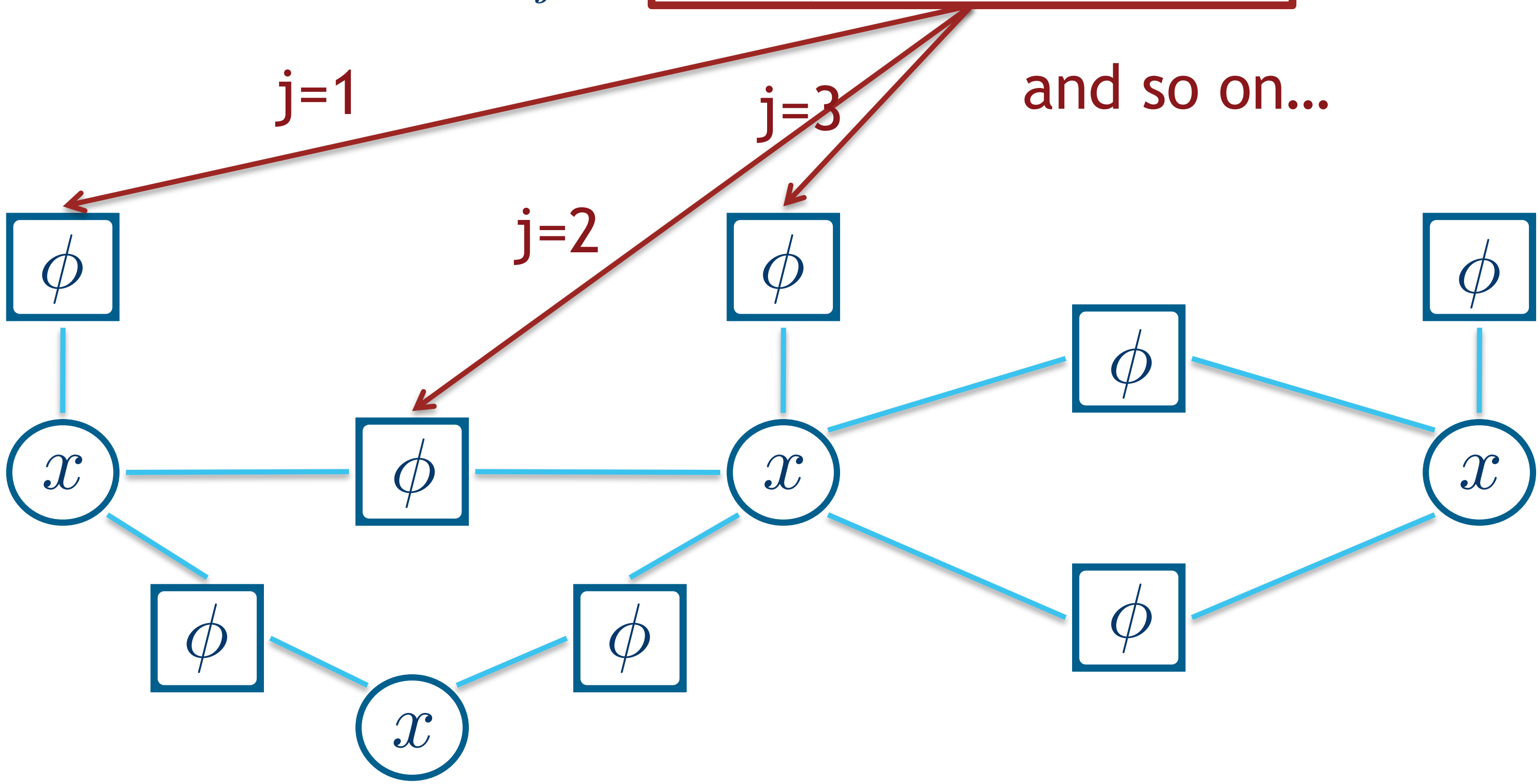
- If we set $p_i = y_i$, a greedy rounding method will find a $\left(1 - \frac{1}{e}\right)$-optimal discrete solution
- If we set $p_i = \frac{1}{2} y_i + \frac{1}{4}$, it improves to ¾-optimal

