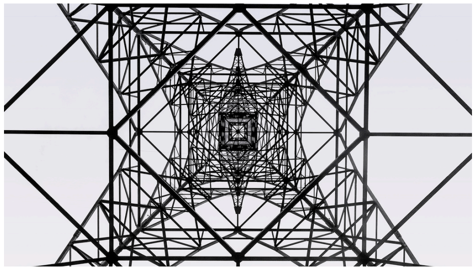
EXPLOITING STRUCTURE FOR META-LEARNING

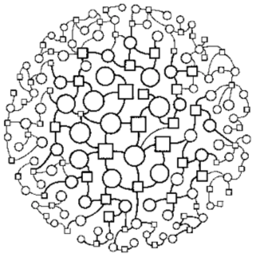
NeurIPS Metalearning Workshop | December 8, 2018

Lise Getoor | UC Santa Cruz
| @lgetoor

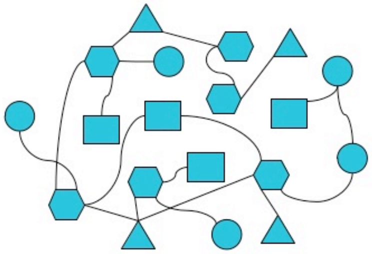
STRUCTURE



STRUCTURE IN
INPUTS



STRUCTURE IN
OUTPUTS



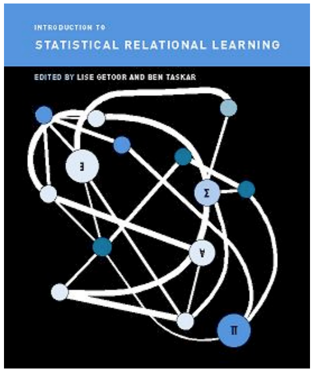
STRUCTURE IN
META-LEARNING MODEL

THIS TALK

Structure & Meta-learning

STATISTICAL RELATIONAL LEARNING

1. Make use of logical structure
2. Handle uncertainty
3. Perform collective inference



[GETOOR & TASKAR ’07]

PROBABILISTIC SOFT LOGIC (PSL)

A probabilistic programming language for collective inference problems

- Predicate = **relationship** or **property**
- Ground Atom = (continuous) random variable
- Weighted Rules = capture **dependency** or **constraint**

PSL Program = Rules + Input DB

psl.linqs.org

KEY REFERENCE: Hinge-Loss Markov Random Fields and Probabilistic Soft Logic, Stephen Bach, Matthias Broecheler, Bert Huang, Lise Getoor, JMLR 2017

COLLECTIVE
Reasoning

outputs depend
on each other

COLLECTIVE

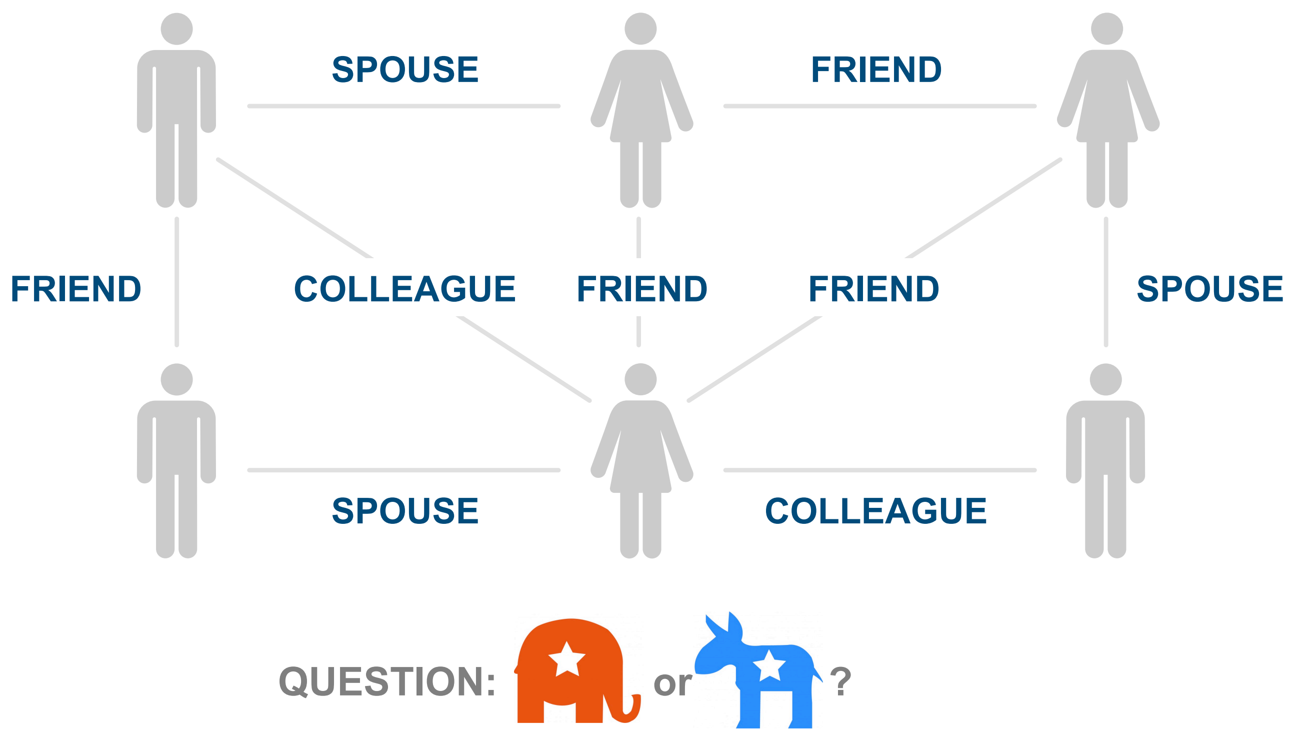
Classification Pattern

```
local-predictor(x,l) → label(x,l)
label(x,l) & link(x,y) → label(y,l)
```

