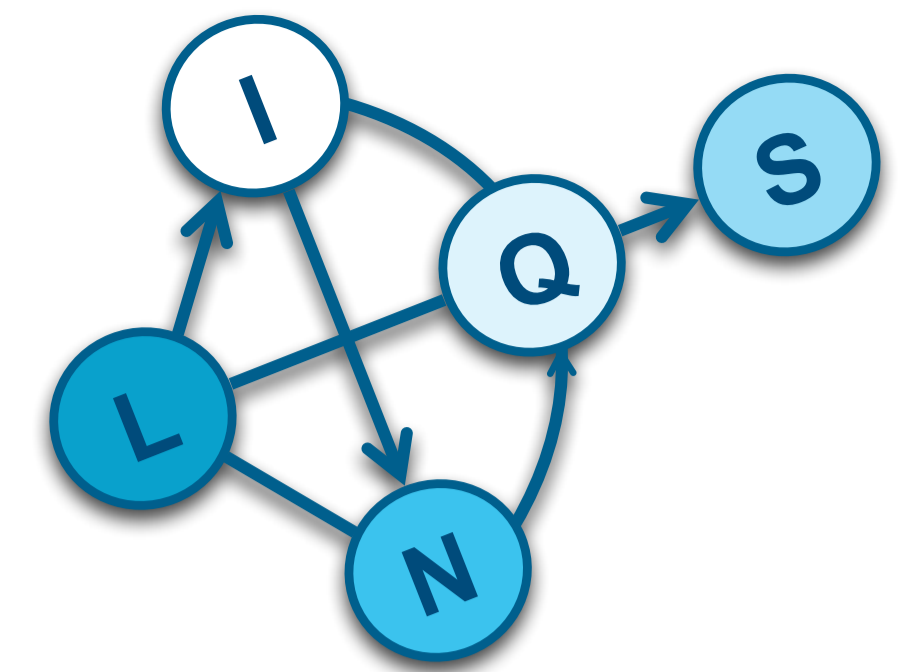


# User Role Prediction in Online Discussion Forums using Probabilistic Soft Logic

Arti Ramesh<sup>1</sup>, Jaebong Yoo<sup>2</sup>, Shitian Shen<sup>2</sup>, Lise Getoor<sup>1</sup> and Jihie Kim<sup>2</sup>  
<sup>1</sup>Department of Computer Science, University of Maryland, College Park  
<sup>2</sup>Information Sciences Institute, University of Southern California



