Social Influence and Sentiment Analysis

—From Sentiment to Emotion Analysis in Social Networks

Jie Tang

Department of Computer Science and Technology

Tsinghua University, China

Networked World

facebook

- **1.65 billion** MAU (users)
- **2.5 trillion** minutes/month

twitter

- **255 million** MAU
- **Peak: 143K** tweets/s

amazon.com

- **304 million** active users
- **14 billion** items/year

WeChat

- **QQ: 800 million** MAU
- **WeChat: 700 million** MAU

新浪微博 weibo.com

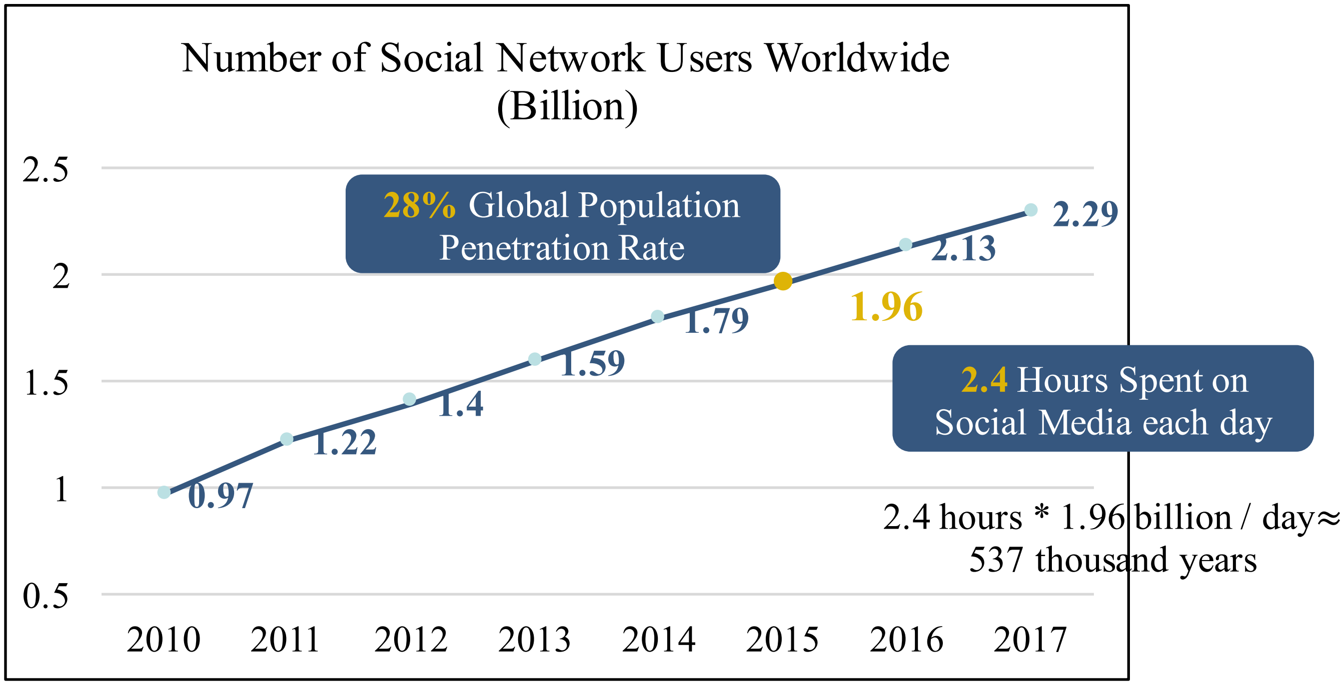
- **220 million** users
- **influencing** our daily life

