Exploiting Social Relations for Sentiment Analysis in Microblogging

Xia Hu[1], Lei Tang[2], Jiliang Tang[1], Huan Liu[1]

[1] Arizona State University

[2]@WalmartLabs

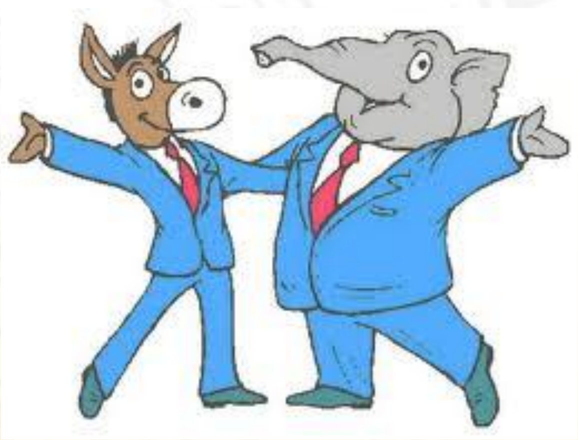
ACM International Conference on Web Search and Data Mining (***WSDM 2013***)

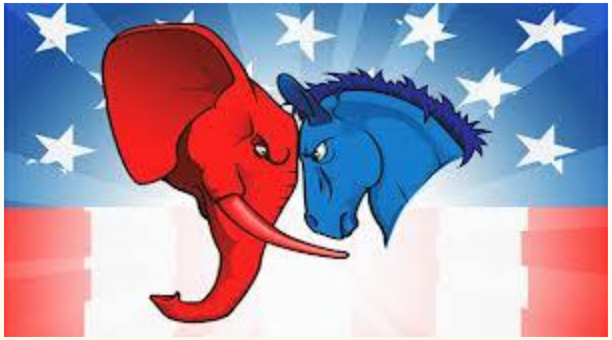
Outline

- Motivation
- Sentiment Analysis with Social Relations
- Experimental Evaluation
- Conclusions and Future Work

Why Sentiment Analysis in Microblogging?

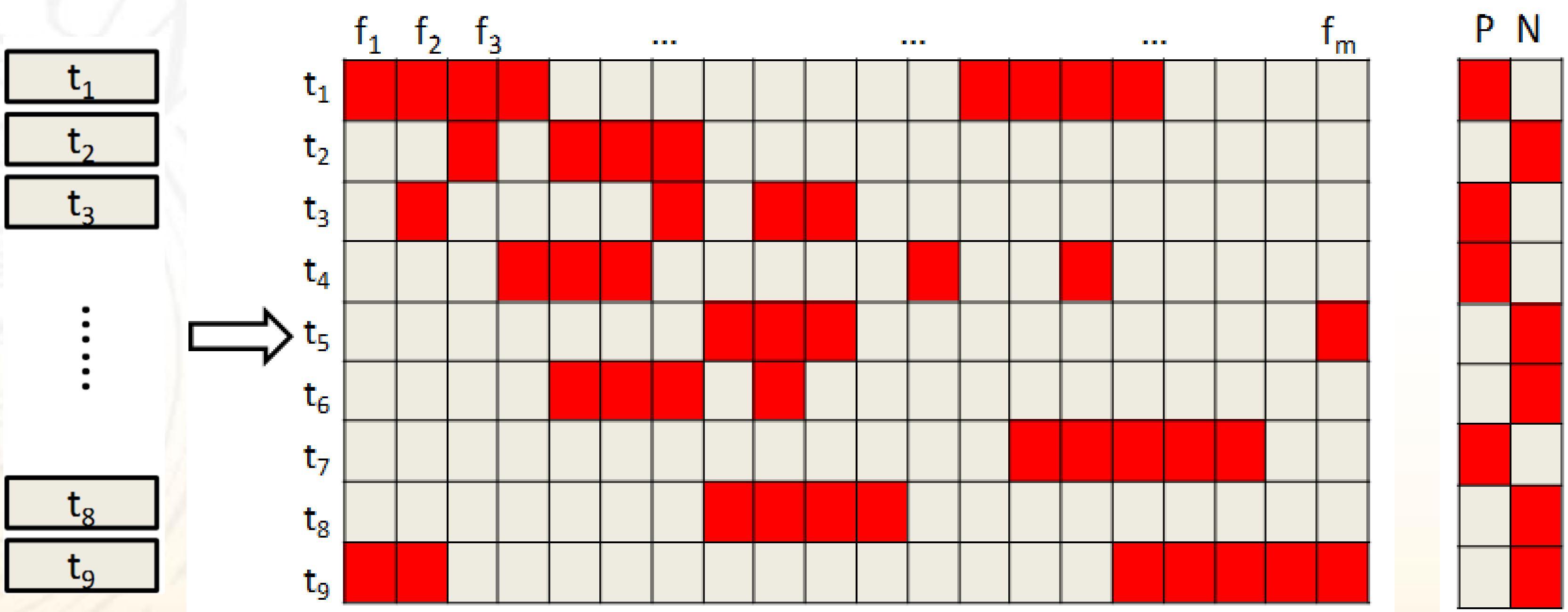
Microblogging has become a popular form of social media, through which users can easily generate comments on breaking news, public events, or products.