## User Role Prediction

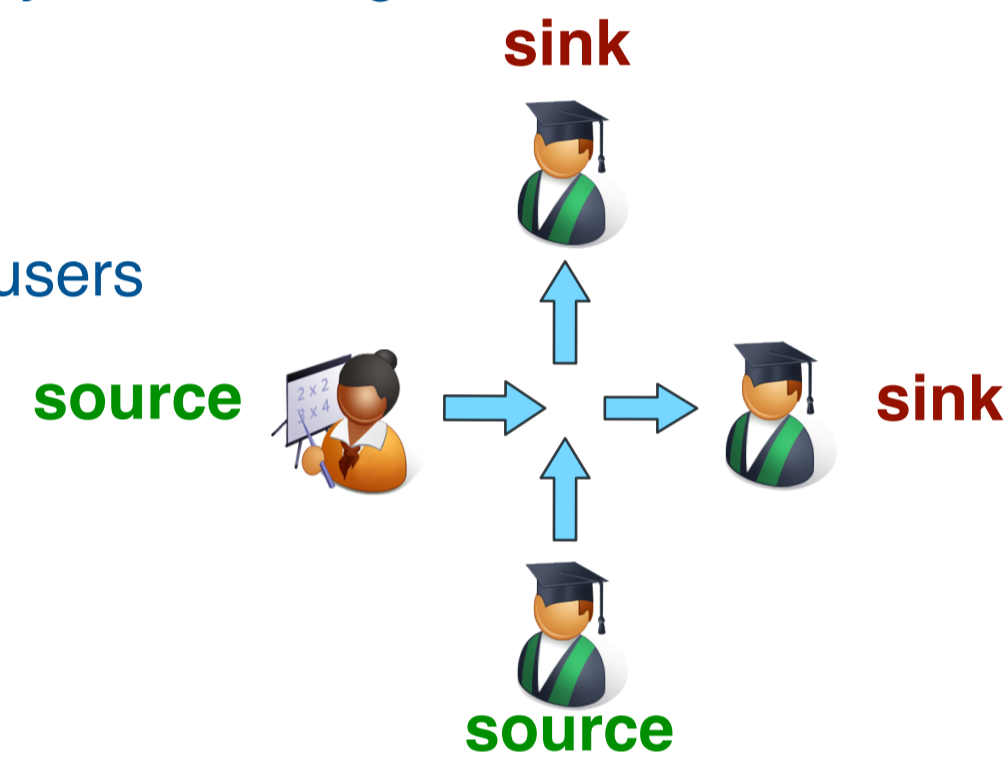
### Motivation

- Online discussions: popular tool for supporting student communication in distance and classroom education
- Identifying **user role** is key to modeling online student discussion

### Goal

Identify two major roles for users

- Information Seeker (**sink**)
- Information Provider (**source**)



## Probabilistic Soft Logic

Declarative language for probabilistic models using **first-order logic (FOL)** syntax

Truth values are relaxed to **soft truth** in [0,1]

Mechanisms for incorporating **similarity functions** and reasoning about **sets**

**Efficient** inference: convex optimization in continuous space

Maximum Likelihood weight learning via voted perceptron

Continuous probability distribution over truth values:

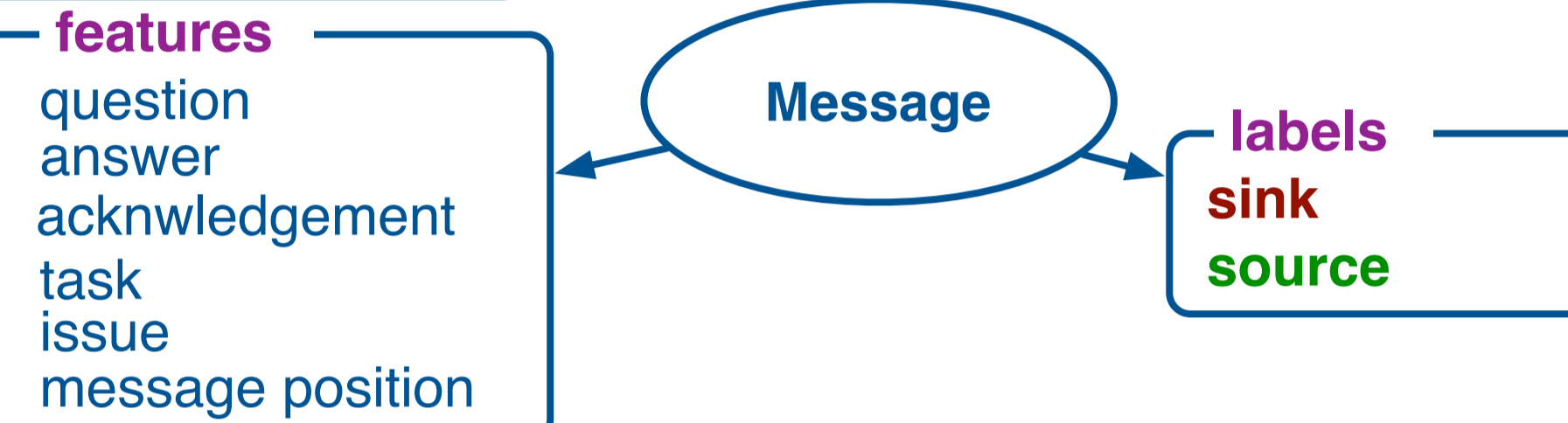
$$\Pr(x) = \frac{1}{Z} \exp \left( - \sum_{r \in P} \sum_{g \in G(r)} w_r (1 - t_g(x)) \right)$$

$P$ : the PSL program     $G(r)$ : set of all groundings of rule  $r$   
 $w_r$ : weight of rule  $r$      $t_g(x)$ : truth value of grounding  $g$