Goemans and Williamson, 1994

Unifying the Relaxations

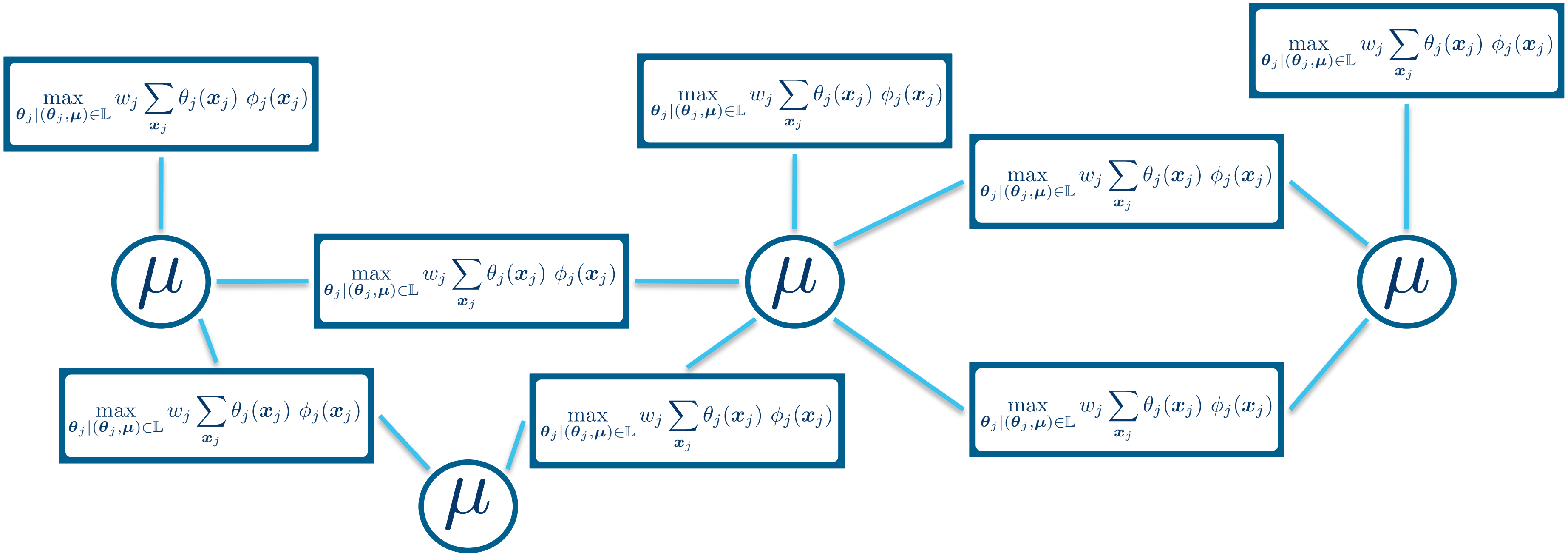
Analysis

$$\arg\max_{(\boldsymbol{\theta},\boldsymbol{\mu})\in\mathbb{L}} \sum_{R_j\in\boldsymbol{R}} w_j \sum_{\boldsymbol{x}_j} \theta_j(\boldsymbol{x}_j)\ \phi_j(\boldsymbol{x}_j)$$



Analysis

$$\arg\max_{\boldsymbol{\mu}\in[0,1]^n} \sum_{R_j\in\boldsymbol{R}} \hat{\phi}_j(\boldsymbol{\mu})$$



Analysis

- We can now analyze each potential's parameterized subproblem in isolation:

$$\hat{\phi}_j(\boldsymbol{\mu}) = \max_{\boldsymbol{\theta}_j | (\boldsymbol{\theta}_j, \boldsymbol{\mu}) \in \mathbb{L}} w_j \sum_{\boldsymbol{x}_j} \theta_j(\boldsymbol{x}_j) \; \phi_j(\boldsymbol{x}_j)$$