Alibaba Group 阿里巴巴集团

- **710 million** trans. on 11/11
- **13.6 billion USD** in 24 hrs

The Era of Big Social Data

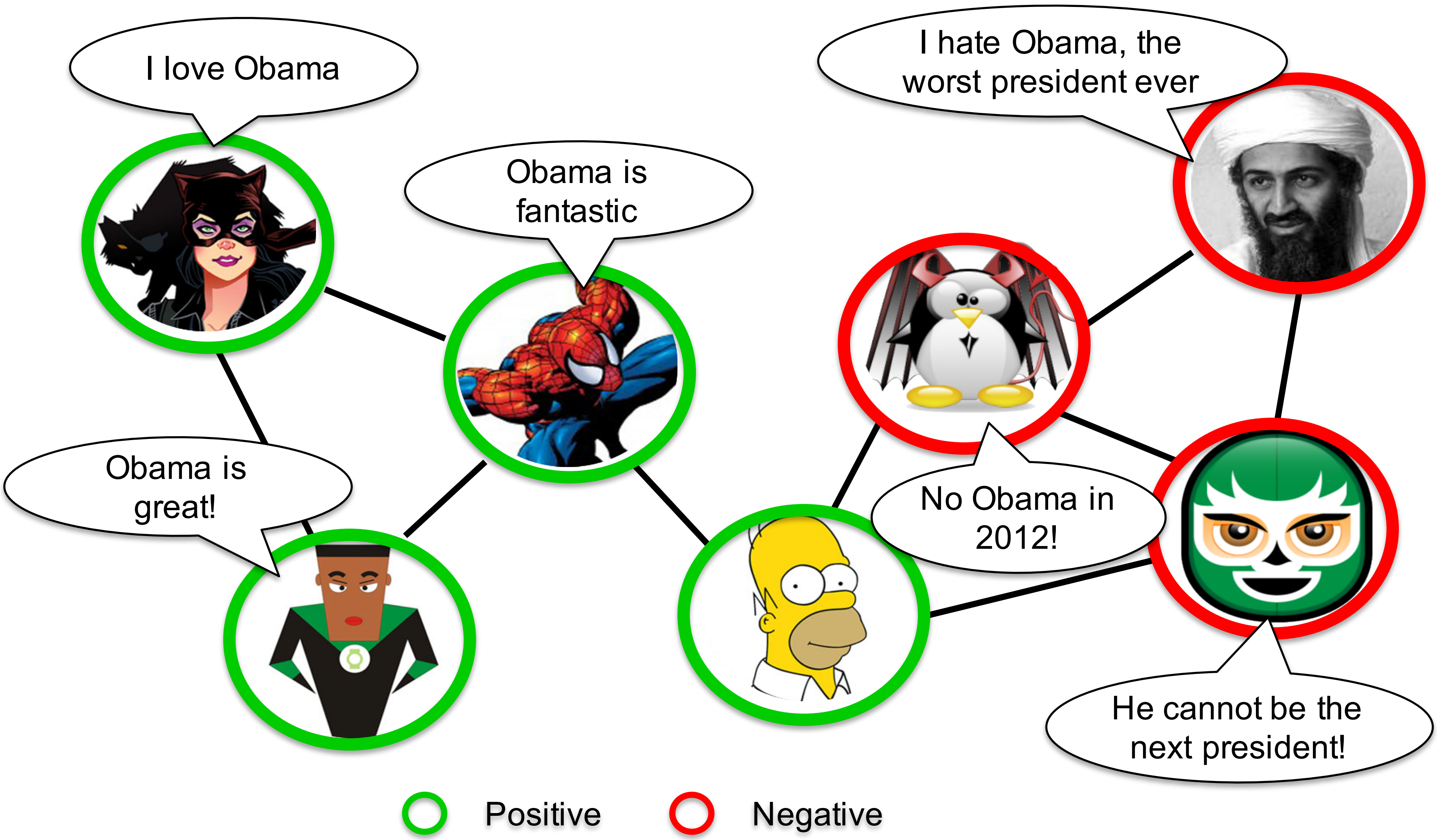
- We generate 2.5×10^{18} byte *big data* per day.



- Big social data:
  - 90% of the data was generated in the past 2 yrs
  - Mining in single data center → mining deep knowledge from multiple data sources

http://www.statista.com/statistics
http://www.globalwebindex.net/

User Opinion and Influence: “Love Obama”



Does Social Influence really matter?

- Case 1: Social influence and political mobilization[1]
  - Will online political mobilization really work?

A controlled trial (with 61M users on FB)

- Social msg group: was shown with msg that indicates one's friends who have made the votes.
- Informational msg group: was shown with msg that indicates how many other.
- Control group: did not receive any msg.

[1] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. Nature, 489:295-298, 2012.