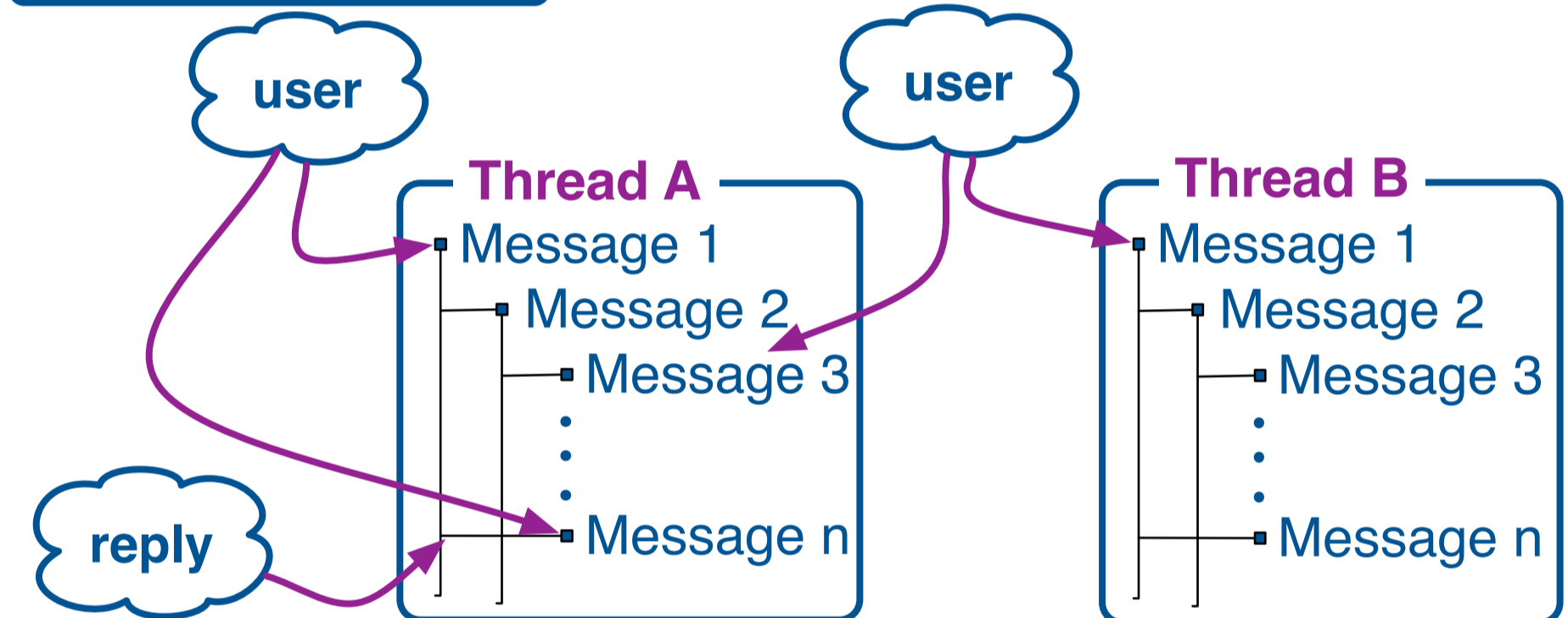
**Software:** <http://psl.umiaccs.umd.edu>

## Online Discussion Data

### Local Features



### Network Features



Online discussion data organized into **threads**; threads broken down into **messages**  
**"reply"** connects messages in a thread  
**"user"** connects messages posted by same user within and across threads

## PSL Rules

Messages –  $M_1, M_2$

### Local Rules

QUESTION( $M_1$ )  $\wedge$  FIRSTMESSAGE( $M_1$ )  $\Rightarrow$  SINK( $M_2$ )  
ANSWER( $M_1$ )  $\wedge$  LASTMESSAGE( $M_1$ )  $\Rightarrow$  SOURCE( $M_1$ )  
TASK( $M_1$ )  $\Rightarrow$  SOURCE( $M_1$ )

