Social Spammer Detection in Microblogging

Xia (Ben) Hu, Jiliang Tang, Yanchao Zhang, Huan Liu
Arizona State University

23rd International Joint Conference on Artificial Intelligence (**IJCAI 2013**)

Outline

- Background and Motivation
- Proposed Sociological Approach
- Experimental Evaluation
- Conclusions and Future Work

Background

- Microblogging has become a widely popular platform for information dissemination and sharing in various scenarios

- With the growing availability of microblogging, social spamming has become rampant. Social spammers are employed to unfairly overpower normal users

Background

- A traditional assumption in spammer detection is that spammers cannot establish an arbitrarily large number of social trust relations with normal users

- However, different from other OSNs, microblogging systems feature unidirectional user binding. Many users simply follow back when they are followed by someone for the sake of courtesy -- ***reflexive reciprocity***

Motivation

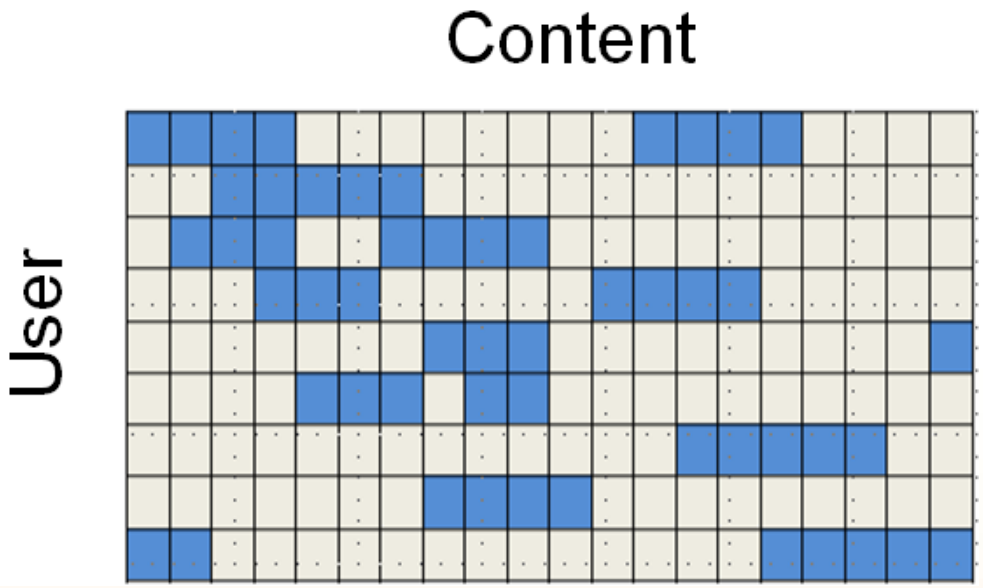
- Spammers can successfully acquire a number of normal followers, especially those referred to as social capitalists who tend to increase their social capital by following back anyone who follows them

- Microblogging provides additional content information
  - Short length
  - Unstructured form
  - Connected texts

