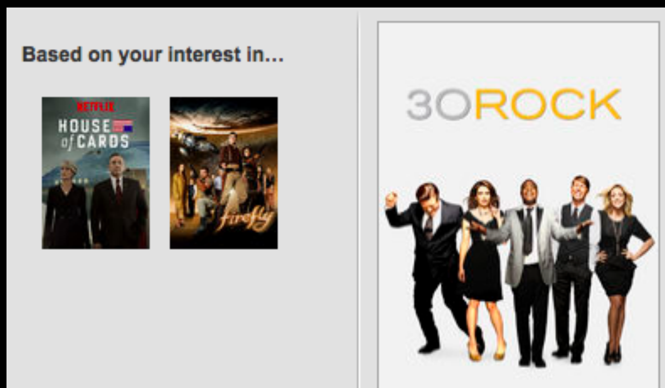
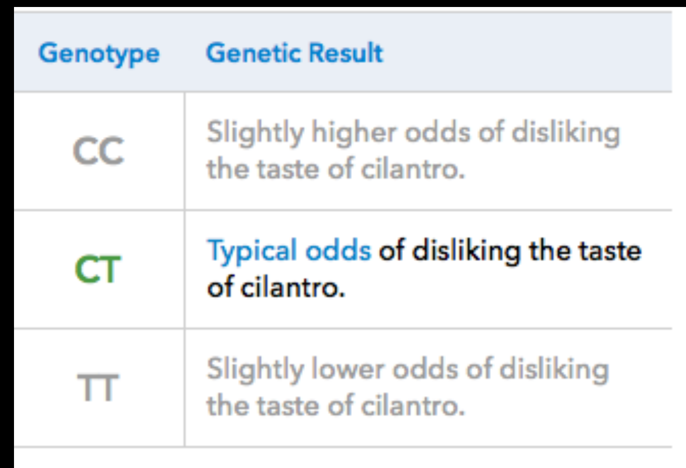
Online Collective Inference

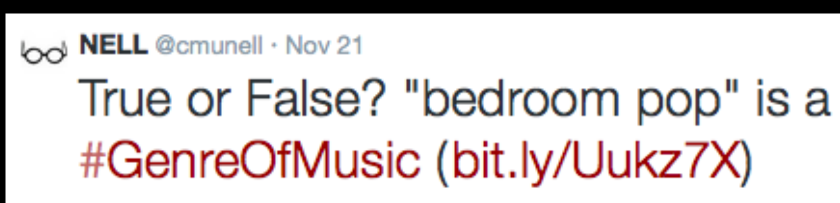
Jay Pujara — U. Maryland, College Park
Ben London — U. Maryland, College Park
Lise Getoor — U. California, Santa Cruz

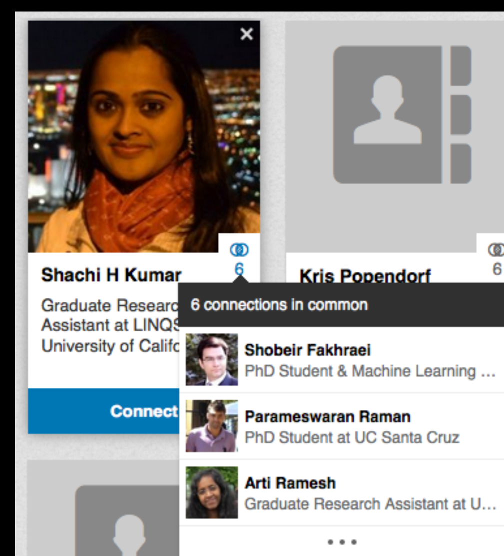
Real-world problems…



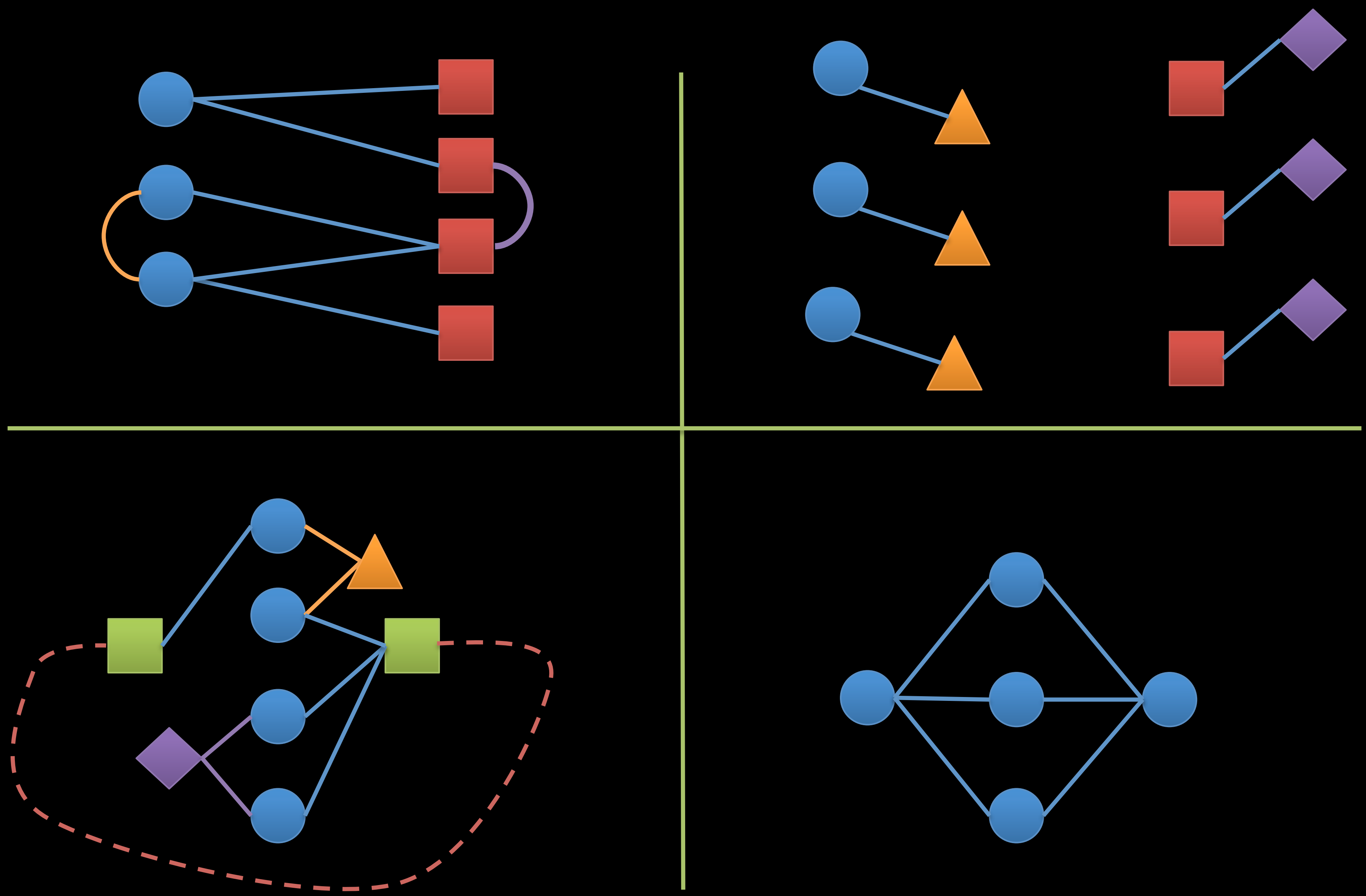


Genotype	Genetic Result
CC	Slightly higher odds of disliking the taste of cilantro.
CT	Typical odds of disliking the taste of cilantro.
TT	Slightly lower odds of disliking the taste of cilantro.

