A Community-Based Pseudolikelihood Approach for Relationship Labeling in Social Networks

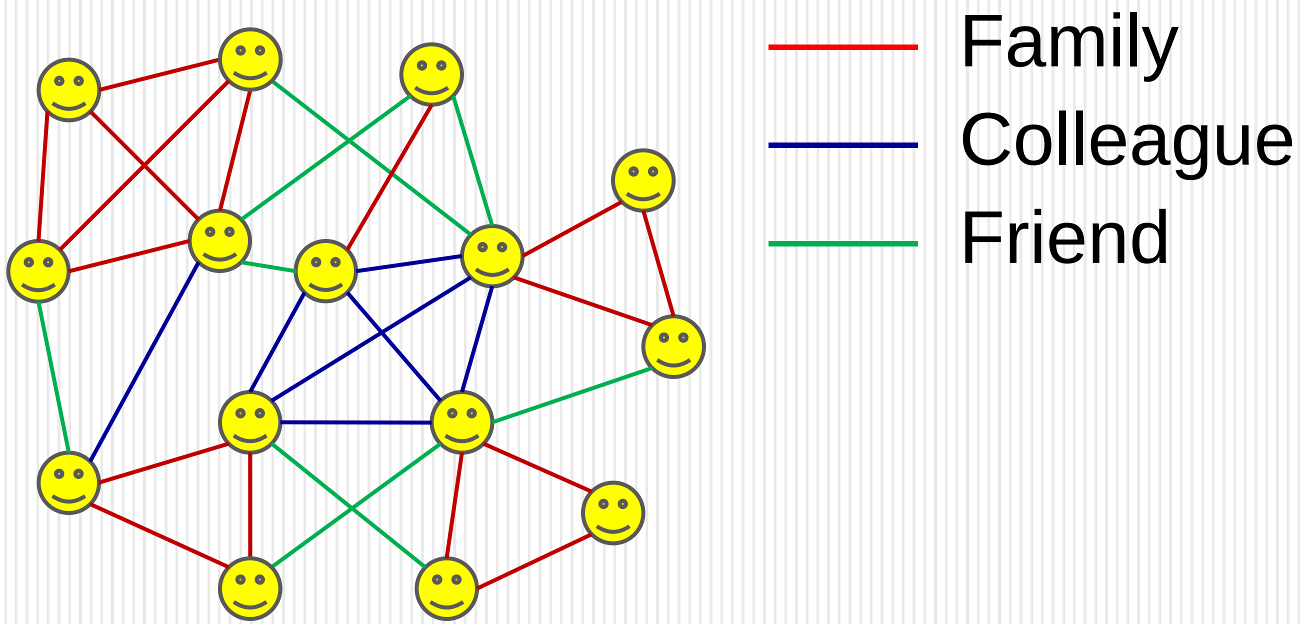
Huaiyu Wan, Youfang Lin, Zhihao Wu, Houkuan Huang