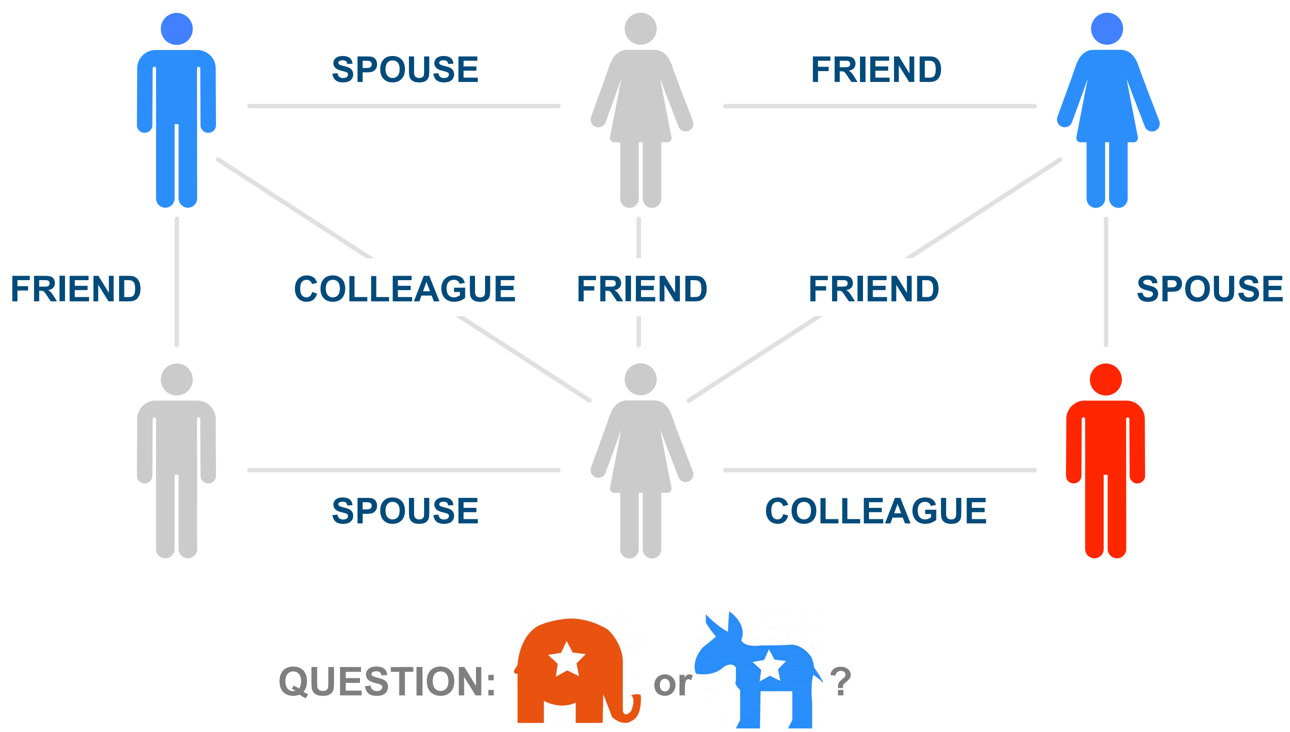
COLLECTIVE
Classification Pattern

```
local-predictor(x,l) → label(x,l)
label(x,l) & link(x,y) → label(y,l)
```