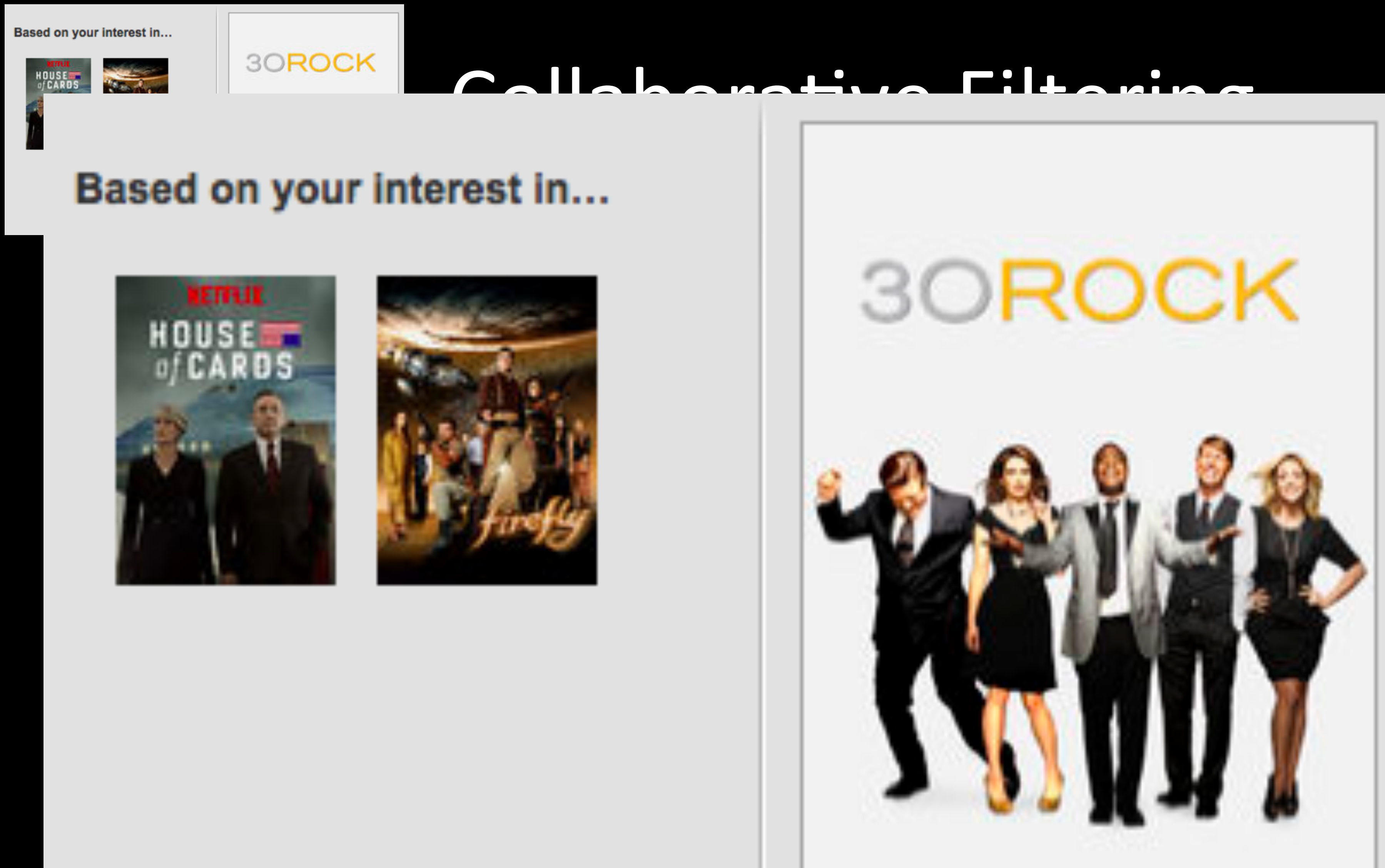


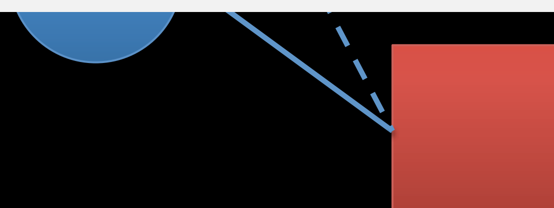


…benefit from relational models



Collaborative Filtering



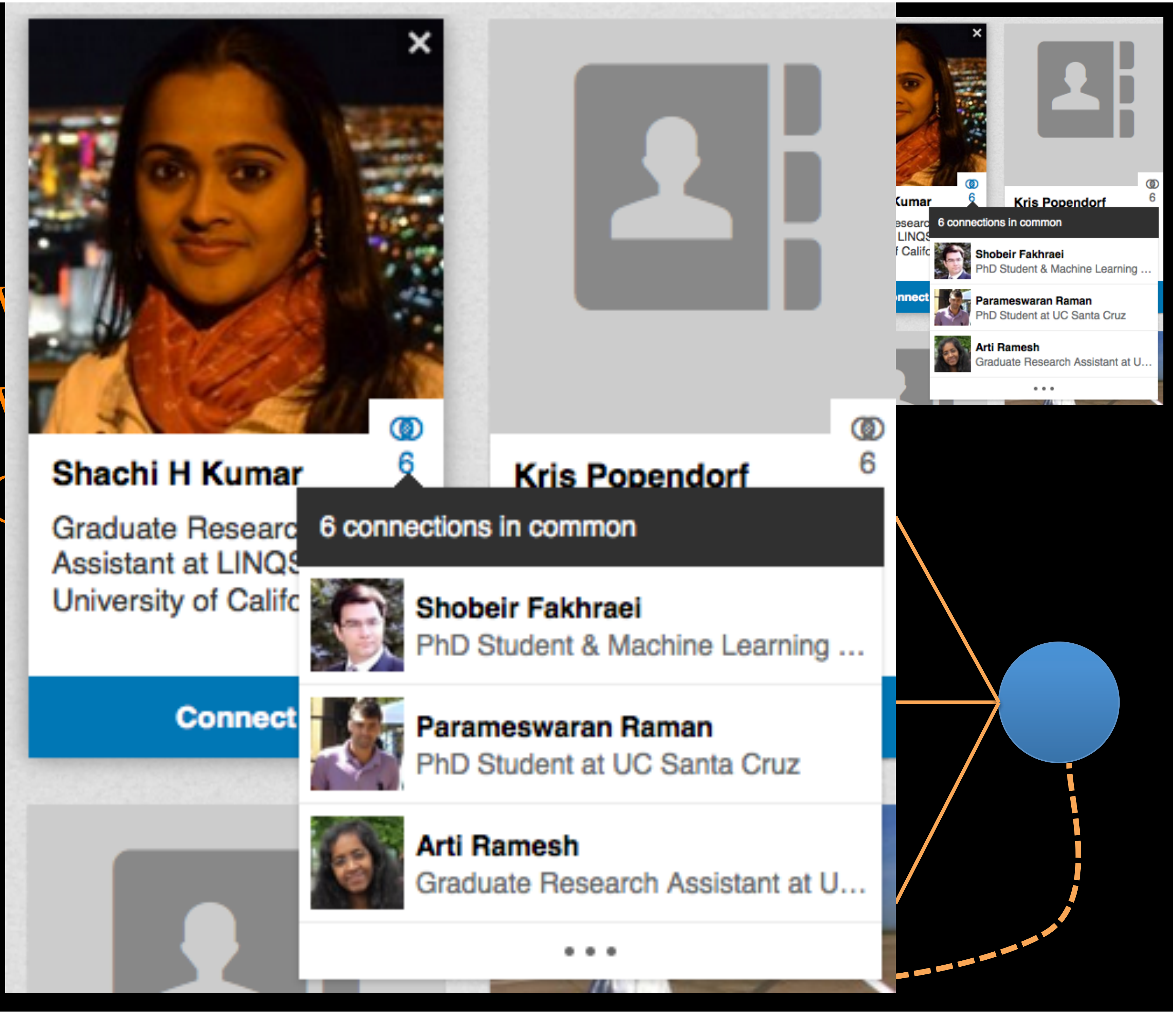


$$\rightarrow \textsc{Likes}(\mathrm{U1}, \mathrm{M2})$$

Cow

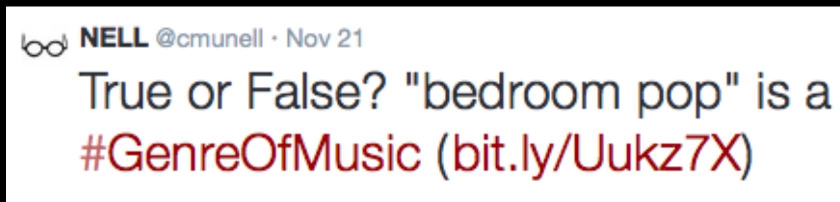
Cow

→ C



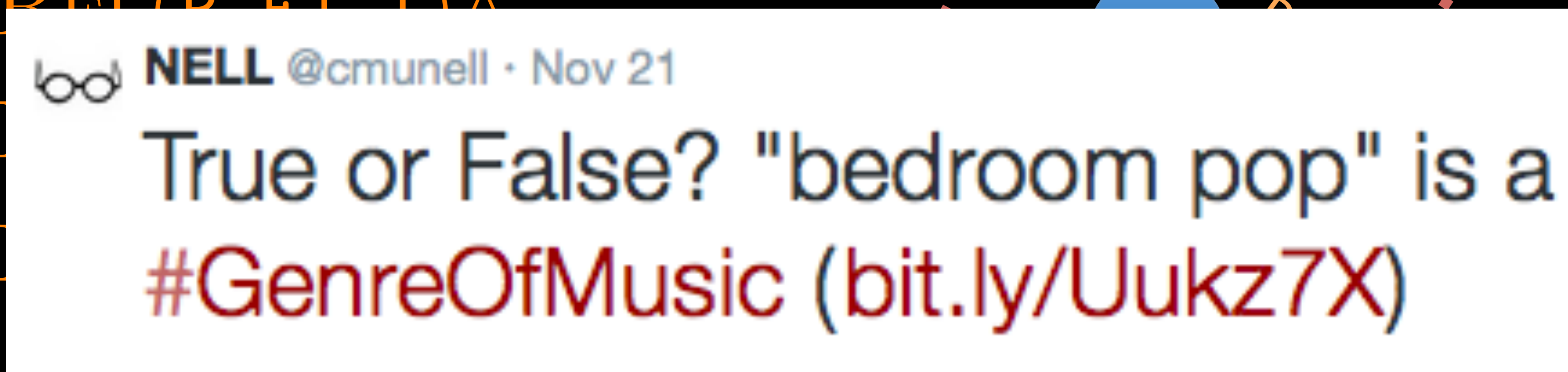
Shachi H Kumar
Graduate Researc
Assistant at LINQS
University of Califo
Connect
6
Kris Popendorf
6
6 connections in common
Shobeir Fakhraei
PhD Student & Machine Learning ...
Parameswaran Raman
PhD Student at UC Santa Cruz
Arti Ramesh
Graduate Research Assistant at U...