School of Computer and Information Technology
Beijing Jiaotong University

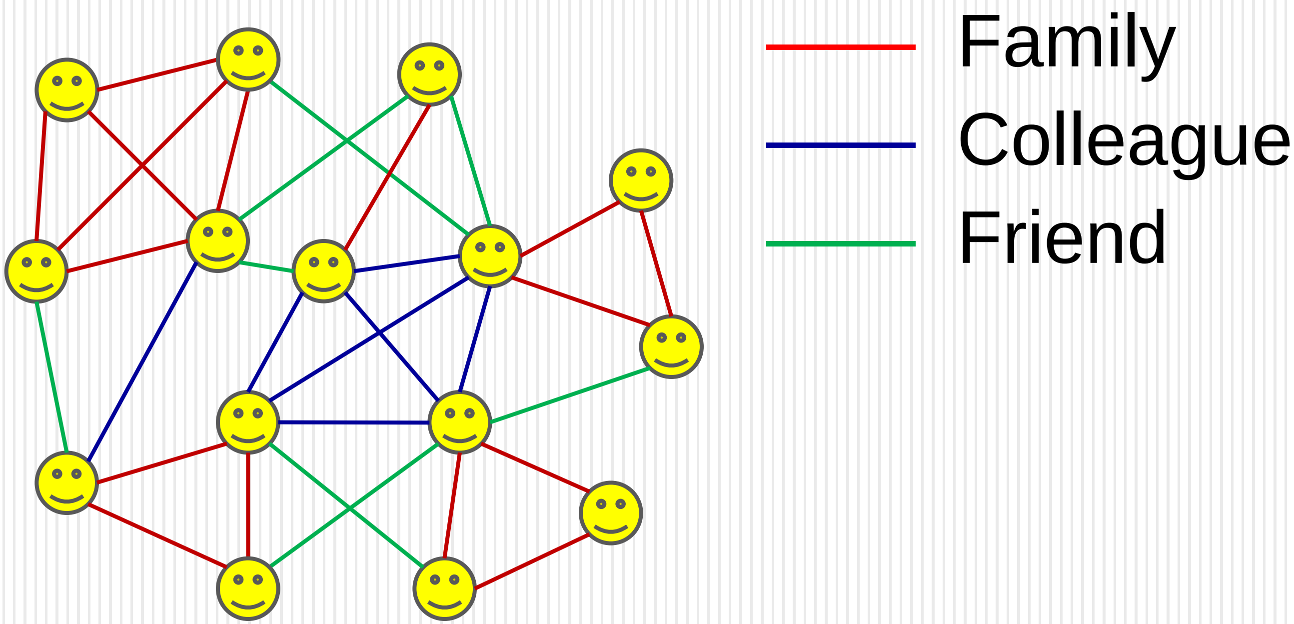
Outline

- Introduction
- Related Works
- Our Proposed Approach
- Experiments & Results
- Conclusion & Future Work

Relationship Labeling Problem



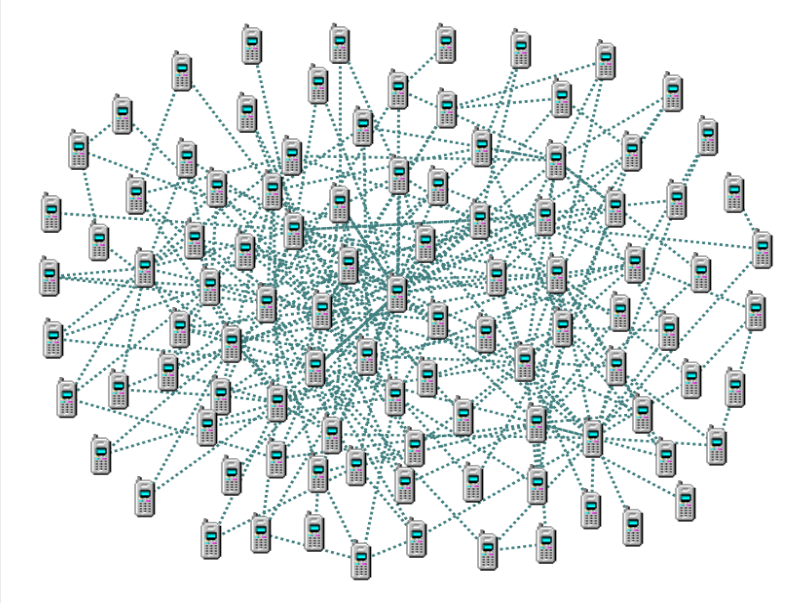
Relationship Labeling Problem



Given a snapshot of a social network, can we infer the types of the relationships between the individuals?