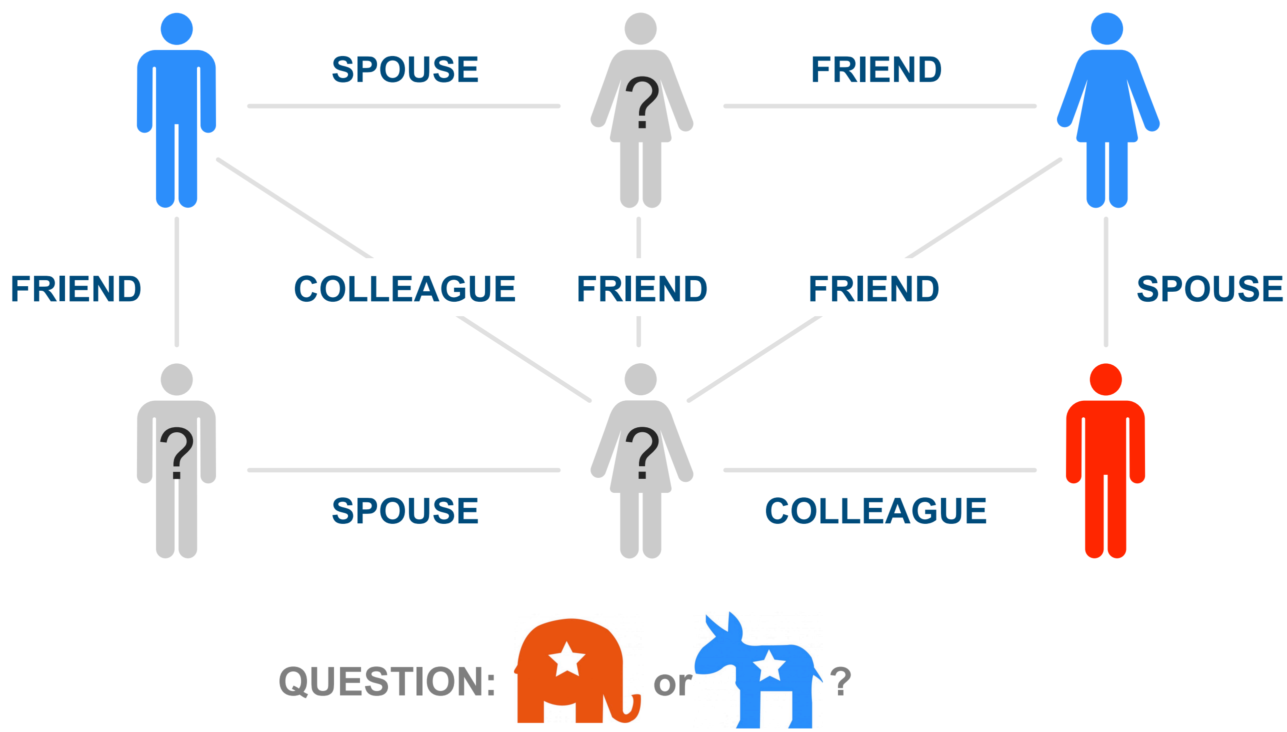
COLLECTIVE CLASSIFICATION



COLLECTIVE CLASSIFICATION



COLLECTIVE CLASSIFICATION



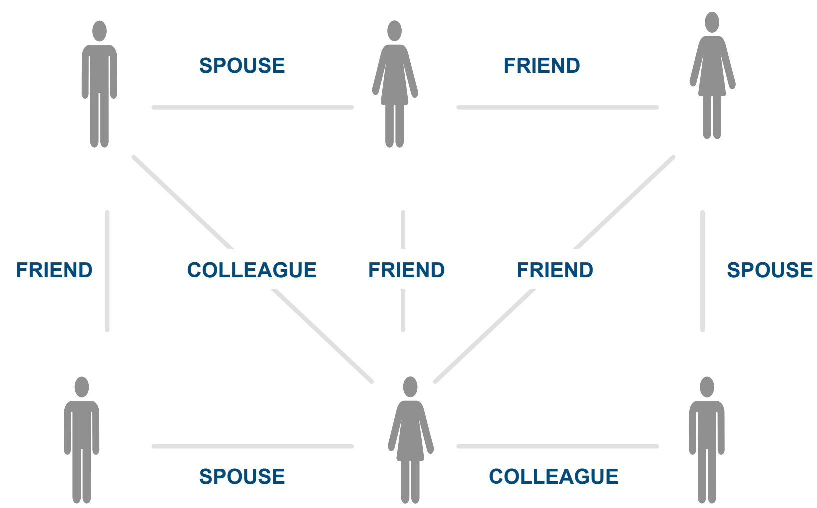
COLLECTIVE CLASSIFICATION

Local rules:

- "If X donates to party P, X votes for P"
- "If X tweets party P slogans, X votes for P"

Relational rules:

- "If X is linked to Y, and X votes for P, Y votes for P"



COLLECTIVE CLASSIFICATION

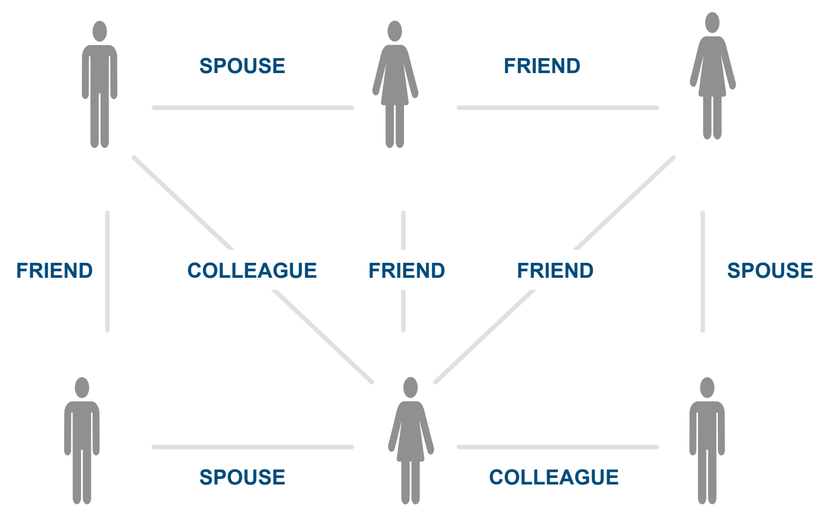
Local rules:

- **“If X donates to party P, X votes for P”**
- “If X tweets party P slogans, X votes for P”

Relational rules:

- “If X is linked to Y, and X votes for P, Y votes for P”

```
Donates(X,P) → Votes(X,P)
```



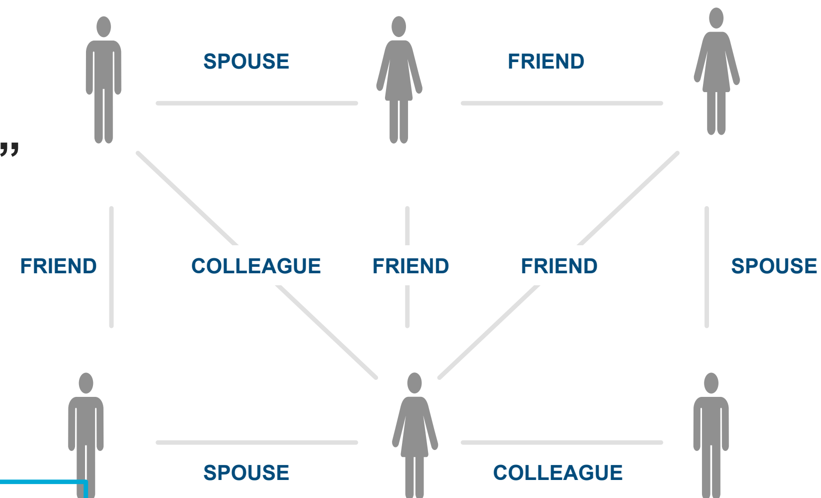
COLLECTIVE CLASSIFICATION

Local rules:

- "If X donates to party P, X votes for P"
- **"If X tweets party P slogans, X votes for P"**

Relational rules:

- "If X is linked to Y, and X votes for P, Y votes for P"




```
Tweets(X,"Affordable Health")
    → Votes(X,"Democrat")
```

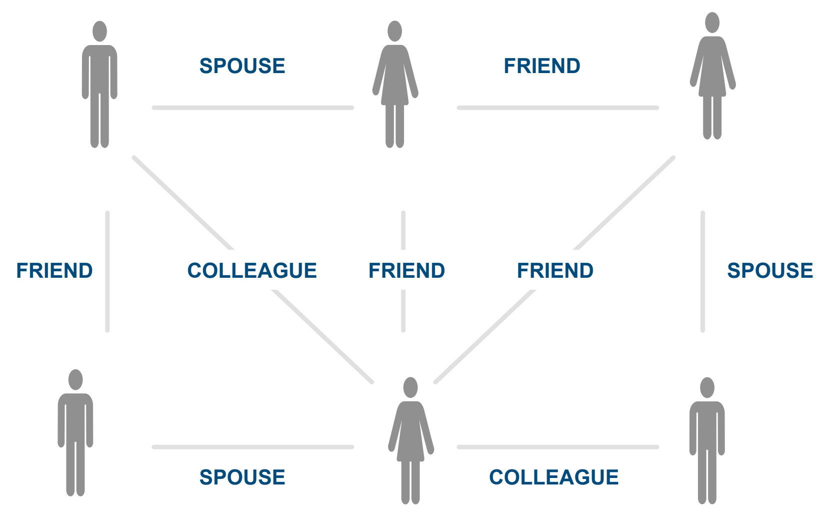
COLLECTIVE CLASSIFICATION

Local rules:

- "If X donates to party P, X votes for P"
- "If X tweets party P slogans, X votes for P"

Relational rules:

- **"If X is linked to Y, and X votes for P, Y votes for P"**



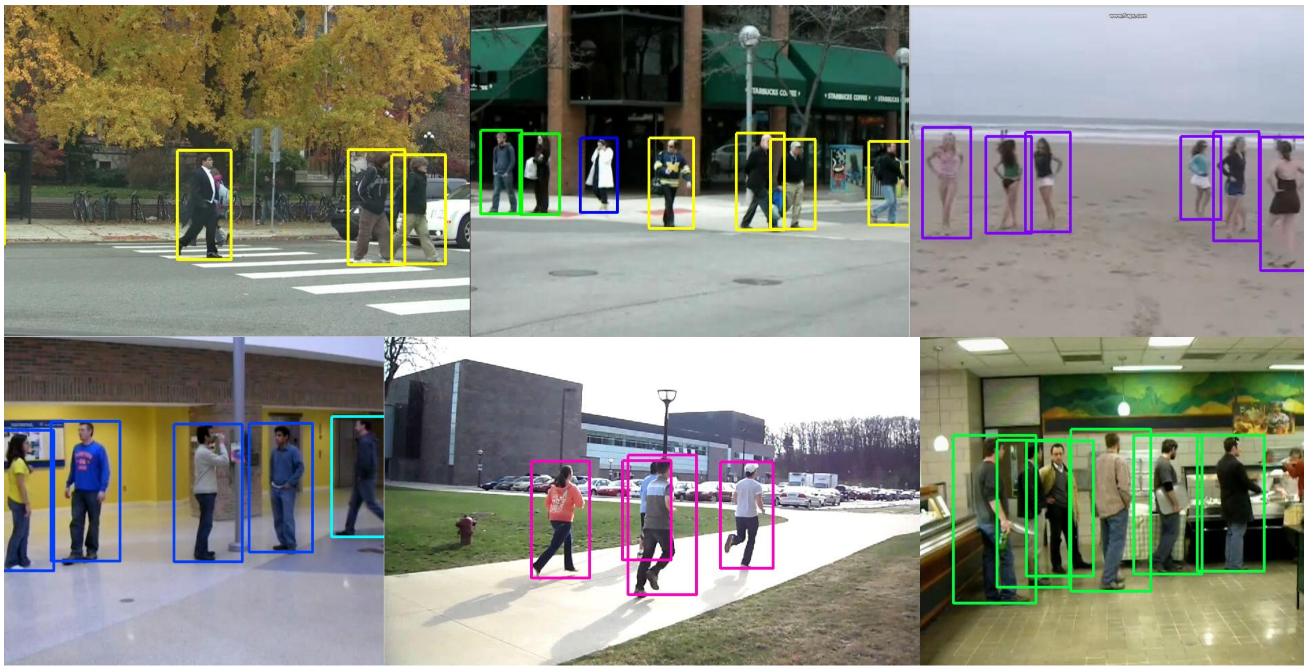

```
Votes(X,P) & Friends(X,Y)→
         Votes(Y,P)
```

```
Votes(X,P) & Spouse(X,Y) →
         Votes(Y,P)
```

COLLECTIVE
Activity Recognition

inferring activities
in video sequence

ACTIVITY RECOGNITION



crossing waiting queueing walking talking dancing jogging

COLLECTIVE
Pattern

```
local-predictor(x,l,f) → activity(x,l,f)
activity(x,l,f) & same-frame(x,y,f) → activity(y,l,f)
activity(x,l,f) & next-frame(f,f') → activity(x,l,f')
```

EMPIRICAL HIGHLIGHTS

Improved activity recognition in video:

	5 Activities		6 Activities	
HOG	47.4%	.481 F1	59.6%	.582 F1
HOG + PSL	59.8%	.603 F1	79.3%	.789 F1
ACD	67.5%	.678 F1	83.5%	.835 F1
ACD + PSL	69.2%	.693 F1	86.0%	.860 F1

London et al., *Collective Activity Detection using Hinge-loss Markov Random Fields*, CVPR WS 13

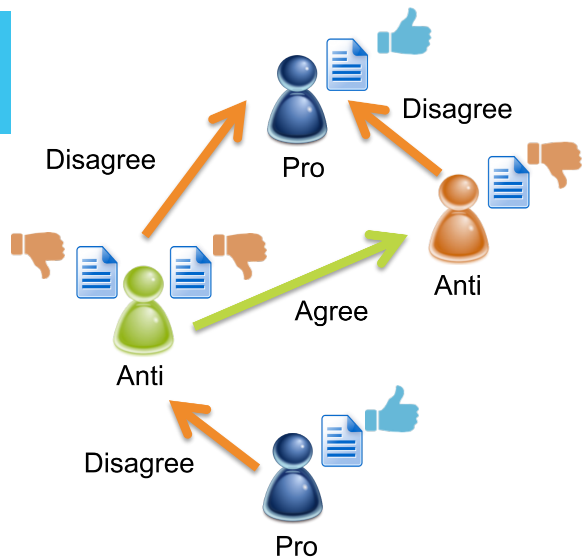
COLLECTIVE
Stance Prediction

Inferring users' stance in online debates

DHANYA SRIDHAR

DEBATE STANCE CLASSIFICATION

TOPIC:
Climate Change



TASK:
Jointly infer users' attitude on topics and interaction polarity

Sridhar, Foulds, Huang, Getoor & Walker, *Joint Models of Disagreement and Stance*, ACL 2015

PSL FOR STANCE CLASSIFICATION

```
// local text classifiers
  w1: LocalPro(U,T)          ->  Pro(U,T)
  w1: LocalDisagree(U1,U2)   ->  Disagrees(U1,U2)

//Rules for stance
  w2: Pro(U1,T) & Disagrees(U1,U2)  -> !Pro(U2,T)
  w2: Pro(U1,T) & !Disagrees(U1,U2) ->  Pro(U2,T)

//Rules for disagreement
  w3:  Pro(U1,T)  &  Pro(U1,T)  ->  !Disagrees(U1,U2)
  w3: !Pro(U1,T)  &  Pro(U2,T)  ->   Disagrees(U1,U2)
```

bitbucket.org/linqs/psl-joint-stance

PREDICTING STANCE IN ONLINE FORUMS

Task: Predict post and user stance from two online debate forums

- 4Forums.com: ~300 users,~6000 posts
- CreateDebate.org: ~300 users, ~1200 posts