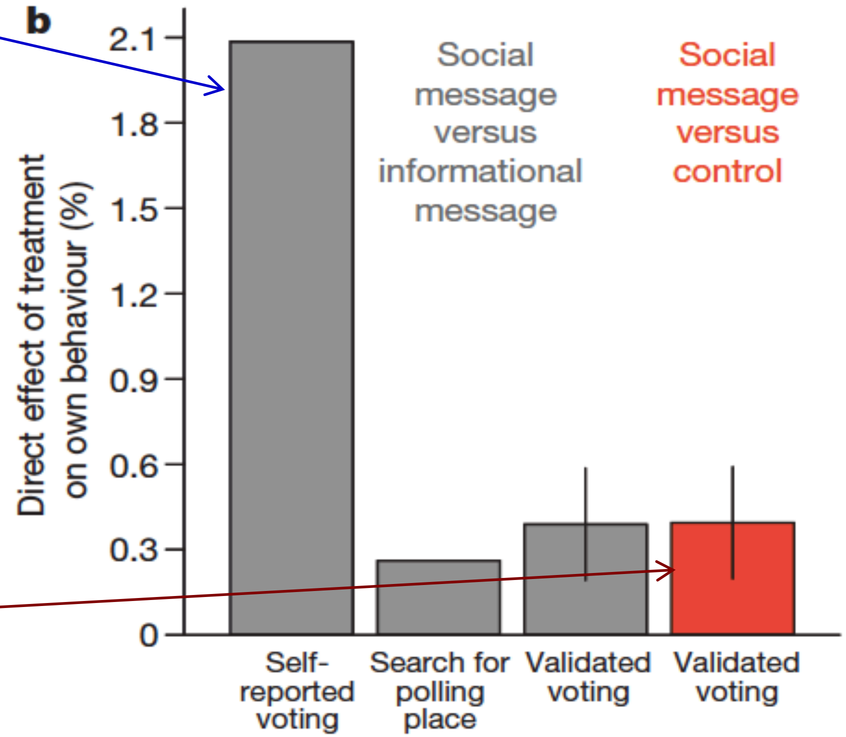
Case 1: Social Influence and Political Mobilization

Social msg group v.s.
Info msg group

Result: The former were 2.08% (*t*-test, *P*<0.01) more likely to click on the "I Voted" button

Social msg group v.s.
Control group

Result: The former were 0.39% (*t*-test, *P*=0.02) more likely to **actually vote** (via examination of public voting records)



[1] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. Nature, 489:295-298, 2012.

Twitter Data

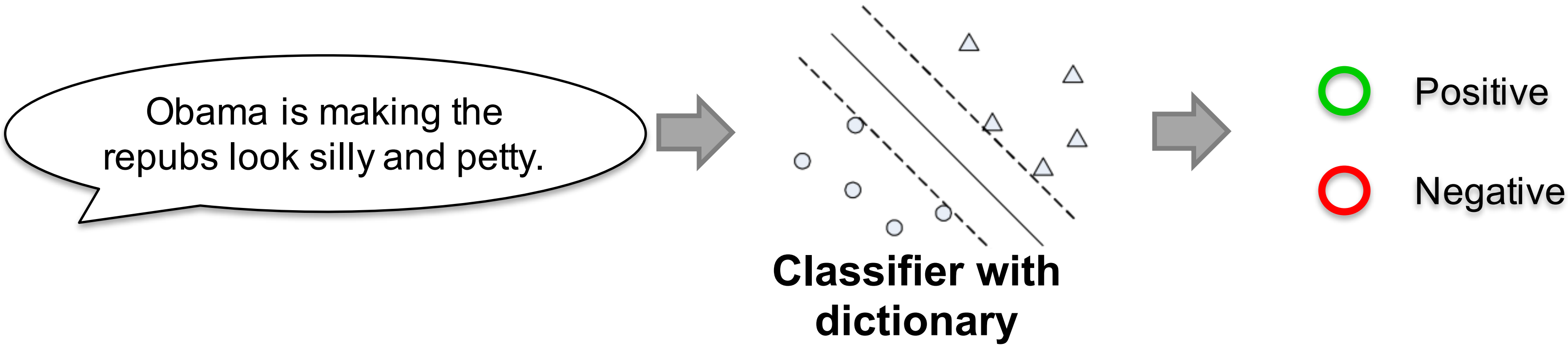
- Twitter
  - 1,414,340 users and 480,435,500 tweets
  - 274,644,047 t-follow edges and 58,387,964 @ edges

Table 1: Statistics for our main datasets.