Knowledge Graph Identification



$\textsc{Rel}(R, E1, T) \wedge$

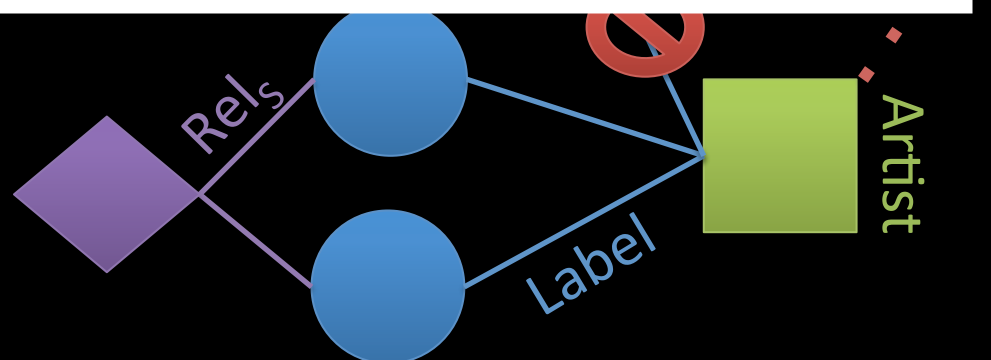
$\rightarrow \textsc{Label}(E2, L)$



$\textsc{MutEx}(L1, L2) \wedge$

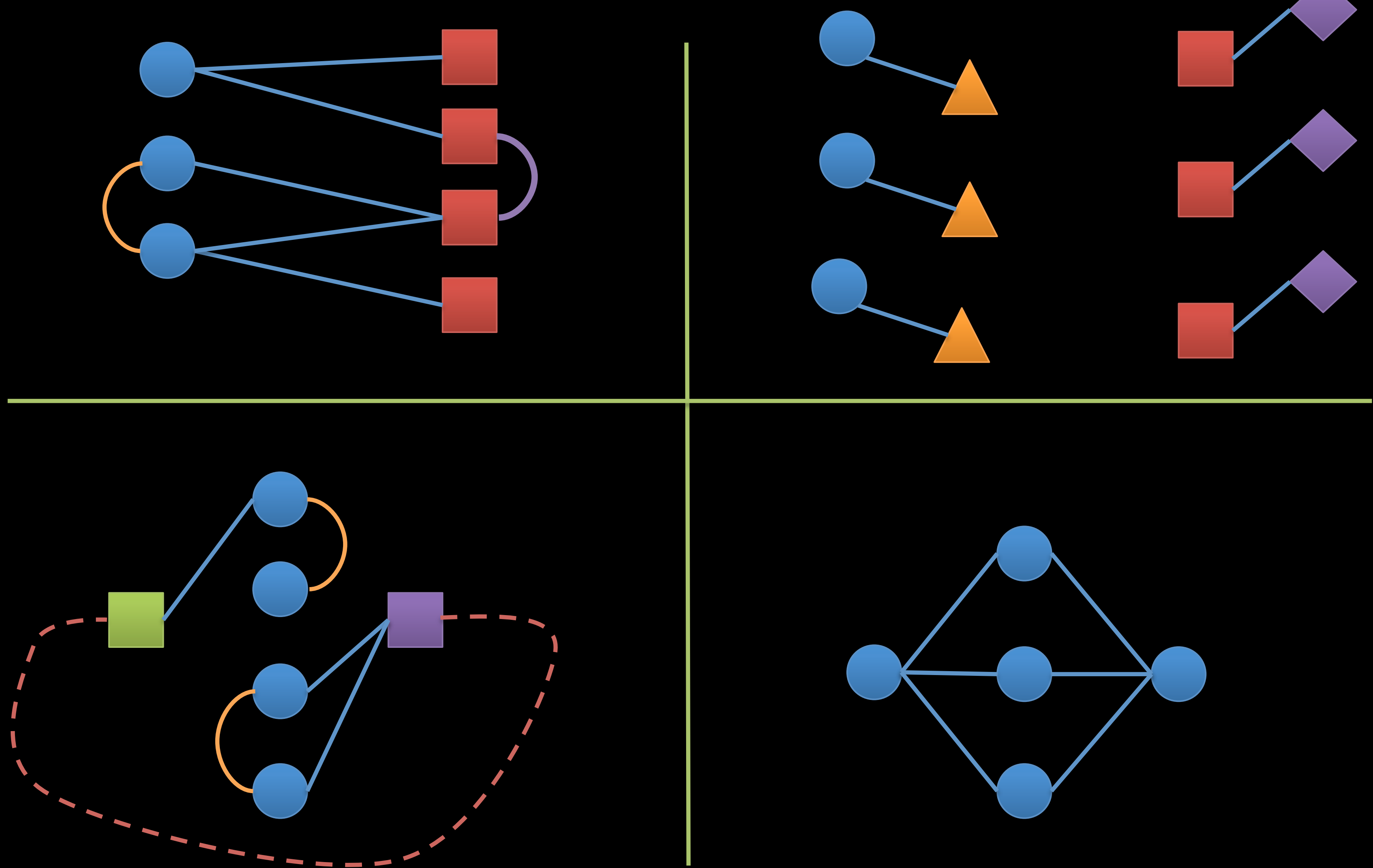
$\textsc{Label}(E, L1) \wedge$

$\rightarrow \neg\textsc{Label}(E, L2)$

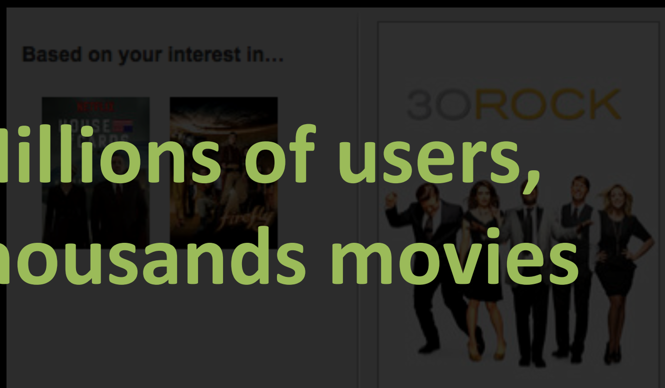


(Jiang et al., ICDM12; Pujara et al., ISWC13)

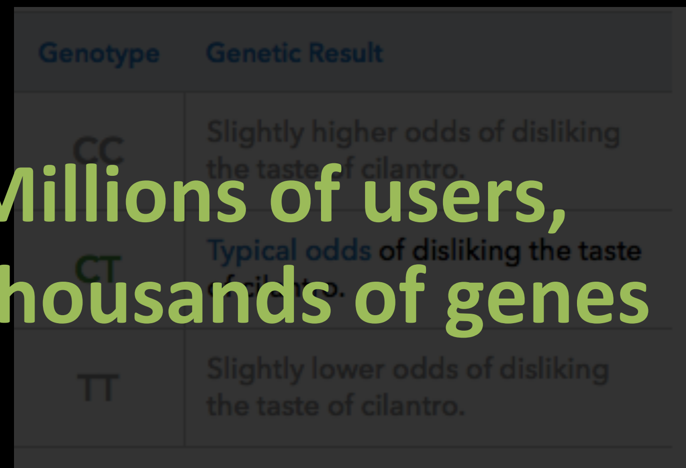
…benefit from relational models



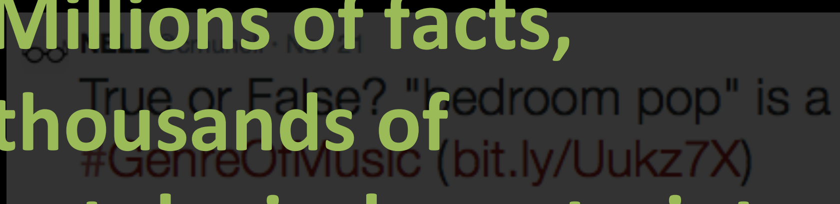
Real-world problems are big!



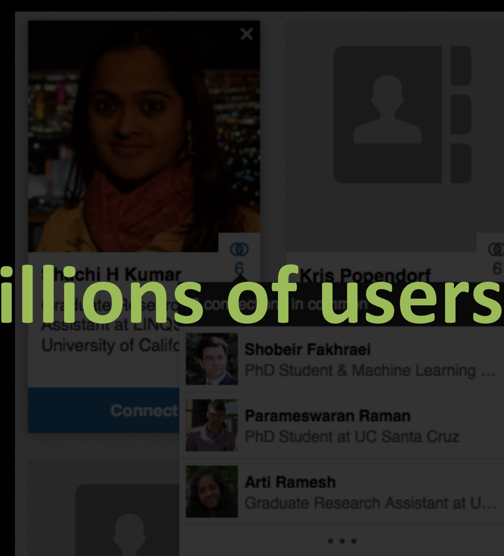
Millions of users, thousands movies



Millions of users, thousands of genes

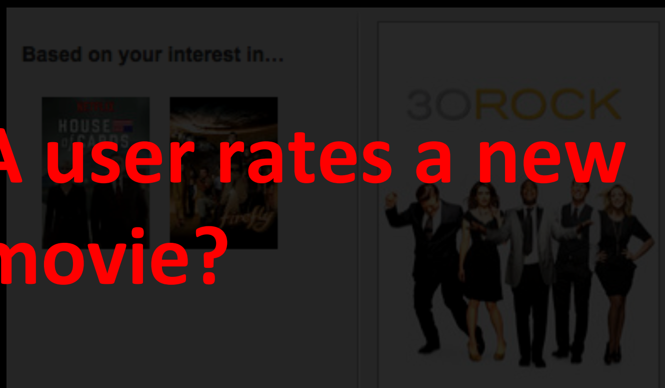


Millions of facts, thousands of ontological constraints

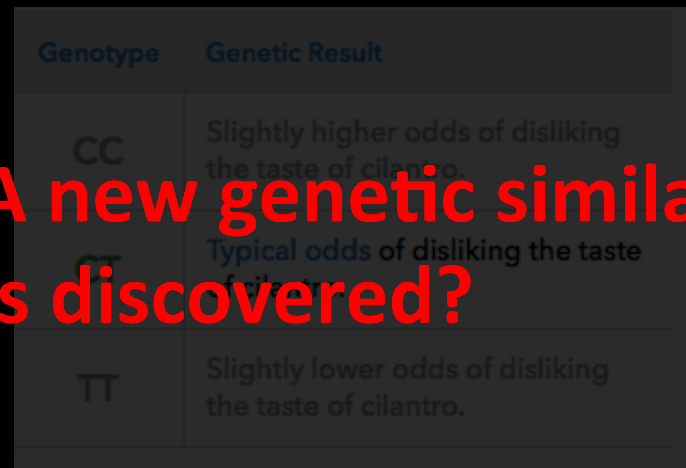


Millions of users

What happens when?

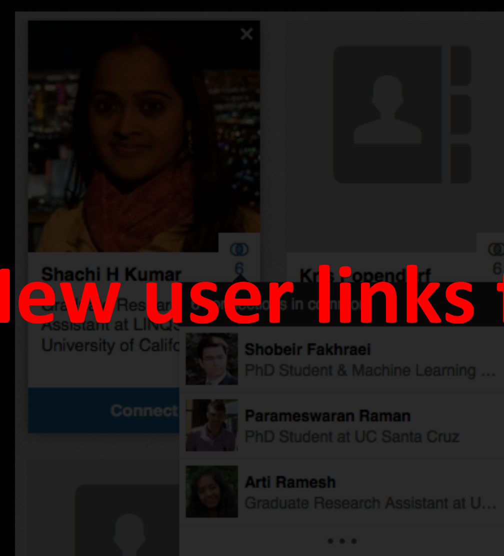


A user rates a new movie?