4FORUMS.COM

	ACCURACY
Text-only Baseline	69.0
PSL	80.3

CREATEDEBATE.ORG

	ACCURACY
Text-only Baseline	62.7
PSL	72.7

Sridhar, Foulds, Huang, Getoor & Walker, *Joint Models of Disagreement and Stance*, ACL 2015

LINK
Prediction Pattern

```
link(x,y) & similar(y,z) → link(x,z)
```

CLUSTERING
Pattern

```
link(x,y) & link(y,z) → link(x,z)
```

MATCHING

Pattern

```
link(x,y) & !same(y,z) → !link(x,z)
```

THIS TALK

Structure & Meta-learning

SRL <-> META-LEARN

SRL Concepts

Templated Models

Weight Learning

Structure Learning

Latent Variables

Logical rules

Meta-learning Concepts

Tied Hyperparameters

Hyperparameter Optimization

Feature & Algorithm Selection

Landmarks

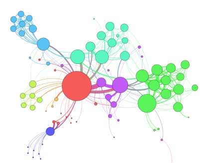
Few/Zero-shot learning

TEMPLATING

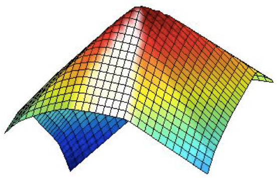
Probabilistic programming language for defining distributions

```
/* Local rules */
w_d: Donates(A, P) -> Votes(A, P)
w_t: Mentions(A, “Affordable Health”) ->
Votes(A, “Democrat”)
w_t: Mentions(A, “Tax Cuts”) -> Votes(A,
“Republican”)

/* Relational rules */
w_s: Votes(A,P) & Spouse(B,A) -> Votes(B,P)
w_f: Votes(A,P) & Friend(B,A) -> Votes(B,P)
w_c: Votes(A,P) & Colleague(B,A) -> Votes(B,P)

/* Range constraint */
Votes(A, “Republican”) + Votes(A, “Democrat”)
= 1.0
```

+



=



LEARN

when structural patterns hold across many instantiations

STRUCTURE LEARNING

- Large subfield of statistical relational learning
  - Friedman et al. IJCAI 99, Getoor et al. JMLR 02, Kok & Domingos ICML05, Mihalkova & Mooney ICML07, DeRaedt et al. MLJ 2008, Khosravi et al AAAI10, Khot et al. ICDM 11, Van Haaren et al. MLJ15, among others
  - NIPS Relational Representation Learning Workshop

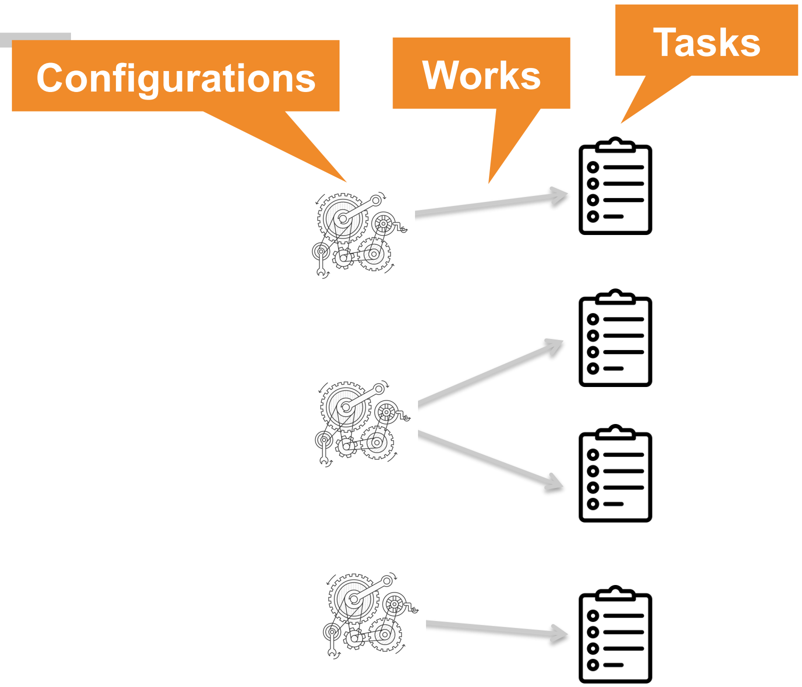
- Basic Idea
  - Search model space
    - Model space is very rich
  - Optimize parameters
    - Information theoretic criteria, likelihood-based, and Bayesian approaches