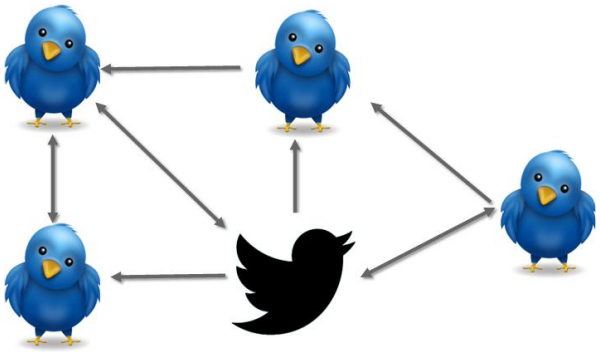
Outline

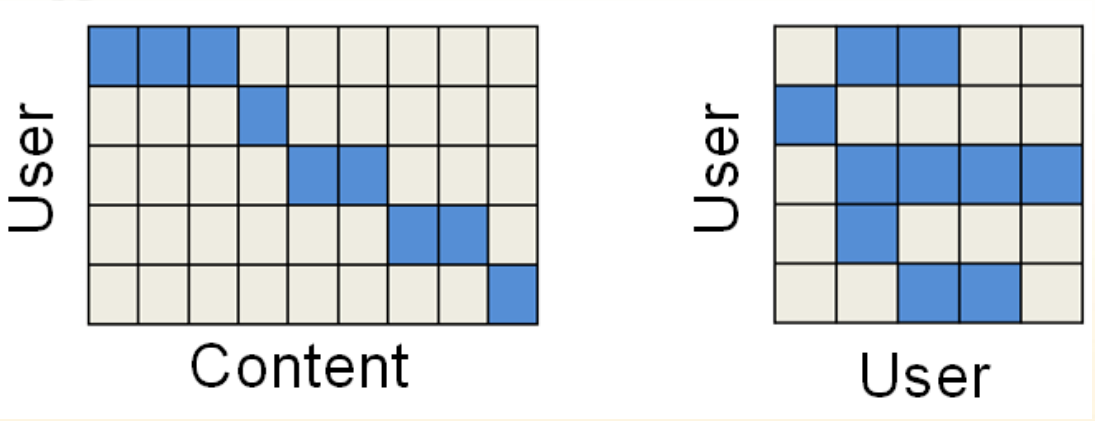
- Background and Motivation
- Proposed Sociological Approach
- Experimental Evaluation
- Conclusions and Future Work