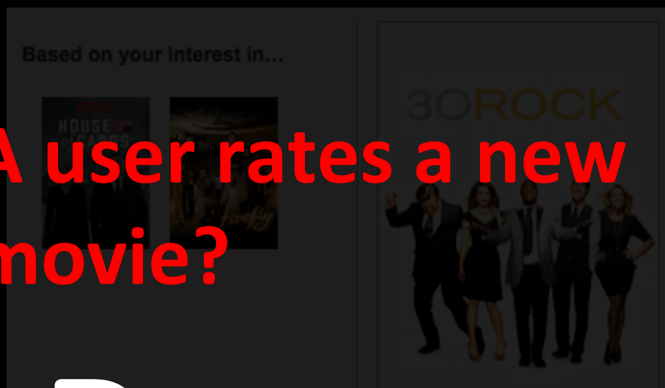
A new genetic similarity is discovered?

New facts are extracted from the Web?

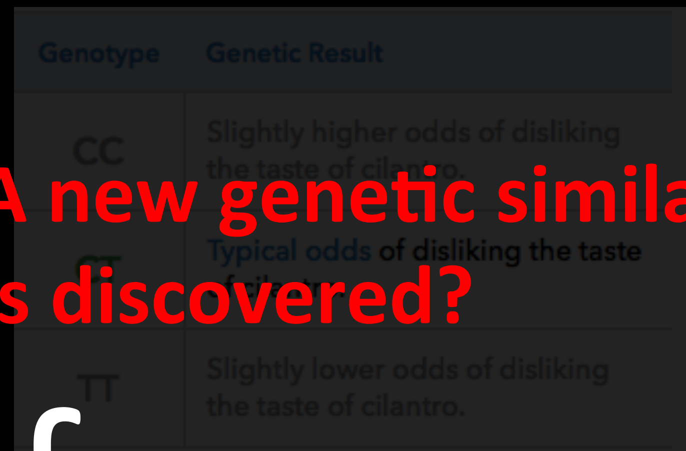


New user links form?

What happens when?



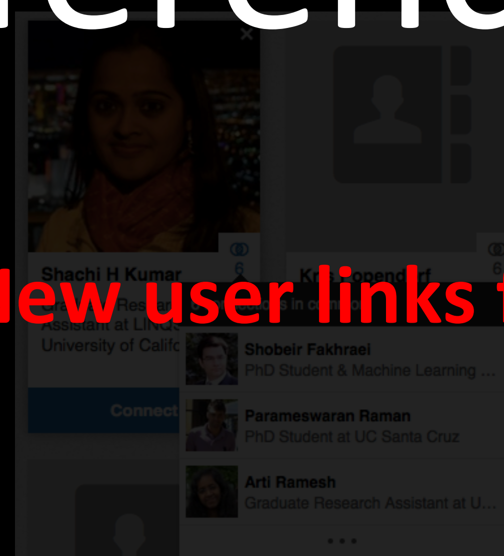
A user rates a new movie?



A new genetic similarity is discovered?

Repeat Inference!

New facts are extracted from the Web?



New user links form?

Why can't we repeat inference?

- We want rich, collective models!
- But, 10M-1B factors = 1-100s hours*
- Ideal: Inference time balances update cycle
- Insanity is doing the same thing over and over…

Online Collective Inference

PROBLEM SETTING

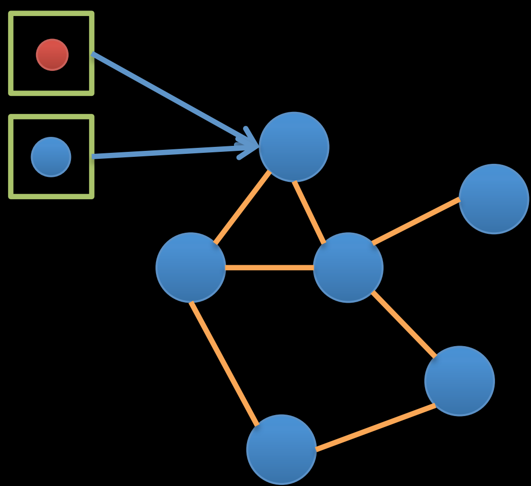
Key Problem

- Real-world problems -> large graphical models
- Changing evidence -> repeat inference

Key Problem

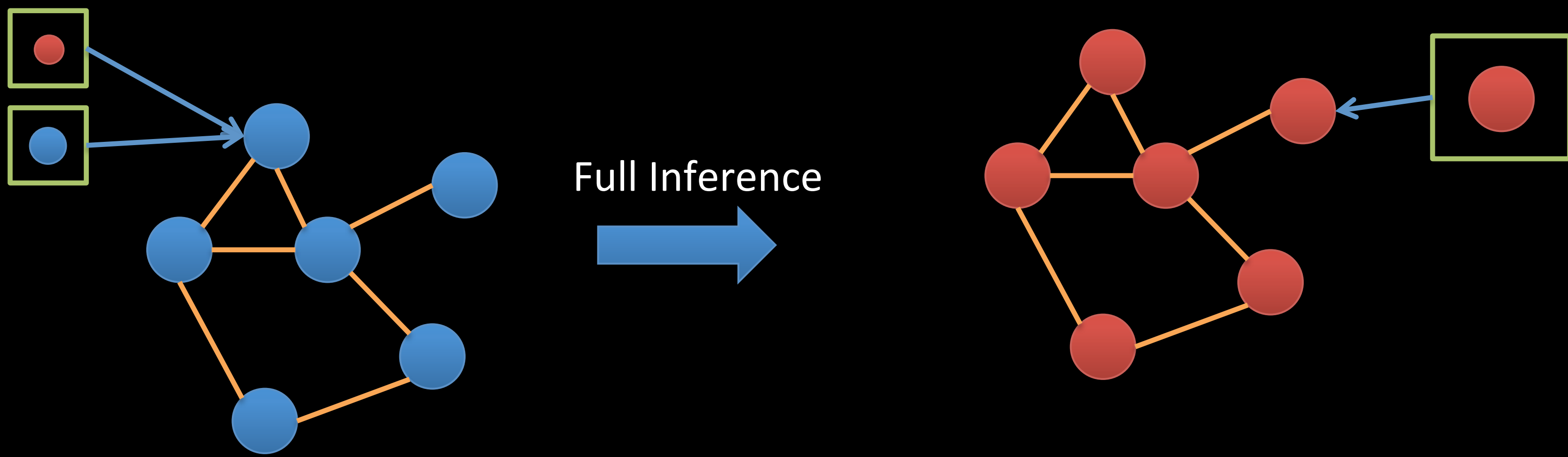
- Real-world problems -> large graphical models
- Changing evidence -> ~~repeat inference~~
- What happens when partially updating inference?
- Can we scalably approximate the MAP state without recomputing inference?

Generic Answer: NO!



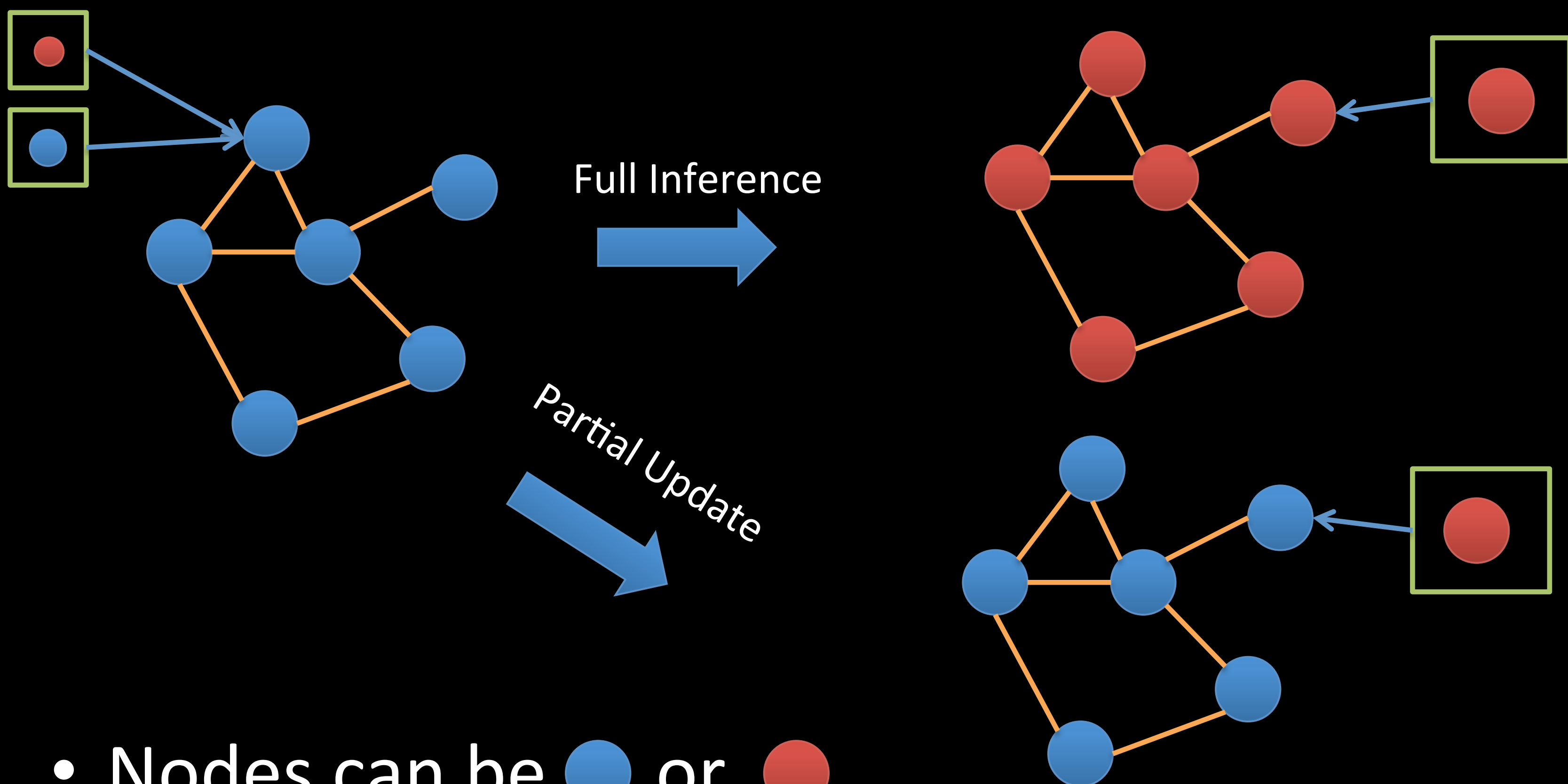
- Nodes can be or
- Model has prob. mass only when nodes same
- Fix some nodes to then observe evidence for

Generic Answer: NO!