Topic	# users	#t-follow edges		#@ edges		# on-topic tweets
		dir.	mutual	dir.	mutual	
Obama	889	7,838	2,949	2,358	302	128,373
Sarah Palin	310	1,003	264	449	60	21,571
Glenn Beck	313	486	159	148	17	12,842
Lakers	640	2,297	353	1,167	127	35.250
Fox News	231	130	32	37	5	8,479

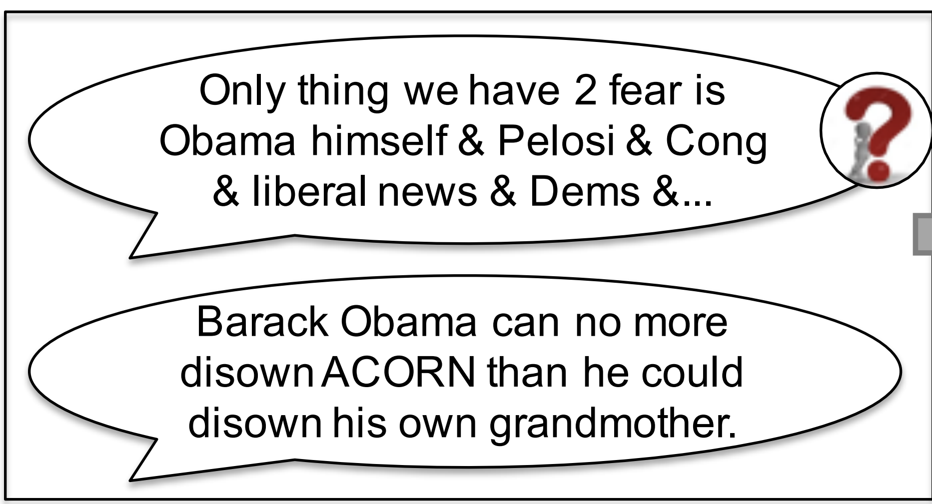
[1] Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In **KDD'11**, pages 1397–1405, 2011.