META

when structural patterns hold
across many learning tasks

LEARN

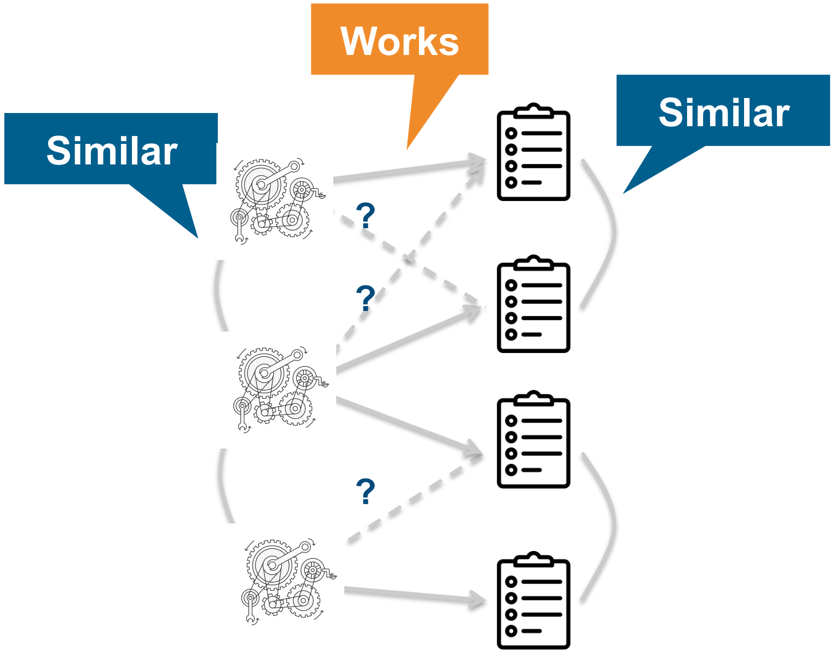
META LEARNING



META LEARNING

Rules express:

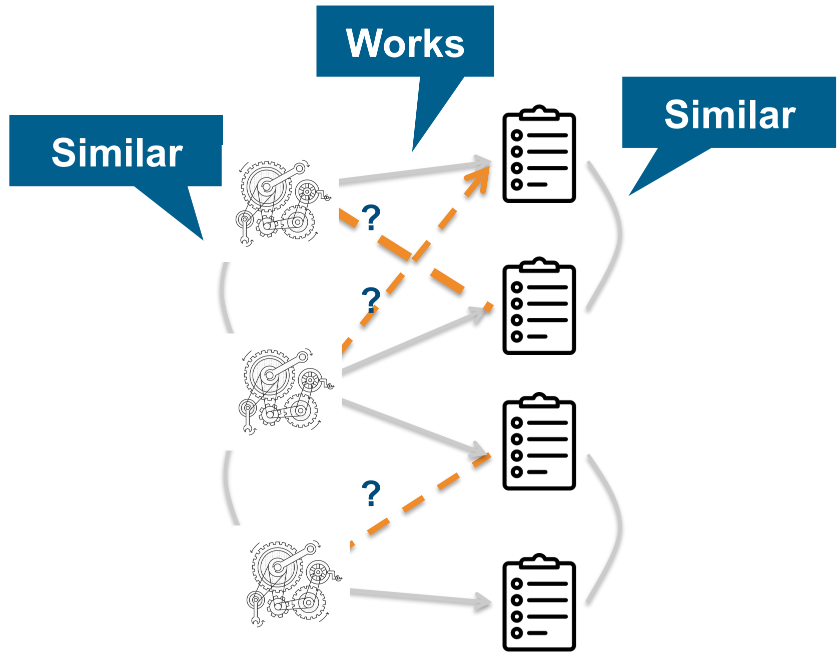
- “If configuration C works well for task T1, and task T2 is similar to T1, C will work well for T2”
- “If configuration C1 works well for task T, and configuration C2 similar to C1, C2 will work well for T”



META LEARNING

Rules express:

- **“If configuration C works well for task T1, and task T2 is similar to T1, C will work well for T2”**
- “If configuration C1 works well for task T, and configuration C2 similar to C1, C2 will work well for T”

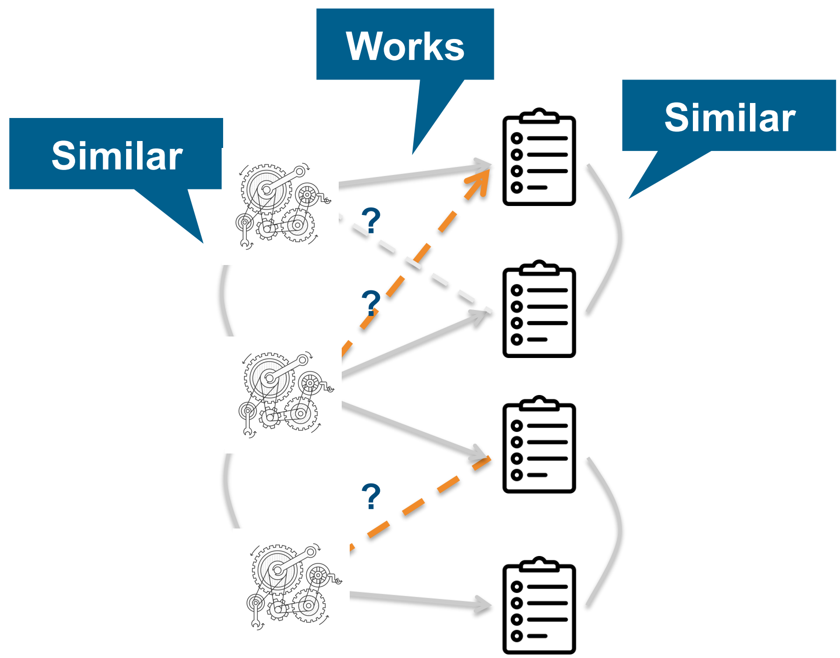



```
Works(C,T1) & SimilarTask(T1,T2) → Works(C,T2)
```