- Nodes can be or
- Model has prob. mass only when nodes same
- Fix some nodes to then observe evidence for

Generic Answer: NO!



- Nodes can be or
- Model has prob. mass only when nodes same
- Fix some nodes to then observe evidence for

Previous Work

- Belief Revision
  - e.g. Gardenfors, 1992
- Bayesian Network Updates
  - e.g. Buntine, 1991; Friedman & Goldszmidt, 1997
- Dynamic / Sequential Models
  - e.g. Murphy, 2002 / Fine et al., 1998
- Adaptive Inference
  - e.g. Acar et al., 2008
- BP Message Passing
  - e.g. Nath & Domingos, 2010
- Collective Stability
  - e.g. London et al., 2013

Problem Setting

- Fixed model: dependencies & weights known
- Online: changing evidence or observations
- Closed world: all variables identified
- Budget: infer only *m* variables in each epoch
- **Strongly-convex inference objective** (e.g. PSL)

Questions:

- What guarantees can we offer?
- Which *m* variables should we infer?

Approach

- Define "regret" for online collective inference
- Introduce regret bounds for strongly convex inference objectives (like PSL!)
- Develop algorithms to *activate* a subset of the variables during inference, given budget

Online Collective Inference

REGRET BOUNDS

Inference Regret

- General inference problem: estimate $P(Y|X)$

- In online collective inference: fix Y_S, infer $Y_{\overline{S}}$

- Regret (learning): captures distance to optimal

- Regret (inference): the distance between the full inference result and the partial inference update (when conditioning on Y_S)

Defining Regret

- Regret: distance between full & approximate inference

$$\mathfrak{R}_n(\mathbf{x}, \mathbf{y}_{\mathcal{S}}; \dot{\mathbf{w}}) \triangleq \frac{1}{n} \left\| h(\mathbf{x}; \dot{\mathbf{w}}) - h(\mathbf{x}, \mathbf{y}_{\mathcal{S}}; \dot{\mathbf{w}}) \right\|_1$$

where

Prior weight

$$h(\mathbf{x}; \dot{\mathbf{w}}) = \arg\min_{\mathbf{y}} \ \mathbf{w} \cdot \mathbf{f}(\mathbf{x}, \mathbf{y}) + \frac{w_p}{2} \|\mathbf{y}\|_2^2 \, .$$

Regret Bound

$$\mathfrak{R}_n(\mathbf{x}, \mathbf{y}_{\mathcal{S}}; \dot{\mathbf{w}}) \leq O\left(\sqrt{\frac{B\|\mathbf{w}\|_2}{n \cdot w_p}\|\mathbf{y}_{\mathcal{S}} - \hat{\mathbf{y}}_{\mathcal{S}}\|_1}\right)$$

Regret Ingredients:

Lipschitz constant

2-norm of model weights

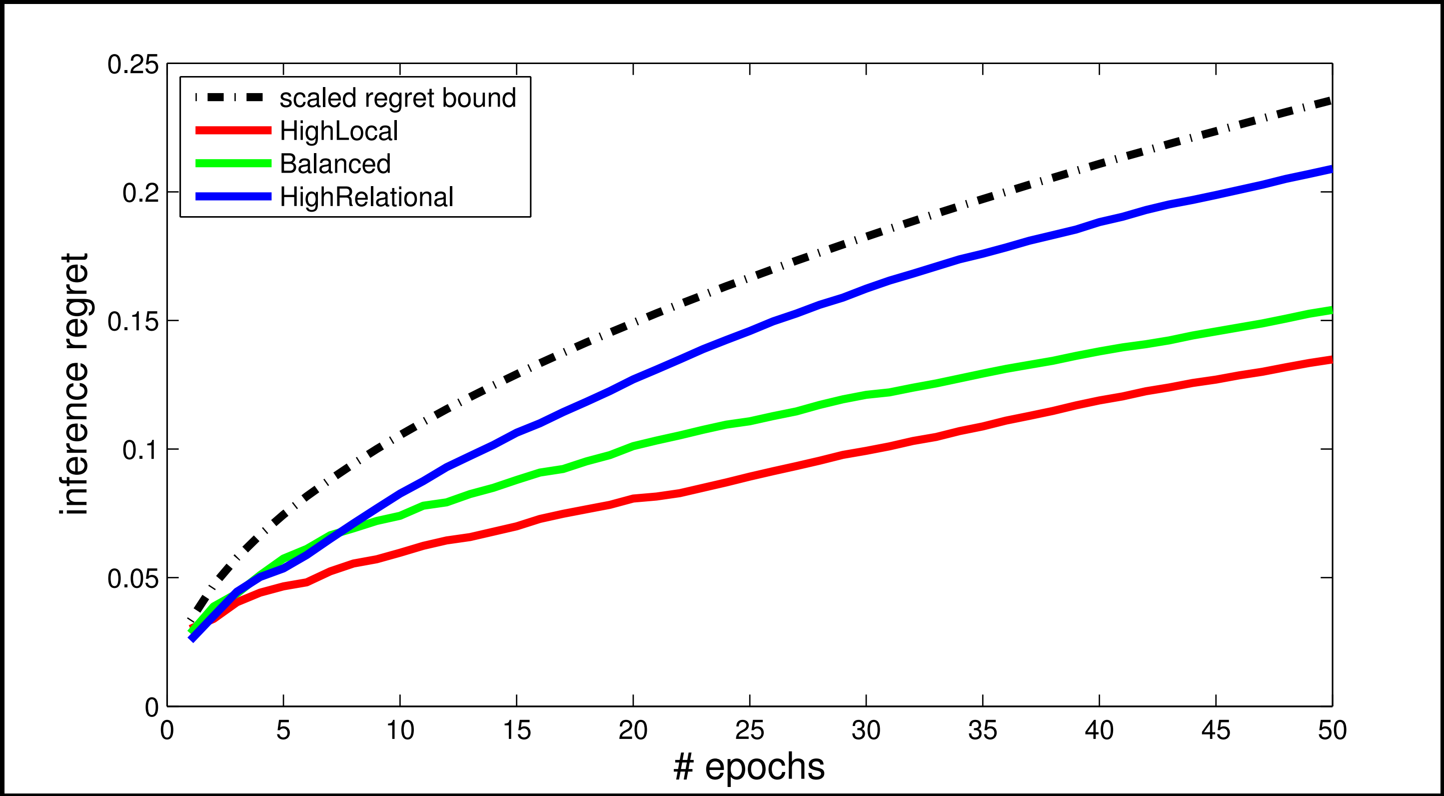
Weight of L_2 prior

L_1 distance fixed variables and values in full inference

Key Takeaway:

Regret depends on L_1 distance between fixed variables & their “true” values in the MAP state

Validating Regret Bounds



Measure regret of no updates versus full inference, varying the importance of relational features

Online Collective Inference

ACTIVATION ALGORITHMS

Which variables to fix?

- Knapsack: combinatorial, regrets/costs, budget

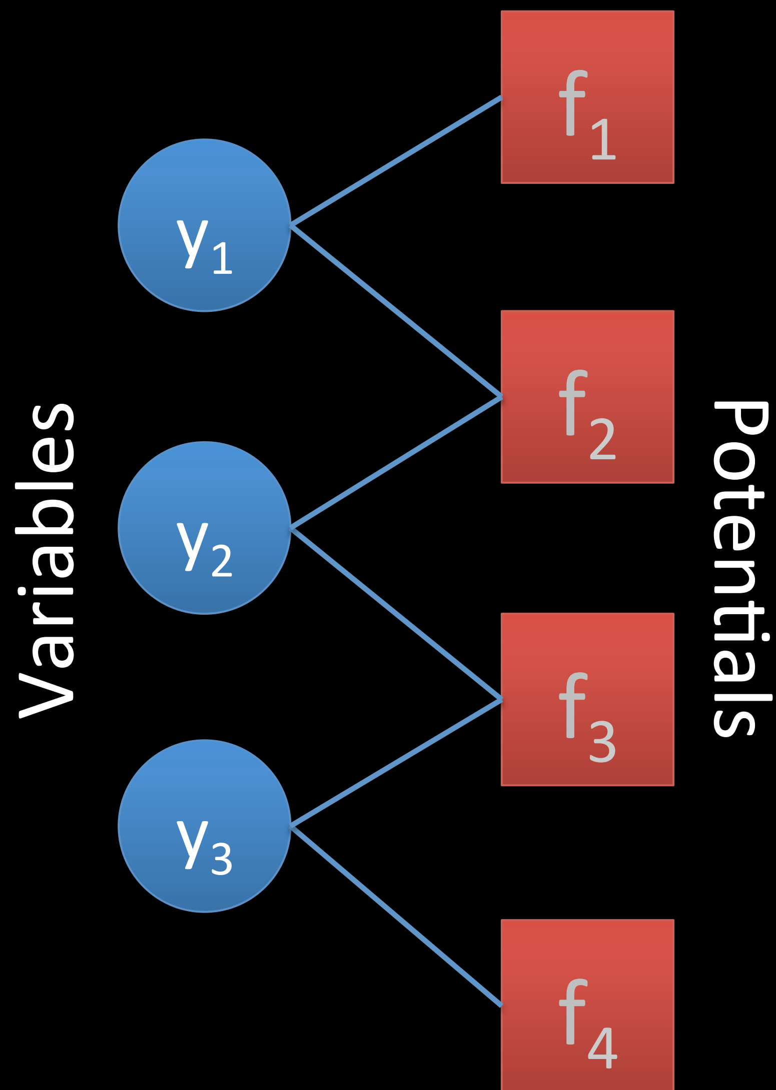
- Theory: fix variables that won't change
- Practice: how can we know what will change?