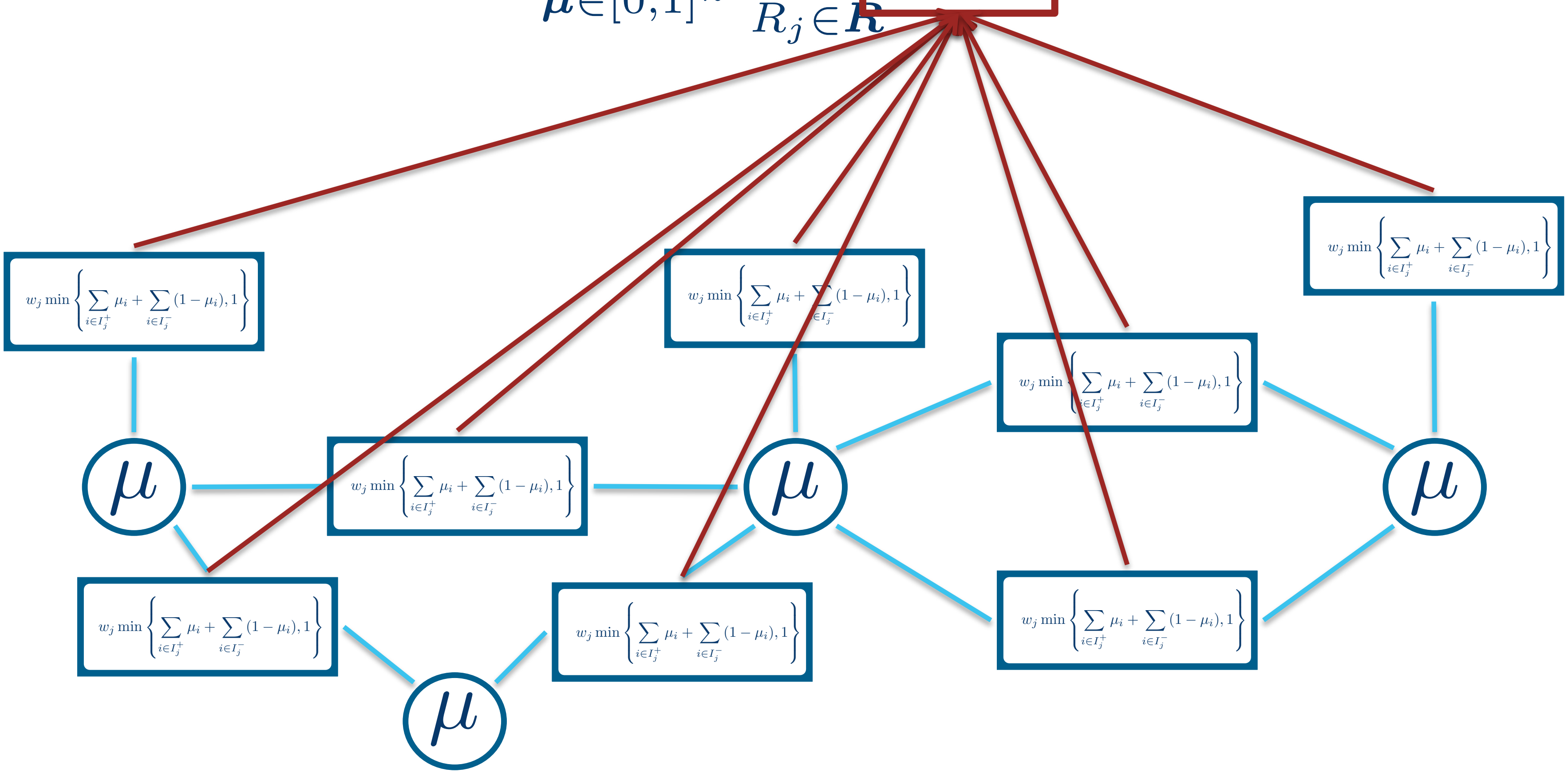
- Using the KKT conditions, we can find a simplified expression for each solution based on the parameters $\boldsymbol{\mu}$:

$$\hat{\phi}_j(\boldsymbol{\mu}) = w_j \min \left\{ \sum_{i \in I_j^+} \mu_i + \sum_{i \in I_j^-} (1 - \mu_i), 1 \right\}$$

Analysis

Substitute back into outer objective

$$\arg\max_{\boldsymbol{\mu}\in[0,1]^n} \sum_{R_j\in\boldsymbol{R}} \hat{\phi}_j(\boldsymbol{\mu})$$

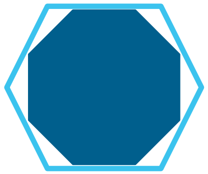


Analysis

- Leads to simplified, projected LCR over $\boldsymbol{\mu}$:

$$\arg\max_{\boldsymbol{\mu}\in[0,1]^n} \sum_{j=1}^{m} w_j \min\left\{ \sum_{i\in I_j^+} \mu_i + \sum_{i\in I_j^-} (1-\mu_i), 1 \right\}$$

Analysis



Local Consistency Relaxation

$$\arg\max_{\boldsymbol{\mu}\in[0,1]^n} \sum_{R_j \in \boldsymbol{R}} w_j \min \left\{ \sum_{i \in I_j^+} \mu_i + \sum_{i \in I_j^-} (1 - \mu_i), 1 \right\}$$