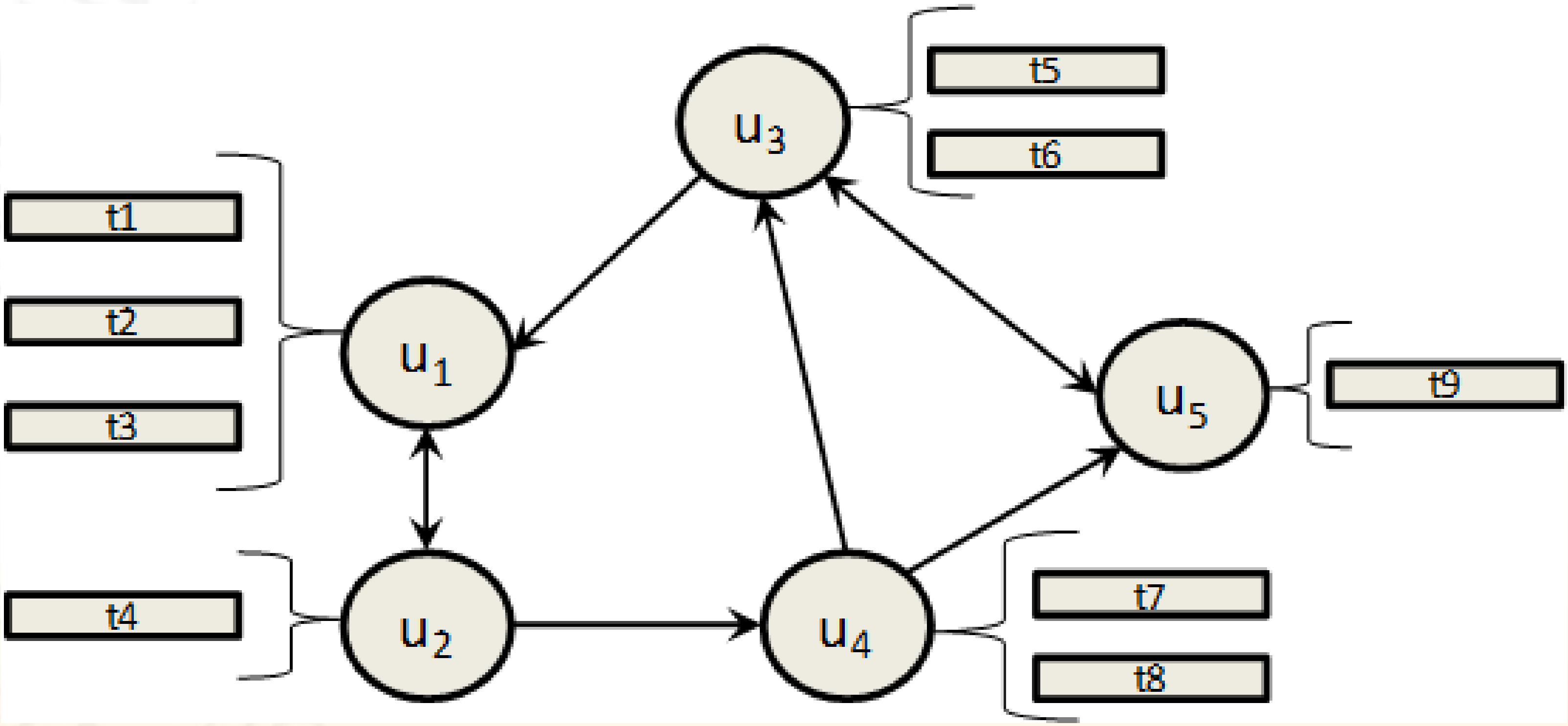
Traditional Data

Pure Textual Content

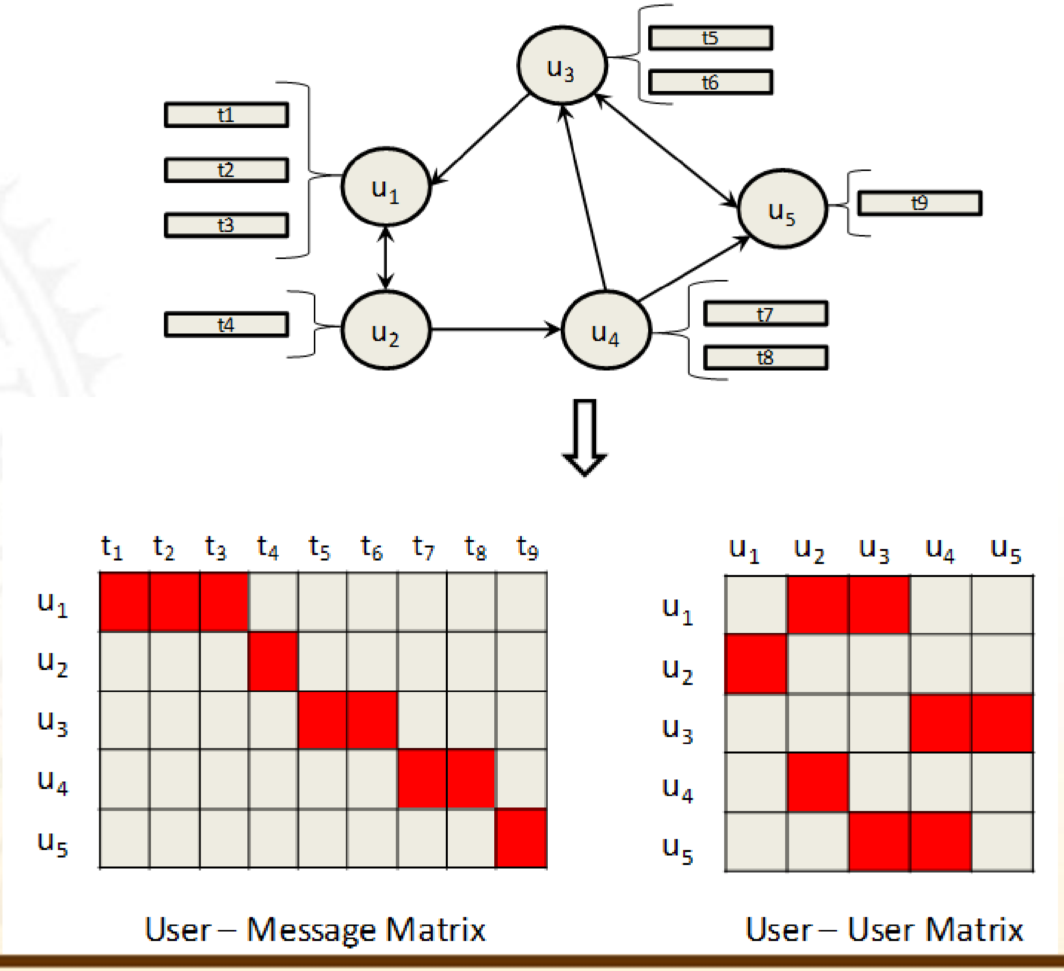


Message – Feature Matrix

Microblogging Data



Microblogging Data



Social Theories

Sentiment Consistency

suggests that the sentiments of two messages posted by the same user are more likely to be consistent than those of two random messages.

Emotional Contagion

reveals that the sentiments of two messages posted by friends are more likely to be similar than those of two random messages.