- Idea: Can we use features of past inferences?

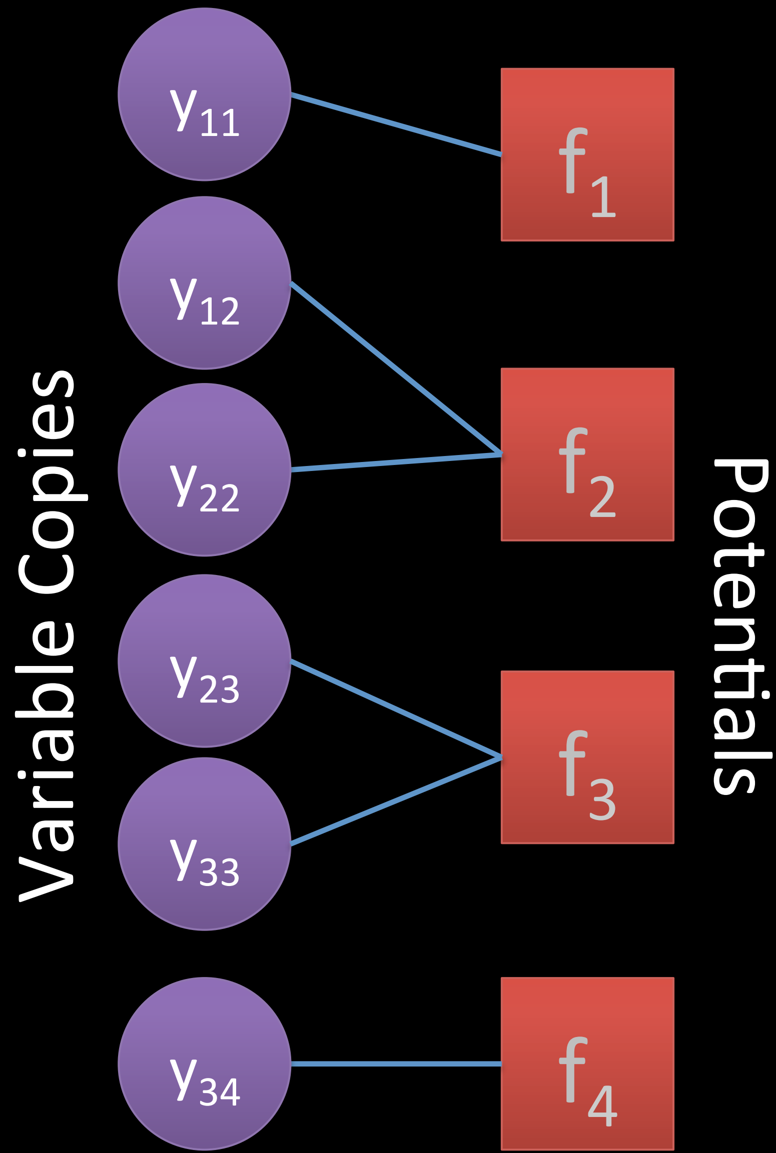
- Explore optimization (case study ADMM & PSL)

ADMM Inference in PSL

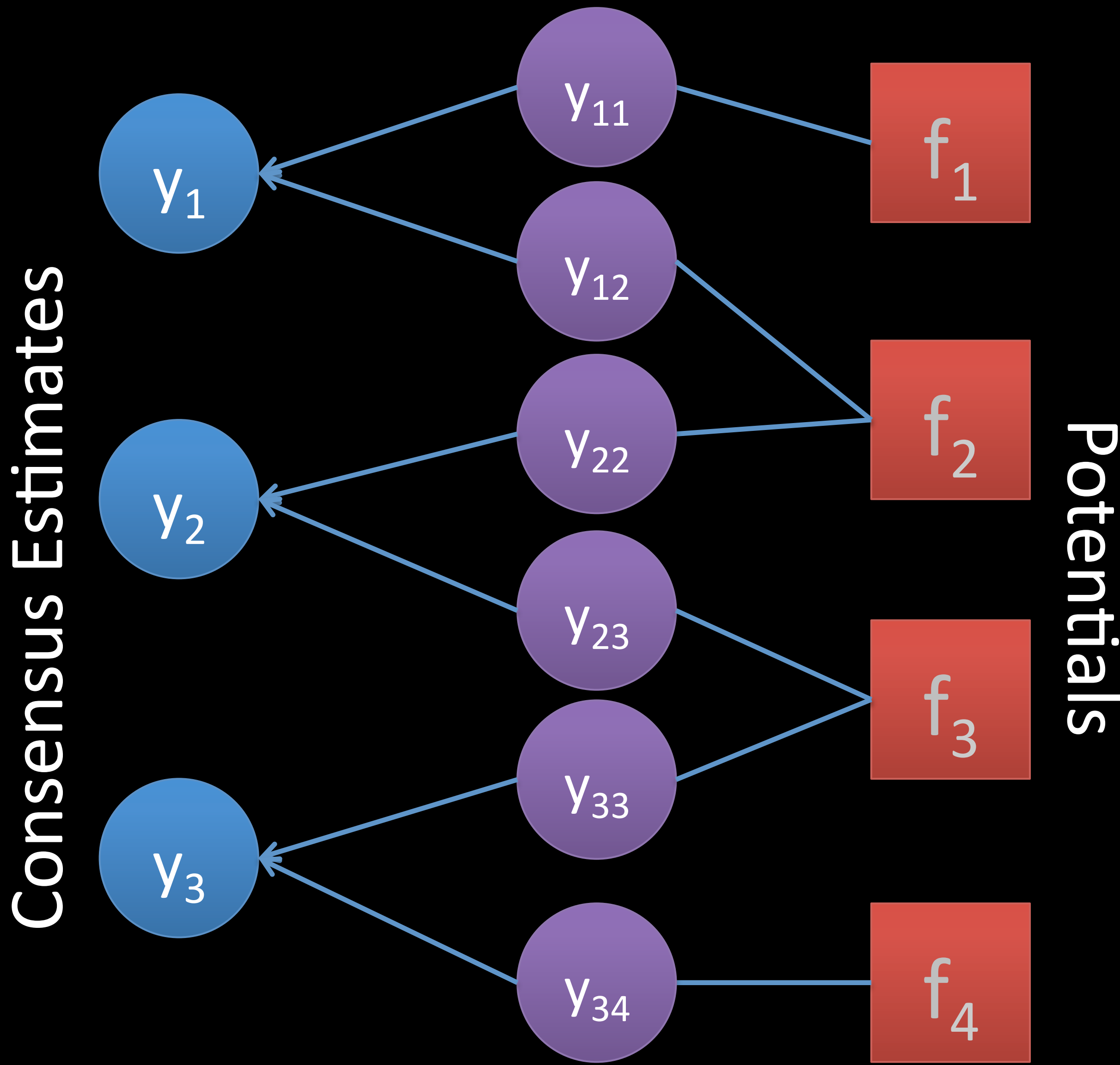
(Boyd et al., 2011; Bach et al. 2012)