From text sentiment to user sentiment

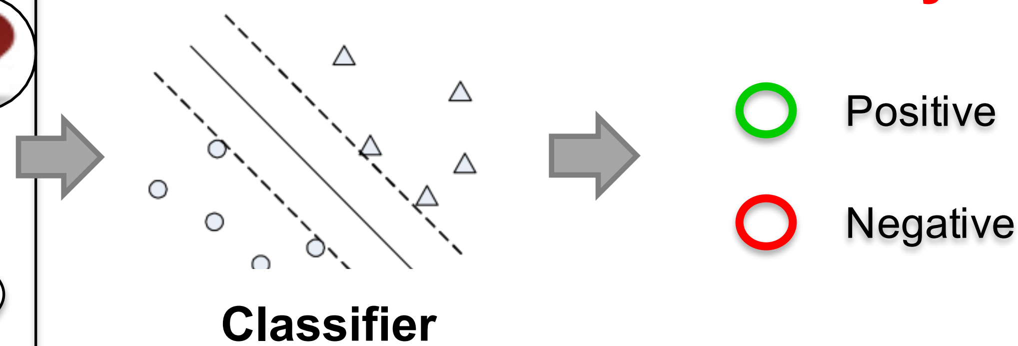


However, the social text is really short and noisy …

User A



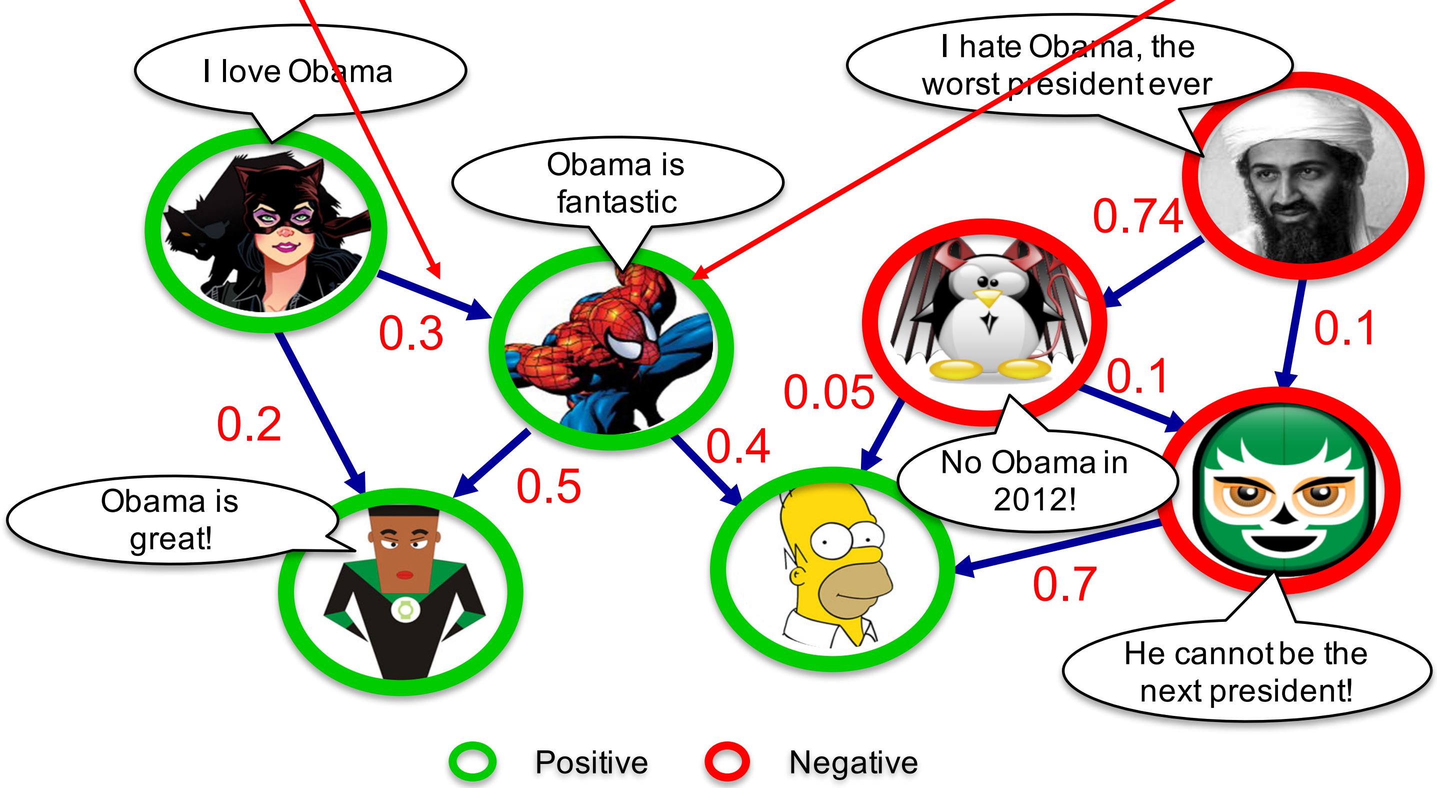
User-level Sentiment Analysis