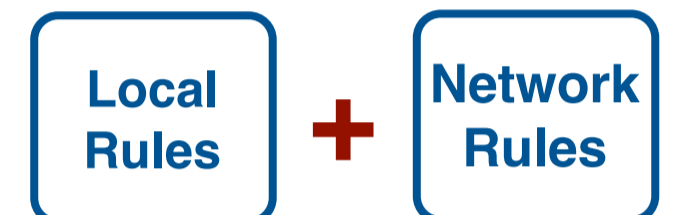
### Network Rules

SAMETHREAD( $M_1, M_2$ )  $\wedge$  SAMEUSER( $M_1, M_2$ )  $\wedge$  SINK( $M_1$ )  $\Rightarrow$  SINK( $M_2$ )  
 $\neg$ SAMETHREAD( $M_1, M_2$ )  $\wedge$  SAMEUSER( $M_1, M_2$ )  $\wedge$  SINK( $M_1$ )  $\Rightarrow$  SINK( $M_2$ )  
REPLY( $M_1, M_2$ )  $\wedge$  SINK( $M_2$ )  $\wedge$   $\neg$ SAMEUSER( $M_1, M_2$ )  $\Rightarrow$  SOURCE( $M_2$ )

## PSL Online Discussion Models

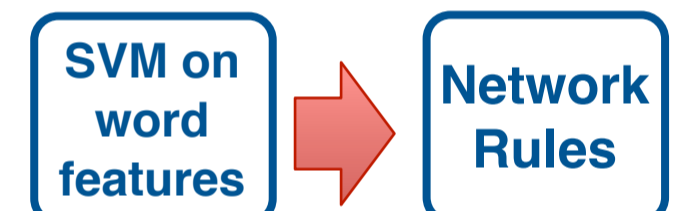
### Model 1

- Combines **local** and **network** rules
- Maximum likelihood weight learning via voted perceptron
- Weight values indicate importance of local and network features

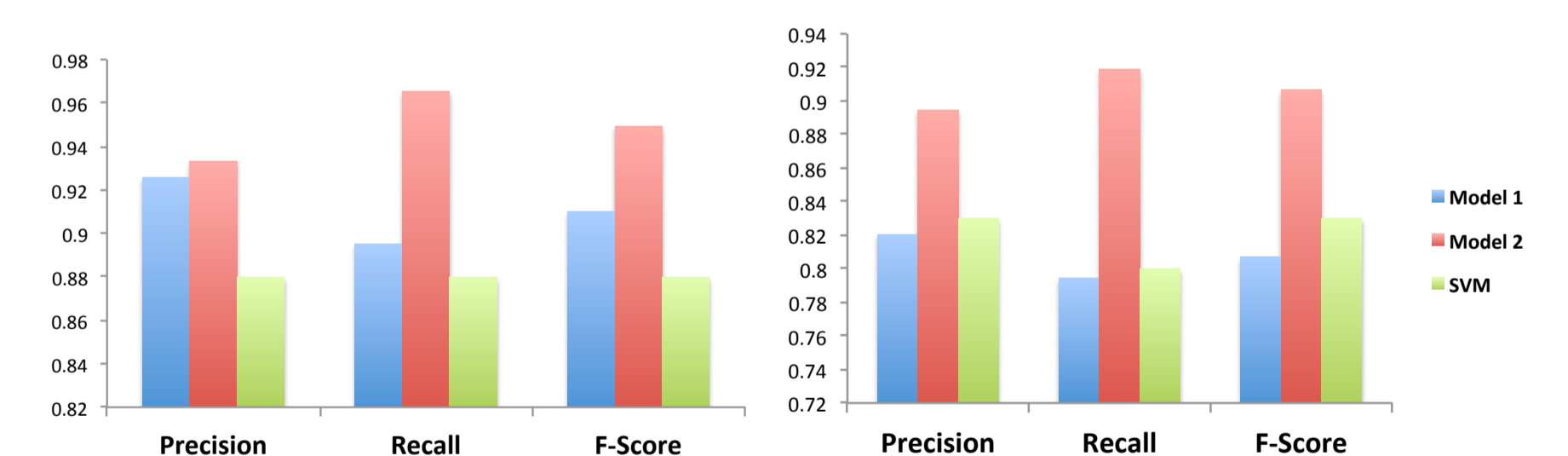


### Model 2

- SVM model using bigram/trigram word features in messages
- PSL network rules act on SVM label outputs to improve predictions



## Evaluation



Model 2 **outperforms** Model 1 and SVM

PSL network rules take advantage of predictions generated by SVM model, propagate beliefs across network better

## Discussion/Future Work

- PSL network rules provide convenient and effective means to model relationships in online discussion domain
- Model can be extended to infer student's understanding of subject, participation, credibility of answers in online forums