Traditional Spammers

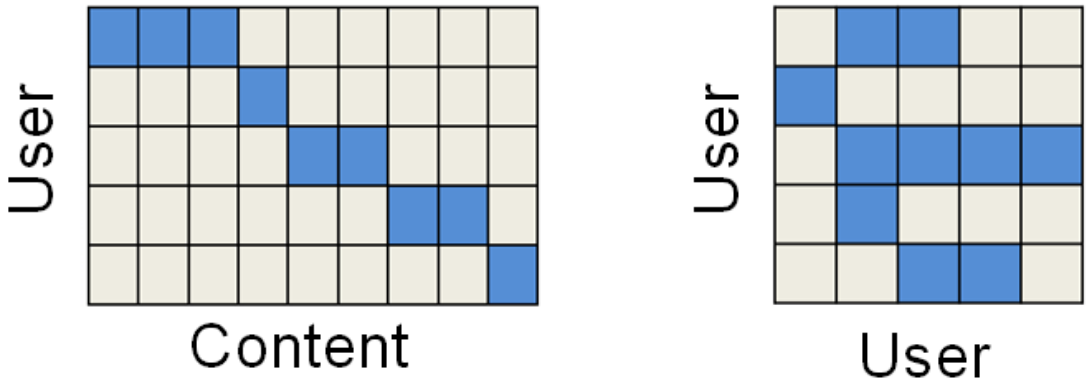


Social Spammers



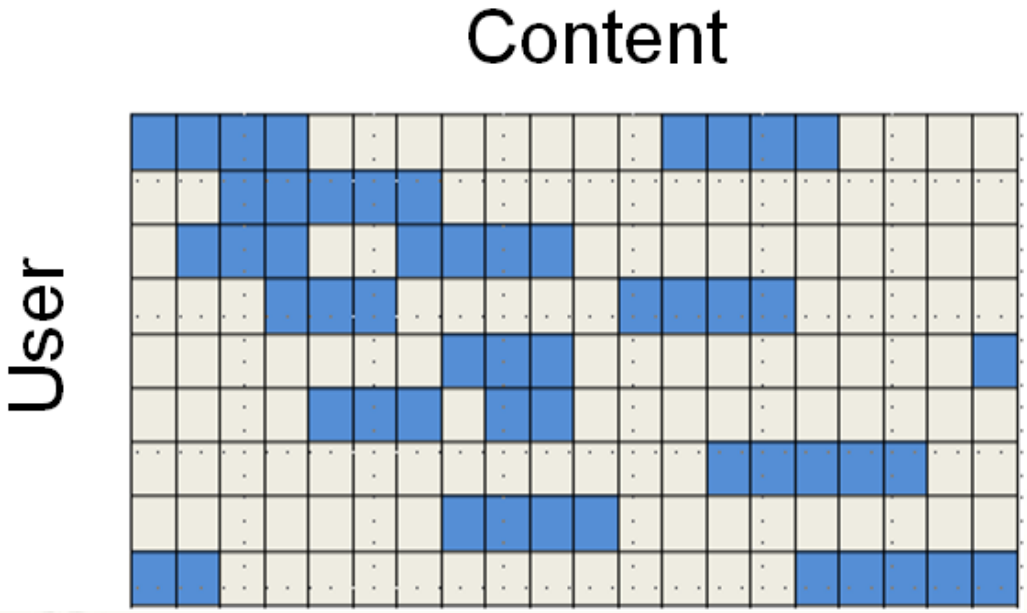


Modeling Social Networks



$$\mathcal{R}_S = \frac{1}{2} \sum_{[u,v] \in E} \boldsymbol{\pi}(u)\mathbf{P}(u,v)\|\hat{\mathbf{Y}}_u - \hat{\mathbf{Y}}_v\|^2,$$

Modeling Content Information



$$\min_{\mathbf{W}} \quad \frac{1}{2}\|\mathbf{X}^T\mathbf{W} - \mathbf{Y}\|_F^2 + \lambda_1\|\mathbf{W}\|_1,$$

Social Spammer Detection

Content Information

$$\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{X}^T\mathbf{W}-\mathbf{Y}\|_F^2+\lambda_1\|\mathbf{W}\|_1+\frac{\lambda_2}{2}\|\mathbf{W}\|_F^2 + \frac{\lambda_s}{2}tr(\mathbf{W}^T\mathbf{X}\mathcal{L}\mathbf{X}^T\mathbf{W}).$$

Social Information

Outline

- Background and Motivation
- Proposed Sociological Approach
- Experimental Evaluation
- Conclusions and Future Work