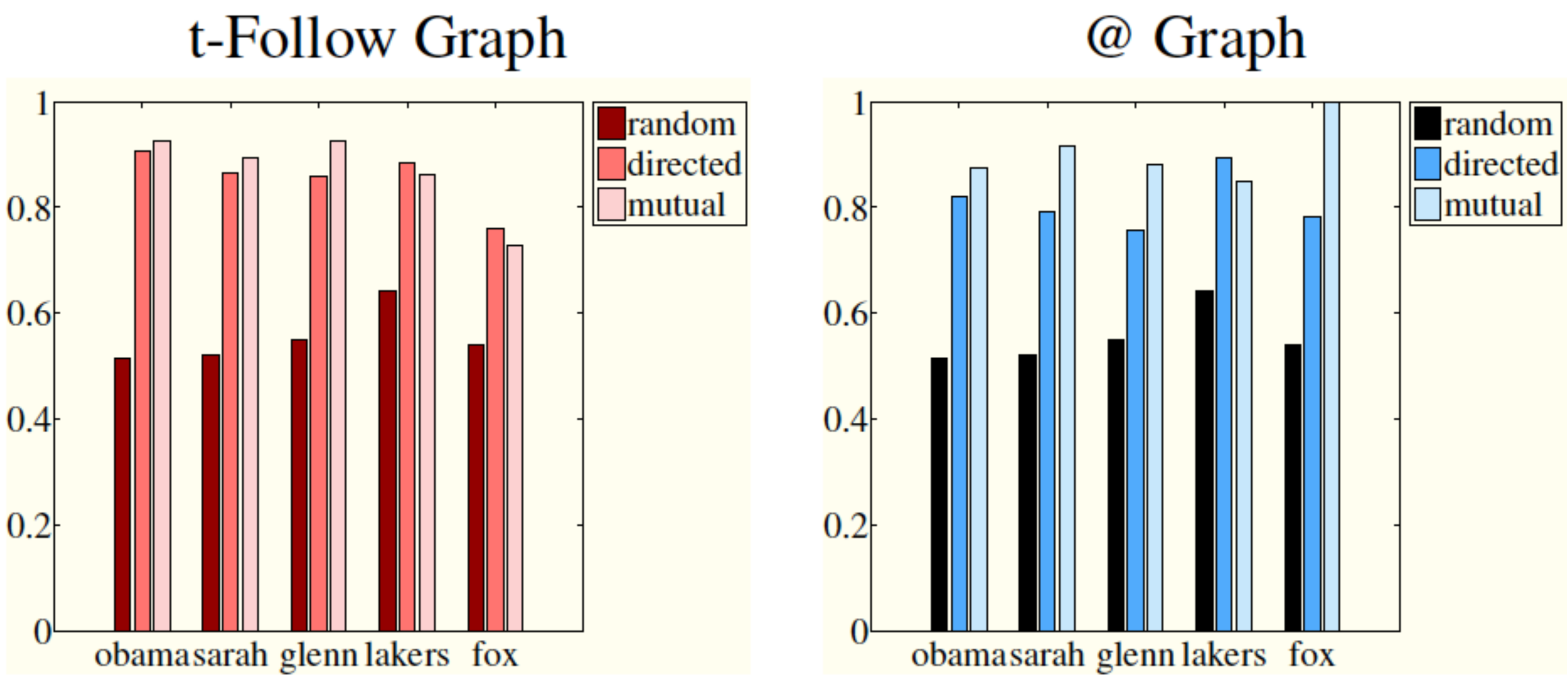
From user sentiment to network sentiment

① Who influenced who? What is the **influence probability**?

② Can we leverage the social influence to help **sentiment analysis**?