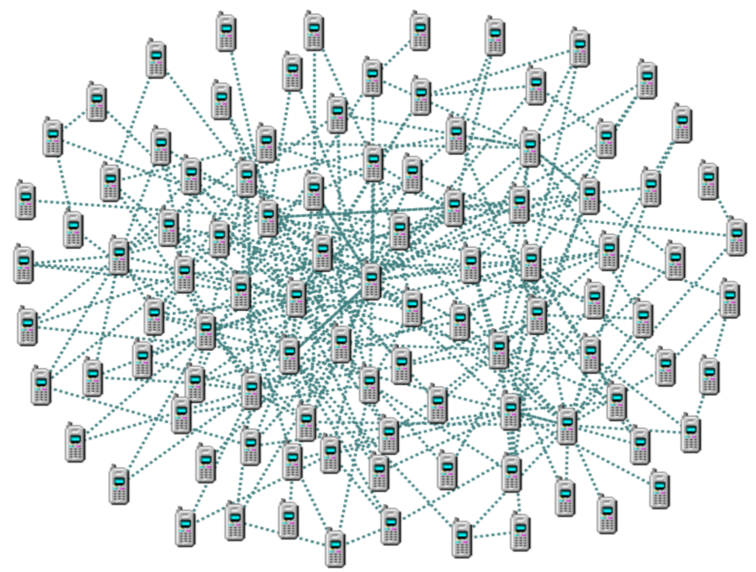
Real-life Applications

- Mobile Social Networks



Real-life Applications

- Mobile Social Networks



- Online Social Networks
- Criminal Networks
- …

How to label relationships?

- Traditional Classifiers
  - use only content attributes
  - require the IID assumption

How to label relationships?

- Traditional Classifiers
  - use only content attributes
  - require the IID assumption
- Relational Classifiers
  - Taskar et al. proposed relational Markov networks (RMNs) and use it for link prediction
  - Zhao et al. use RMNs to label the relationships in terrorist social networks

Advantages of RMNs

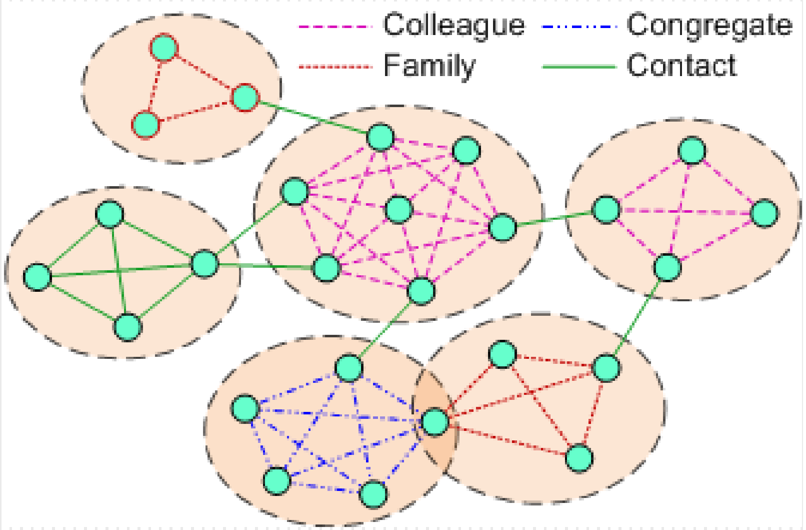
- Collective Classification
  - not require the IID assumption
  - more accurate then separate classification
- Undirected Graphical Model
  - not require the acyclicity constraint
- Discriminative Training
  - more accurate over generative training given enough training examples

Disadvantages of RMNs

- High Computational Complexity
  - multiple rounds of approximate inference
- Uncertain Prediction Accuracy
  - the accuracy is directly dependent on the definition of relational clique templates

Ideas of Our Approach

- Using Community Structure to Assist Collective Classification
  - community structure is one of the most important properties
  - “birds of a feather flock together”



Ideas of Our Approach

- Using the Pseudolikelihood Technique to Drop the Computational Complexity
  - an effective alternative of likelihood
  - successfully used in relational learning field