META LEARNING

Rules express:

- "If configuration C works well for task T1, and task T2 is similar to T1, C will work well for T2"
- **"If configuration C1 works well for task T, and configuration C2 similar to C1, C2 will work well for T"**




```
Works(C1,T) & SimilarConfig(C1,C2) → Works(C2,T)
```

META-LEARNING

- Challenge: defining similarity

- Advantages:
  - can make use of multiple similarity measures
  - can use domain knowledge for defining task and configuration similarity

- Research questions:
  - Are there benefits from using this approach?
  - What are opportunities for collective reasoning?

LANDMARKING

- Can be described using latent variables
- E.g., Task-Area and Learner-Expertise as latent variables
- Research questions:
  - Are there benefits from using SRL approach?
  - What are opportunities for collective reasoning?

ALGORITHM & MODEL SELECTION

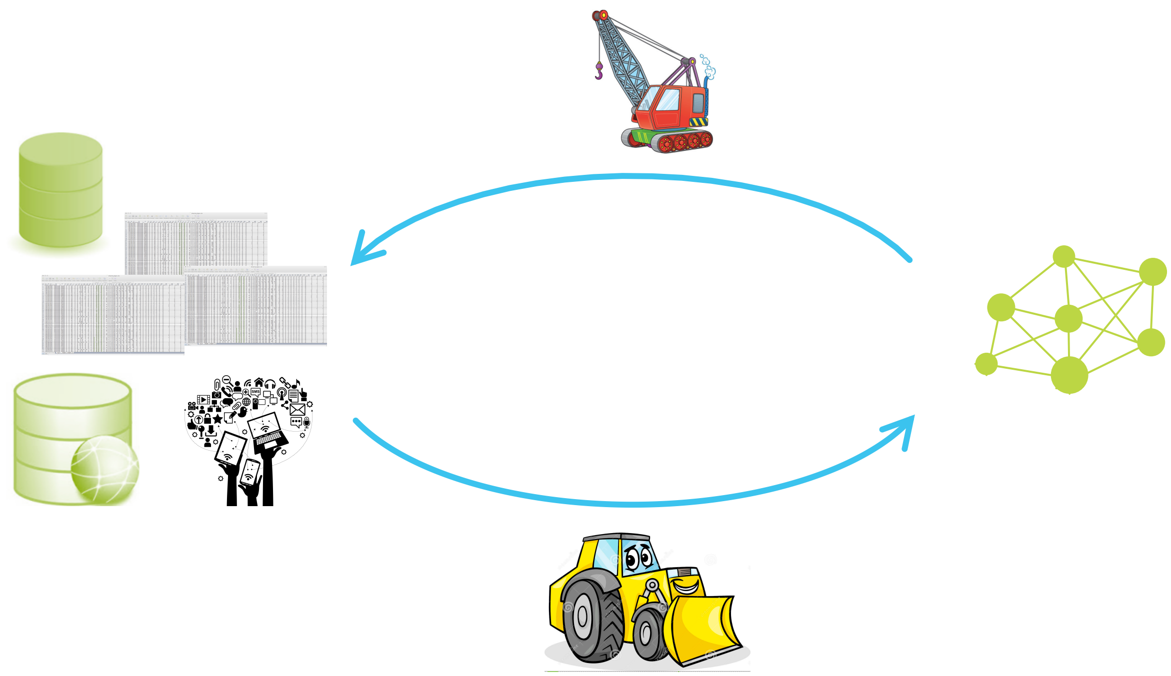
- Can be described using (probabilistic/soft) logical rules
- Research questions:
  - Are there benefits from using SRL approach?
  - What are opportunities for collective reasoning?

PIPELINE CONSTRUCTION

- Can be described using logical rules and constraints
- Research questions:
  - Are there benefits from using SRL approach?
  - What are opportunities for collective reasoning?

CLOSING

STRUCTURE AND META-LEARNING



CLOSING THE LOOP

CLOSING COMMENTS

Provided some examples of structure and collective reasoning

Opportunity for Meta-Learning methods that can mix:

- probabilistic & logical inference
- data-driven & knowledge-driven modeling
- Meta-modeling for meta-modeling

OPPORTUNITY!

Compelling applications abound!

THANK YOU!

PROBABILISTIC
SOFT
LOGIC

psl.linqs.org

Contact information:

getoor@ucsc.edu

| @lgetoor

PSL SUMMARY IN A SLIDE

- MAP Inference in PSL translates into convex optimization problem → **inference is really fast**
- Inference further enhanced with state-of-the-art optimization and distributed graph processing paradigms →**inference even faster**
- **Learning** methods for **rule weights** & **latent variables**
- PSL is **open-source**, code, data, tutorials available online

psl.linqs.org