Hypothesis Testing (I)

Datasets

Table 1: Statistics of the Datasets

	STS	*OMD*
# of Tweets	22,262	1,827
# of Users	8,467	735
Max Degree of the Users	897	138
Min Degree of the Users	1	1
Avg. Tweets per User	2.63	2.49

Hypothesis Testing (II)

Sentiment Difference Score

$$\mathbf{T}_{ij} = ||\mathbf{y}_i - \mathbf{y}_j||_2$$

Verifying Sentiment Consistency

$$H_0 : \mathbf{sc}_t = \mathbf{sc}_r$$
$$H_1 : \mathbf{sc}_t < \mathbf{sc}_r$$

Hypothesis Testing (III)

Sentiment Difference Score

$$\mathbf{T}_{ij} = ||\mathbf{y}_i - \mathbf{y}_j||_2$$

Verifying Sentiment Contagion

$$H_0 \; : \; \mathbf{ec}_t \; = \; \mathbf{ec}_r$$

$$H_1 \; : \; \mathbf{ec}_t \; < \; \mathbf{ec}_r$$

Outline

- Motivation
- Sentiment Analysis with Social Relations
- Experimental Evaluation
- Conclusions and Future Work

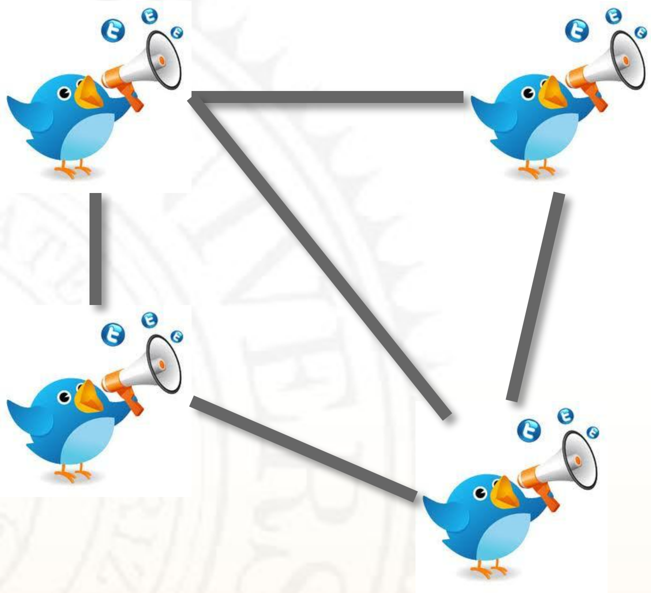
Sentiment Analysis

Sentiment (TWEET)

= $\boxed{\text{Coefficients}}$ × FeatureVector(TWEET)

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

Modeling Social Relations



$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}\|\hat{\mathbf{Y}}_{i*}-\hat{\mathbf{Y}}_{j*}\|^2$$

$$= \quad tr(\mathbf{X}^T\mathbf{W}\mathcal{L}\mathbf{X}^T\mathbf{W}),$$

Sentiment Analysis with Social Relations

Sentiment (TWEET)

= $\boxed{\text{Coefficients}}$ × FeatureVector(TWEET)

$$\min_{\mathbf{W}} \boxed{\frac{1}{2}||\mathbf{X}^T\mathbf{W} - \mathbf{Y}||_F^2} + \boxed{\frac{\alpha}{2}||\mathbf{W}^T\mathbf{X}\mathcal{L}^{\frac{1}{2}}||_F^2}$$

Textual Information

Social Relations

A Sparse Formulation

Sentiment (TWEET)

$=\boxed{\text{Coefficients}}$ × FeatureVector(TWEET)

Textual Information

Social Relations

$$\min_{\mathbf{W}} \boxed{\frac{1}{2}||\mathbf{X}^T\mathbf{W} - \mathbf{Y}||_F^2} + \boxed{\frac{\alpha}{2}||\mathbf{W}^T\mathbf{X}\mathcal{L}^{\frac{1}{2}}||_F^2}$$

$$\boxed{+\ \beta||\mathbf{W}||_1,}$$

Sparse Regularization

Outline

- Motivation
- Sentiment Analysis with Social Relations
- Experimental Evaluation
- Conclusions and Future Work

Experiments (I)