Detailed Steps of Our Approach

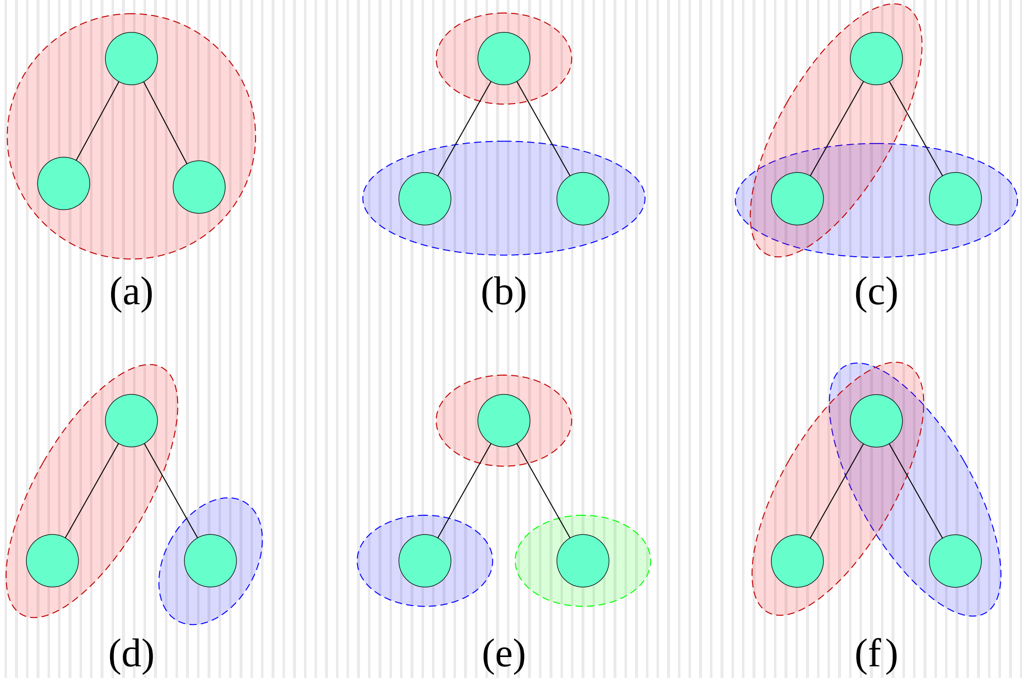
- Step 1: Community Detection
  - to find a meaningful division of nodes in a social network
  - non-overlapping & overlapping algorithms

Detailed Steps of Our Approach

- Step 2: Constructing CRFs
  - goal: define the cliques of CRFs (i.e., establish the links between the relationship labels)
  - just consider pairwise CRFs here

Detailed Steps of Our Approach

- Step 2: Constructing CRFs
  - goal: define the cliques of CRFs (i.e., establish the dependencies between the relationship labels)
  - just consider pairwise CRFs here



(a) (b) (c) (d) (e) (f)

Detailed Steps of Our Approach

➢ Step 3: The Pseudolikelihood Model

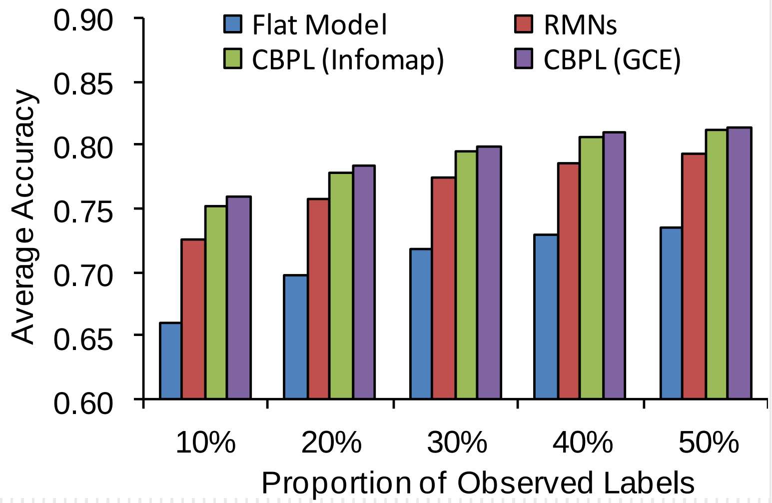
	Pseudolikelihood	Likelihood
Joint Probability	$P(\mathbf{y} \mid \mathbf{x},\mathbf{r}) = \prod_{i=1}^{n} P(y_i \mid MB(y_i))$ $P(y_i \mid MB(y_i)) = \frac{1}{Z_i(\mathbf{x}_i,\mathbf{r}_i)} \prod_{v_j \in MB(y_i)} \phi(y_i, v_j)$	$P(\mathbf{y} \mid \mathbf{x},\mathbf{r}) = \frac{1}{Z(\mathbf{x},\mathbf{r})} \prod_{C \in \mathrm{C}} \prod_{c \in C(I)} \phi_C(\mathbf{x}_c, \mathbf{y}_c)$
Partition Function	$Z_i(\mathbf{x}_i,\mathbf{r}_i) = \sum_{y_i'} \prod_{v_j \in MB(y_i')} \phi(y_i', v_j)$	$Z(\mathbf{x},\mathbf{r}) = \sum_{\mathbf{y}'} \prod_{C \in \mathrm{C}} \prod_{c \in C(I)} \phi_C(\mathbf{x}_c, \mathbf{y}'_c)$