Dataset

Table 2: Summary of the Experimental Dataset

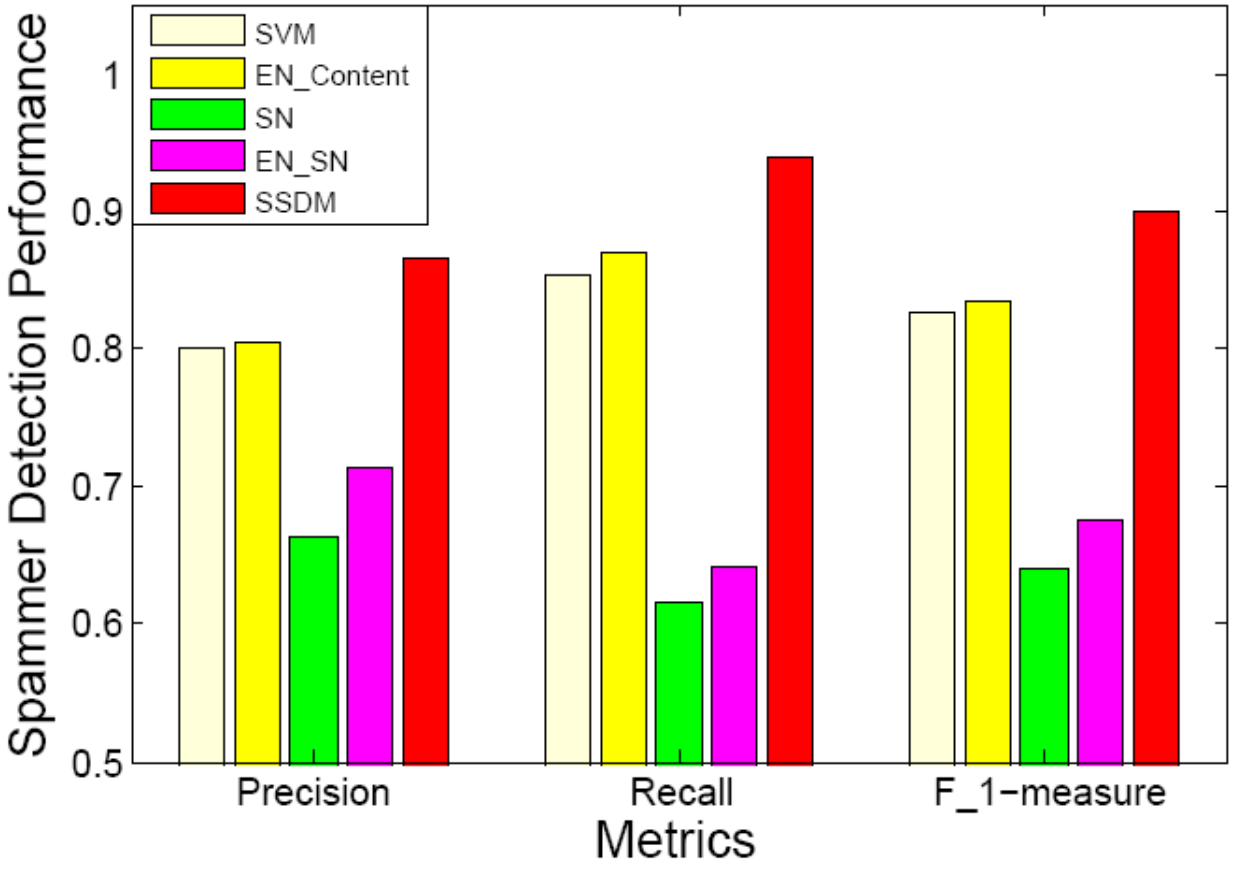
# Spammers	# Normal Users	Max Degree of Users
2,118	10,335	1,025
# Tweets	# Unigrams	Min Degree of Users
380,799	21,388	3

Social Spammer Detection Results

Table 1: Social Spammer Detection Results

	50% of the Training Data			*100% of the Training Data*		
	Precision	Recall	F_1-measure (gain)	Precision	Recall	F_1-measure (gain)
LS_Content_SN	0.786	0.843	0.813 (N.A.)	0.793	0.850	0.821 (N.A.)
EN_Content_SN	0.801	0.872	0.835 (+2.69%)	0.836	0.891	0.863 (+5.09%)
SMF_UniSN	0.804	0.889	0.845 (+3.87%)	0.844	0.915	0.878 (+6.92%)
SSDM	0.852	0.896	0.873 (+7.40%)	0.865	0.939	0.901 (+9.73%)

Social Spammer Detection Results



Outline

- Background and Motivation
- Proposed Sociological Approach
- Experimental Evaluation
- Conclusions and Future Work

Conclusion

- We formally define the problem of social spammer detection in microblogging with both network and content information

- We propose a unified model to effectively integrate both social network information and content information for the problem we are studying

- We empirically evaluate the proposed framework on a real-world Twitter dataset and elaborate the effects of each type of information for social spammer detection

Future Work

- It would be interesting to investigate other social activities, like retweet behavior and emotion status, for social spammer detection

- Sparse learning can generate a number of important textual features with the model

- Conducting behavior and linguistic analysis across social media sites to better understand motivations of the social spammers with the textual features might be also a promising direction

Questions

Acknowledgments: This work is, in part, sponsored by ONR (N000141110527) and (N000141010091). Comments and suggestions from DMML members and reviewers are greatly appreciated.