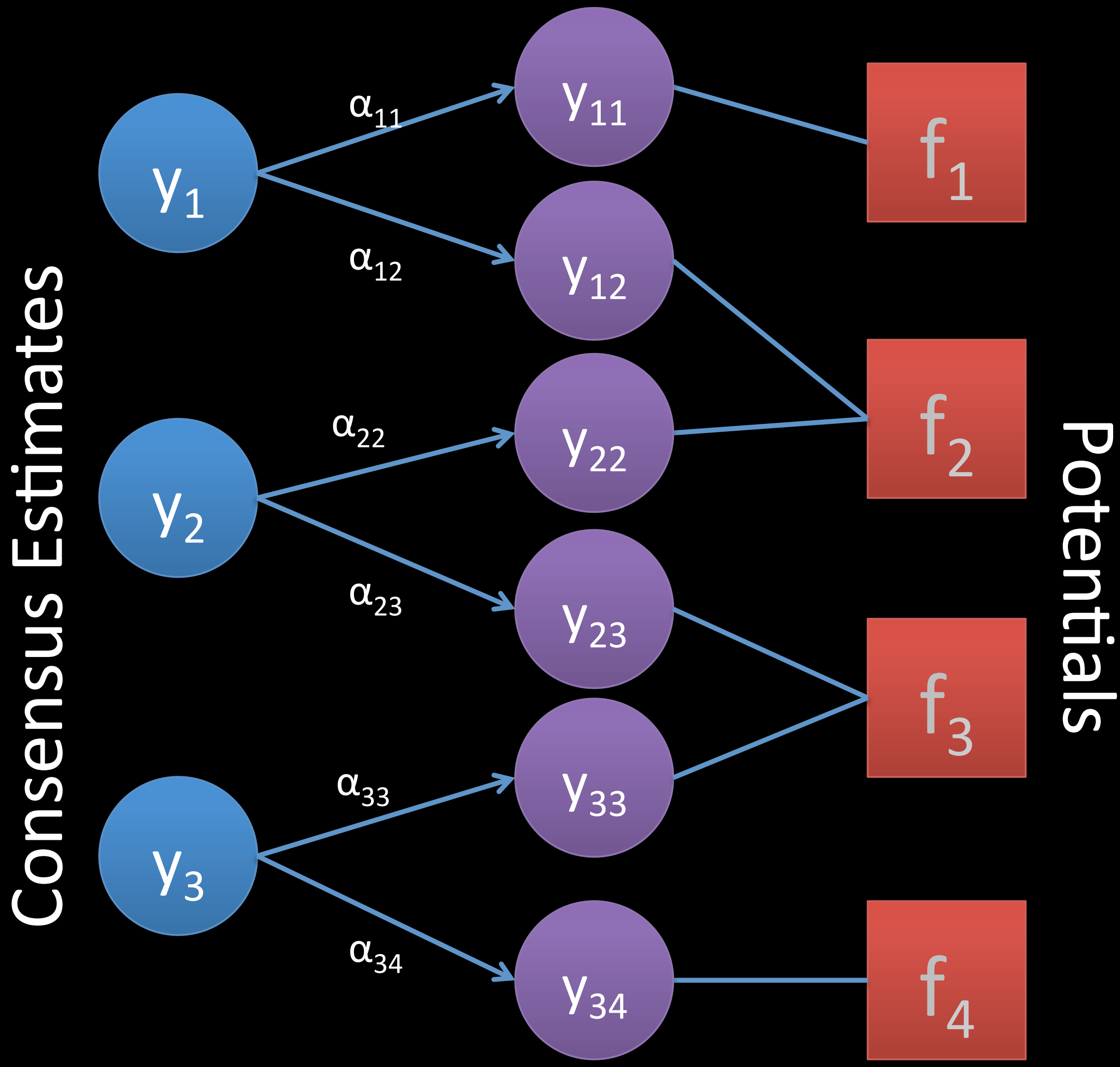
ADMM Inference in PSL



ADMM Inference in PSL



ADMM Inference in PSL



ADMM Features

$$\min_{\tilde{\mathbf{y}}_g} \; w_g \, f_g(\mathbf{x}, \tilde{\mathbf{y}}_g) + \frac{\rho}{2} \left\| \tilde{\mathbf{y}}_g - \mathbf{y}_g + \frac{1}{\rho} \boldsymbol{\alpha}_g{}^2 \right\|_2$$

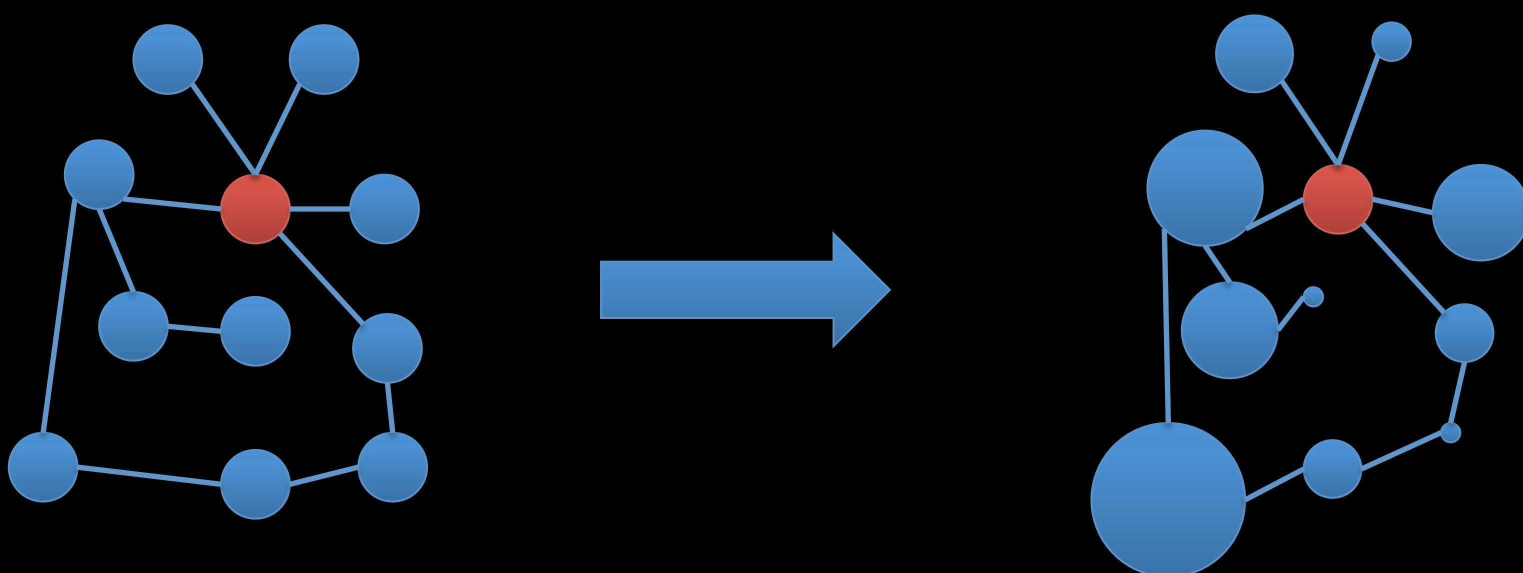
- Weight: how important is the potential?
- Potential: what loss do we incur?
- Consensus: what is the variable's value?
- Lagrange Multiplier: how much disagreement is there across potentials?

Two heuristics for activation

- Truth-Value: Variable value near 0.5
- Weighted Lagrangian: rule weight x Lagrange multipliers high

Using Model Structure

- Variable dependencies matter!
- Perform BFS, starting with new evidence
- Use heuristics + decay to prioritize exploration



EXPERIMENTAL EVALUATION

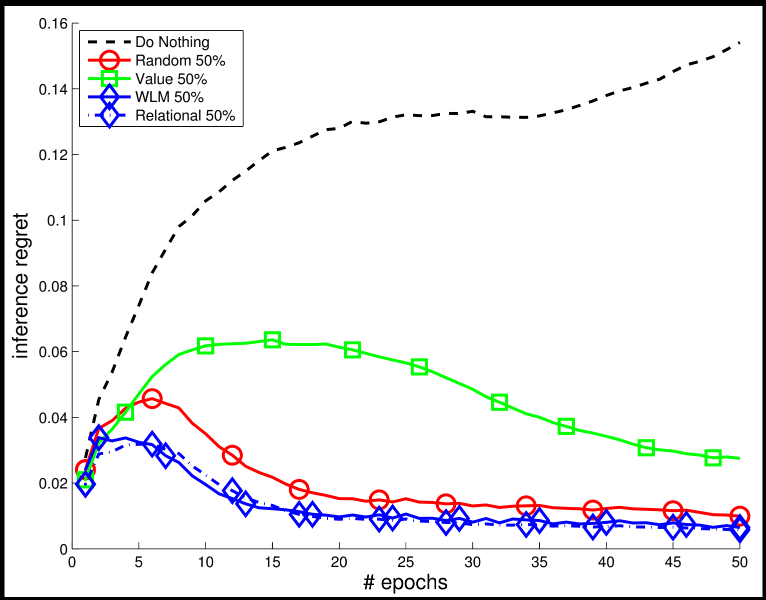
Two Online Inference Tasks

- Collective Classification (Synthetic)
  - Infer attributes of users in a social network as progressively more information is shared
- Collaborative Filtering (Jester; Goldberg et al. 2001)
  - Infer user ratings of jokes as users provide ratings for an increasing number of jokes

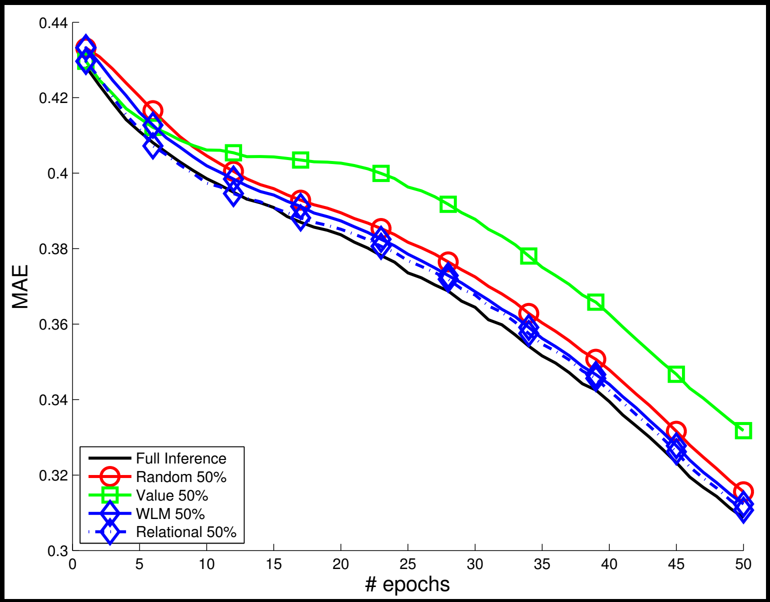
Two Online Inference Tasks

- Collective Classification (Synthetic)
  - 100 total trials (10 networks x 10 series)
  - Network evolves from 10% to 60% observed
  - Fix 50% of variables at each epoch
- Collaborative Filtering (Jester)
  - 10 trials, 100 users, 100 jokes
  - Evolve from 25% to 75% revealed ratings
  - Fix {25,50,75}% of variables at each epoch