MAX SAT Relaxation

$$\arg\max_{\boldsymbol{y}\in[0,1]^n} \sum_{R_j \in \boldsymbol{R}} w_j \min \left\{ \sum_{i \in I_j^+} y_i + \sum_{i \in I_j^-} (1 - y_i), 1 \right\}$$

Evaluation

New Algorithm: Rounded LP

- Three steps:
  - Solves relaxed MAP inference problem
  - Modifies pseudomarginals
  - Rounds to discrete solutions

- We use the alternating direction method of multipliers (ADMM) to implement a message-passing approach
  [Glowinski and Marrocco, 1975; Gabay and Mercier, 1976]

- ADMM-based inference for MAX SAT form of problem was originally developed for hinge-loss MRFs
  [Bach et al., 2015]

Evaluation Setup

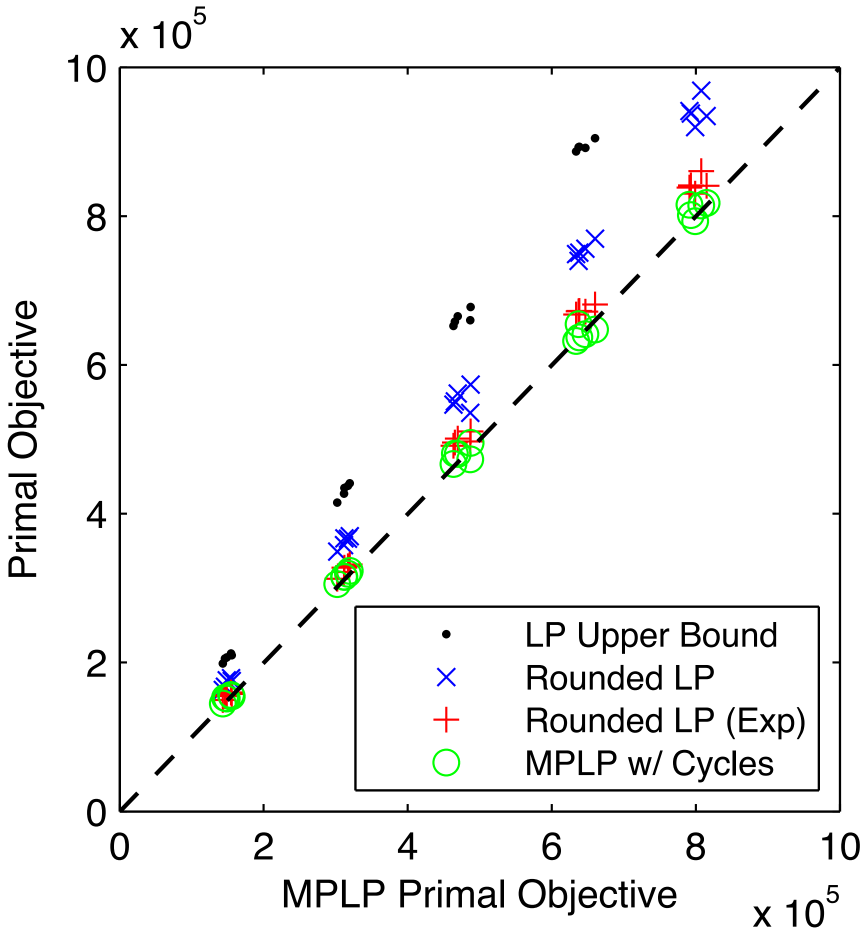
- Compared with
  - MPLP
  - MPLP with cycle tightening

[Globerson and Jaakkola, 2007; Sontag et al. 2008, 2012]

- MPLP uses coordinate descent dual decomposition, so rounding not applicable
- Solved MAP in social-network opinion models with super- and submodular features
- Measured primal score, i.e., weighted sum of satisfied clauses

Results

- Expected scores of Rounded LP are significantly better
- Rounded LP's final scores are even better
- Cycle tightening has limited effect
- Rounded LP does 20% better than MPLP, and only takes 1 minute for 1 million clauses



Conclusion

Conclusion

- Uniting local consistency and MAX SAT relaxation **combines the benefits** of both: scalability and accuracy
- Ro
  qu

Thank You!

bach@cs.umd.edu @stevebach

- Many applications to structured and relational data:
  - Social network analysis
  - Bioinformatics
  - Recommender systems
  - Text and video understanding