Experiments

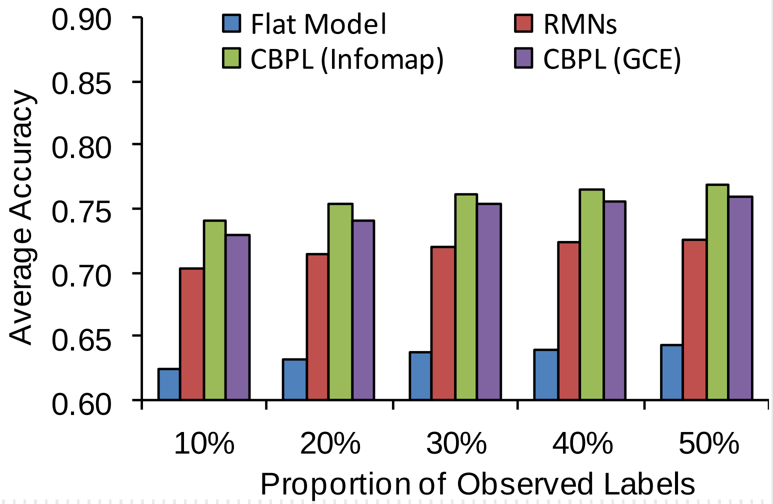
- Datasets
  - a terrorist networks
    - ✓ 244 nodes and 840 relationships
    - ✓ 1224 features for each relationship
  - a mobile social networks
    - ✓ 1623 nodes & 4295 relationships
    - ✓ take the communication information between two users as the features of their relationships
    - ✓ use service packages (family or group package) to label the dataset

Experiments

- Results
  - average accuracy



The terrorist social network



The mobile social network

Experiments

- Results
  - average training times

The terrorist social network

Approach	Proportion of Observed Labels				
	10%	20%	30%	40%	50%
Flat Model	0.81	2.06	3.64	7.79	11.83
RMNs	4.49	25.86	96.41	289.05	820.60
CBPL (Infomap)	**2.01**	**6.02**	**15.48**	**34.85**	**51.27**
CBPL (GCE)	**2.53**	**7.91**	**18.96**	**39.58**	**58.73**

The mobile social network

Approach	Proportion of Observed Labels				
	10%	20%	30%	40%	50%
Flat Model	0.79	1.64	2.45	3.31	5.87
RMNs	6.53	33.62	133.84	437.76	1362.54
CBPL (Infomap)	**1.26**	**4.92**	**12.85**	**27.41**	**46.05**
CBPL (GCE)	**1.67**	**5.63**	**15.39**	**32.27**	**53.72**

Conclusion & Future Works

- proposed a relationship labeling approach which:
  - uses community structure to assist construct CRFs
  - uses the pseudolikelihood technique to drop the computational complexity

Conclusion & Future Works

- proposed a relationship labeling approach which:
  - uses community structure to assist construct CRFs
  - uses the pseudolikelihood technique to drop the computational complexity
- Future Works:
  - the quantification of community structure
  - the generalization of our approach to multipartite or even multimode networks

Thank you!

Huaiyu Wan, Youfang Lin, Zhihao Wu, Houkuan Huang

School of Computer and Information Technology
Beijing Jiaotong University