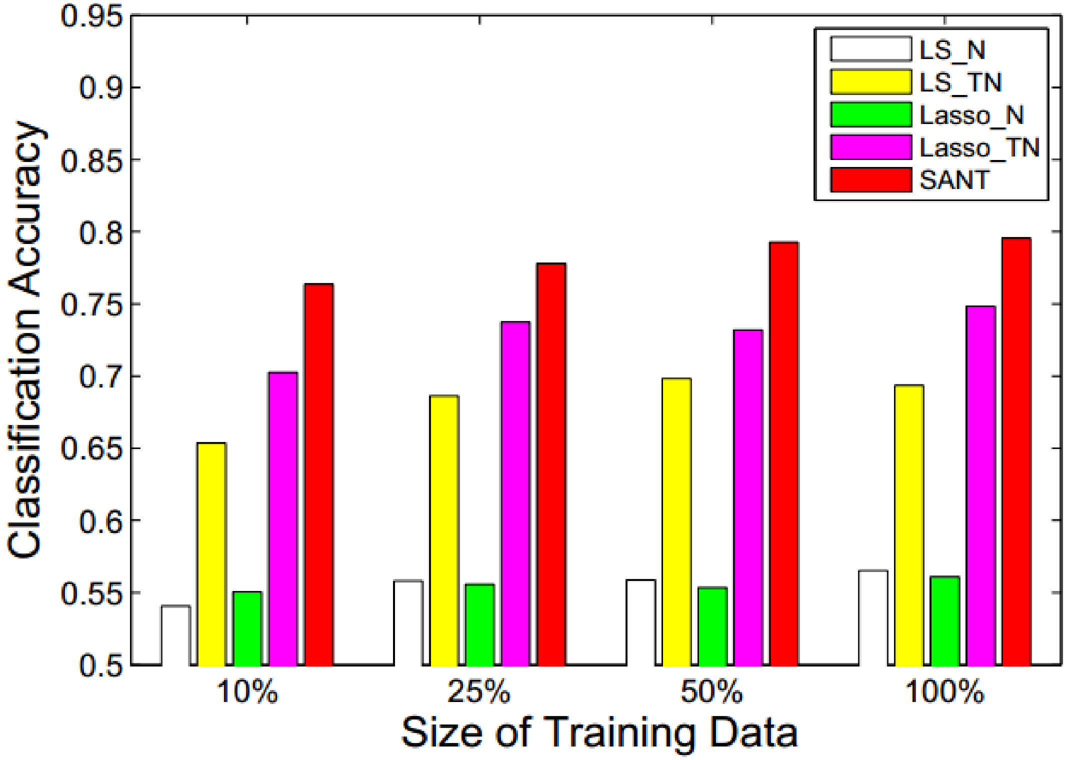
Comparison with Text-based Methods

Table 2: Sentiment Classification Accuracy on STS Dataset

	$D_{10\%}$ (gain)	$D_{25\%}$ (gain)	$D_{50\%}$ (gain)	$D_{100\%}$ (gain)
LS	0.670 (N.A.)	0.704 (N.A.)	0.720 (N.A.)	0.713 (N.A.)
Lasso	0.699 (+4.22%)	0.722 (+2.56%)	0.746 (+3.50%)	0.759 (+6.38%)
MinCuts	0.677 (+0.93%)	0.705 (+0.27%)	0.727 (+0.89%)	0.757 (+6.10%)
LexRatio	0.699 (+4.25%)	0.746 (+5.97%)	0.753 (+4.55%)	0.763 (+6.94%)
SANT	**0.764 (+13.90%)**	**0.778 (+10.56%)**	**0.793 (+10.02%)**	**0.796 (+11.52%)**

Experiments (II)

Incorporating Social Relations



Outline

- Motivation
- Sentiment Analysis with Social Relations
- Experimental Evaluation
- Conclusions and Future Work

Conclusions

•We formally define the problem of sentiment analysis with social relations in microblogging

•By verifying the existence of two social theories in microblogging, we build sentiment relations between messages via social relations

•We present a novel supervised method to tackle the high-dimensional texts by integrating sentiment relations between the texts

Future Work

- Contextual information, like spatial-temporal patterns, could be potentially useful to measure the sentiment consistency of people as well

- We can further explore how sentiments diffuse in the social network and how people's sentiments correlate with internal (their friends) and external (public events) factors

Questions

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