Collective Classification: Approximate Inference



regret vs. epochs

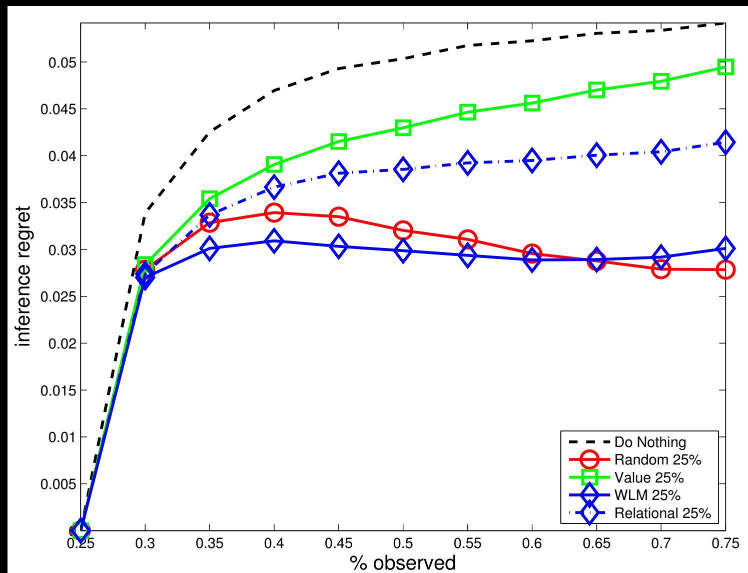


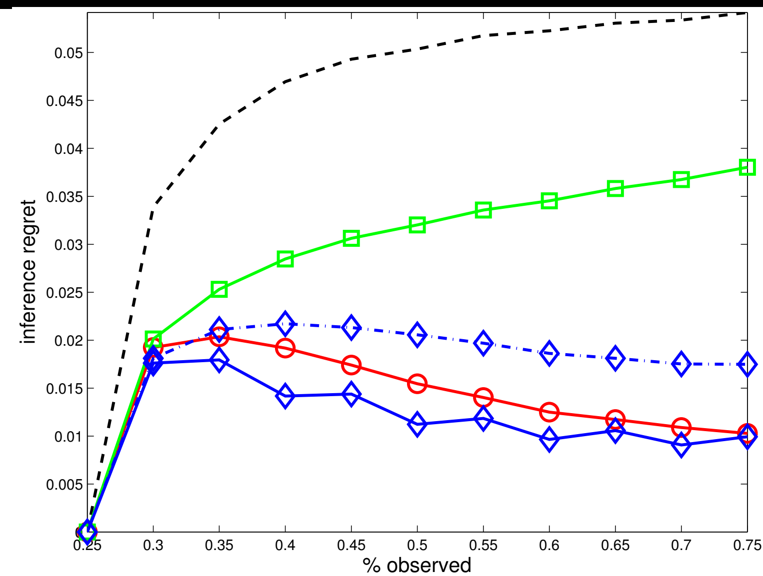
error vs. epochs

- Regret diminishes over time
- Error decreases, approaching full inference
- 69% reduction in inference time

Collaborative Filtering

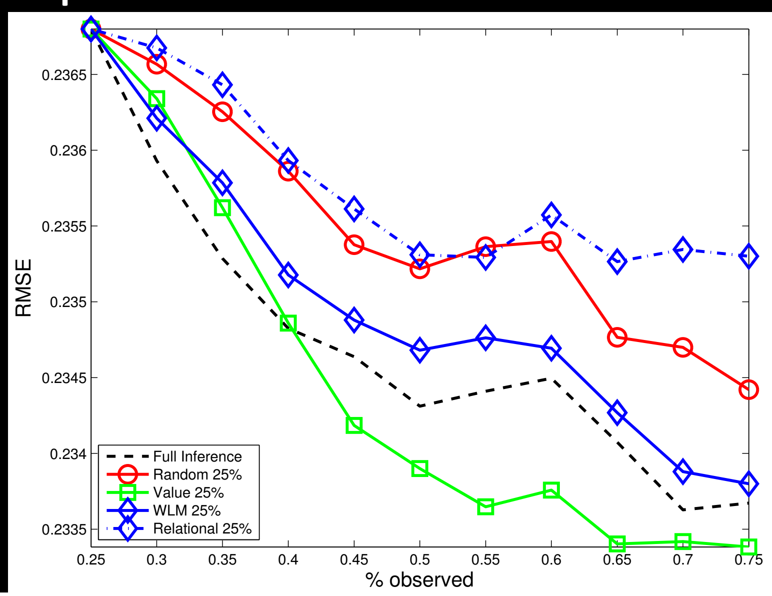
Regret

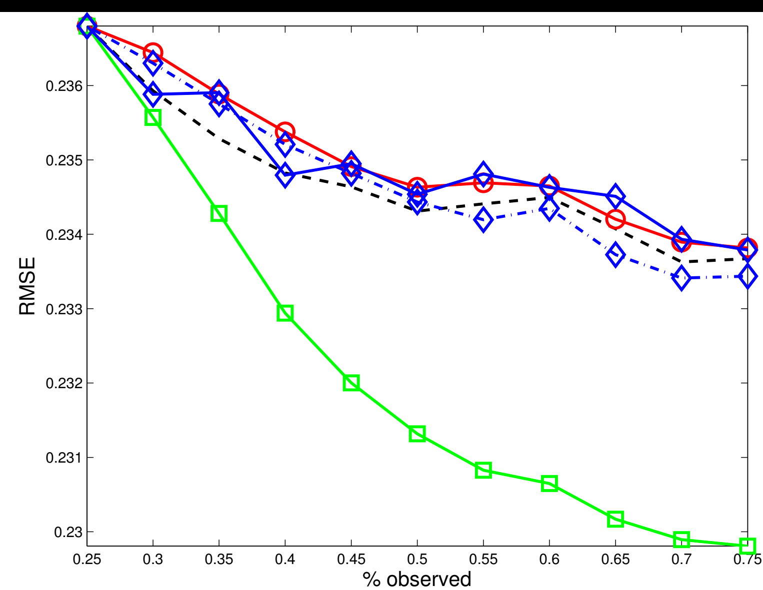




Epochs 25% activated 50% activated

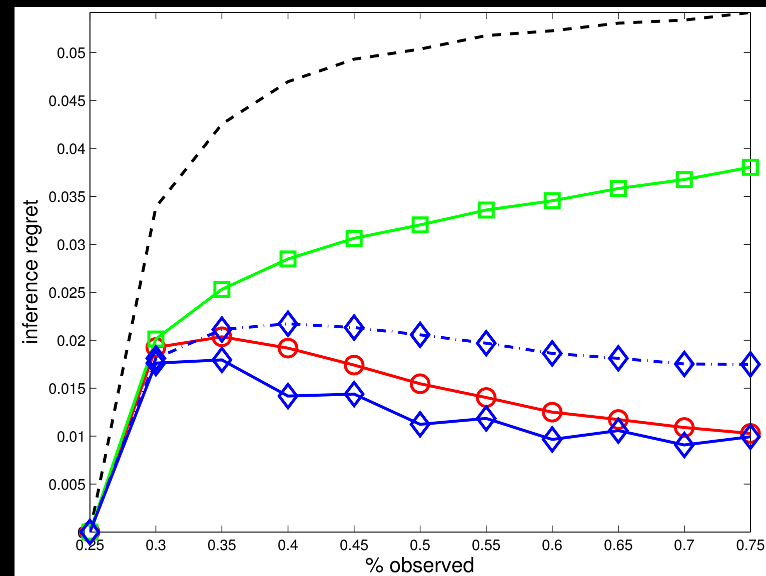
RMSE





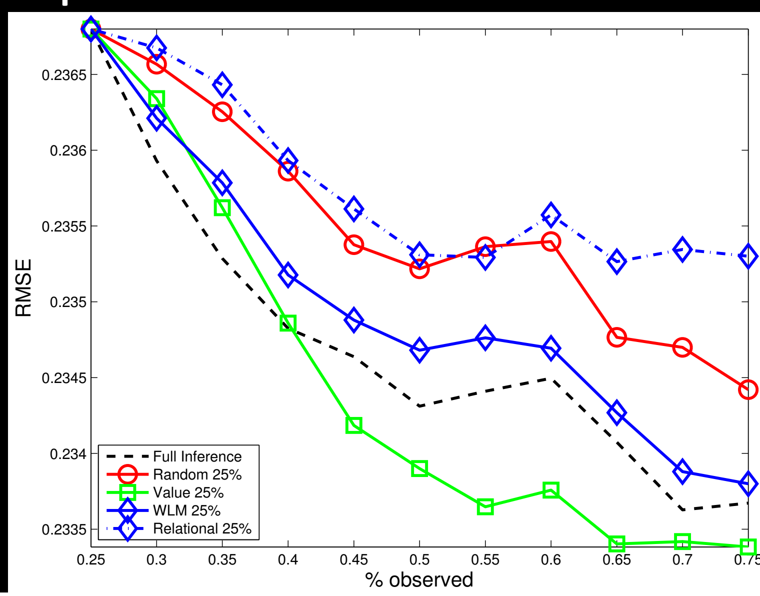
Collaborative Filtering

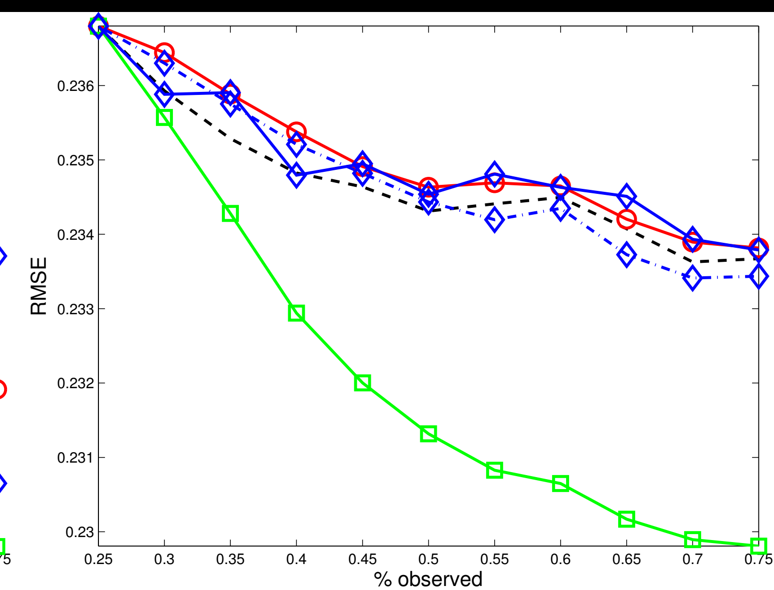
- Value: high regret, but lower error than full inference
  - Preserves polarized ratings
- 66% reduction in time for approximate inference



Epochs 25% activated 50% activated

RMSE





Online Collective Inference

CONCLUSION

Summary

- Extremely relevant to modern problems
- Necessity: approximate MAP state in PGMs
- Inference regret: bound approximation error
- Approx. algos: use optimization features
- Results: low regret, low error, faster
- New possibilities: rich models, fast inference

Future Work

- Better bounds for approximate inference?
- Dealing with changing models/weights
- Explicitly modeling change in models
- Applications:
  - Drug targeting
  - Knowledge Graph construction
  - Context-aware mobile devices