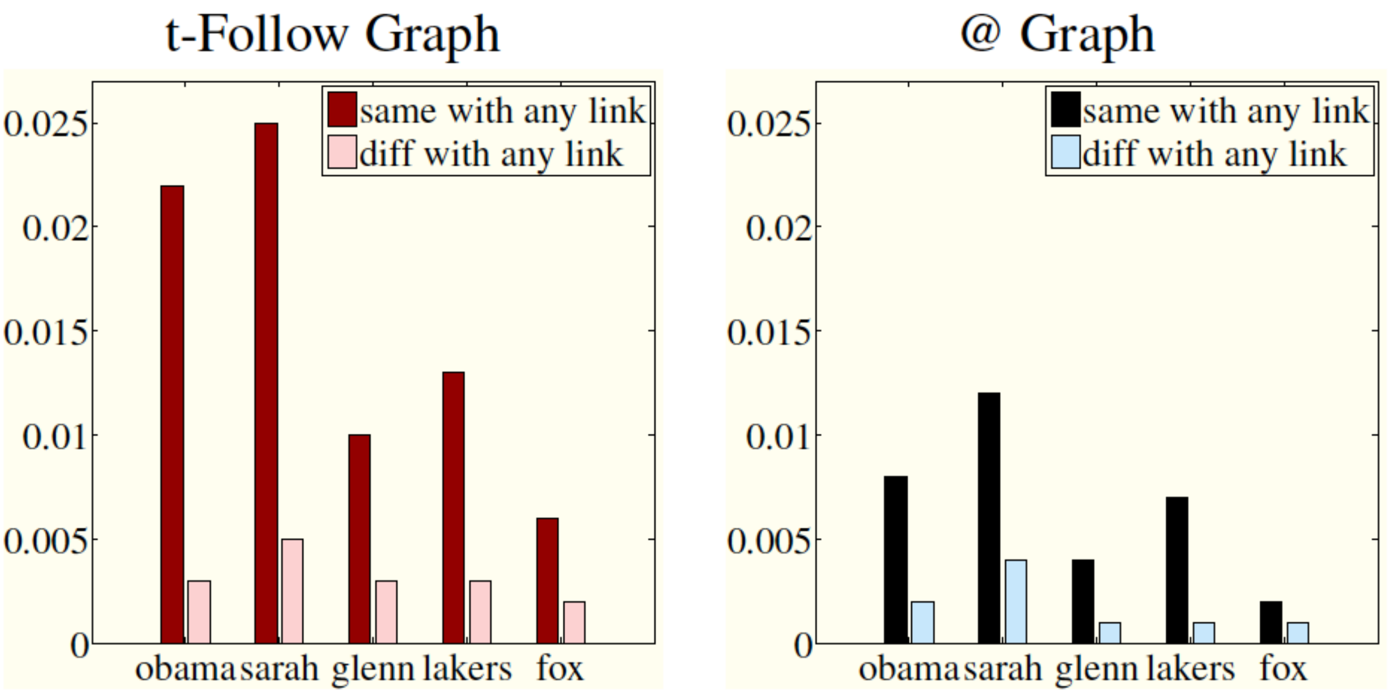
Sentiment Influence in Twitter



Shared sentiment conditioned on type of connection.
—people tend to follow the opinion of their friends

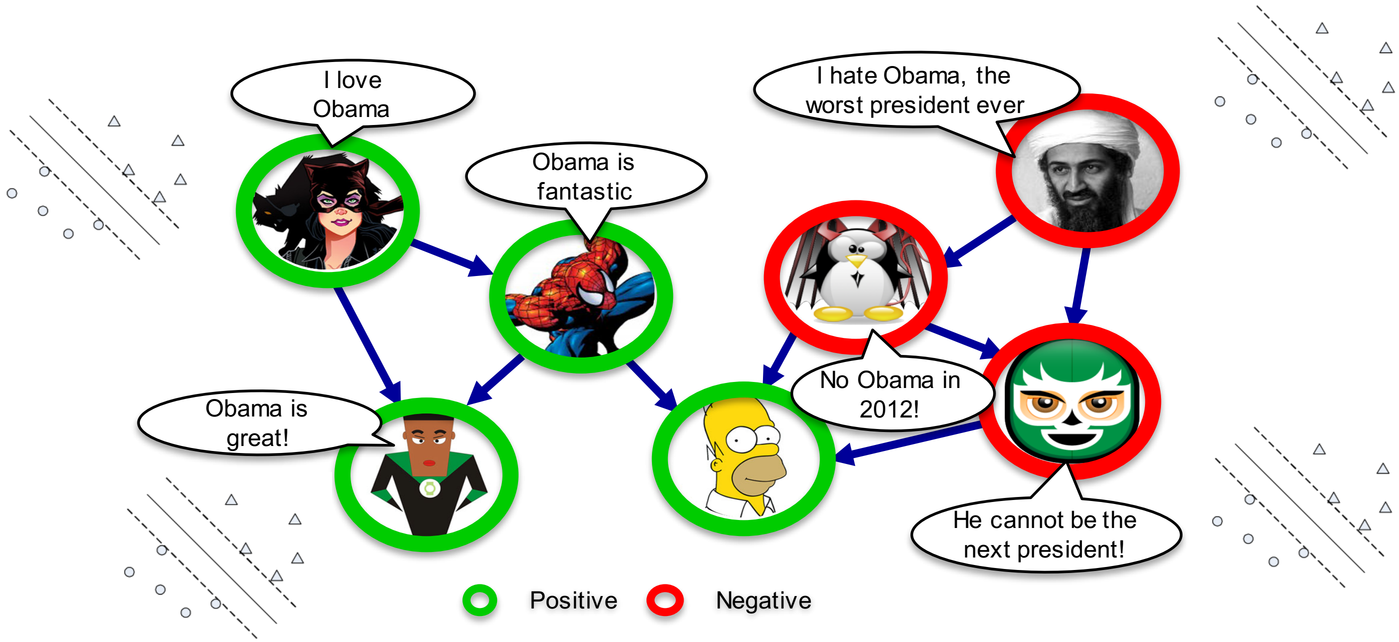
[1] Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In **KDD'11**, pages 1397–1405, 2011.

Selection



Connectedness conditioned on labels
—people tend to create relationships with other people who share the same opinion with them

Learning for network sentiment analysis

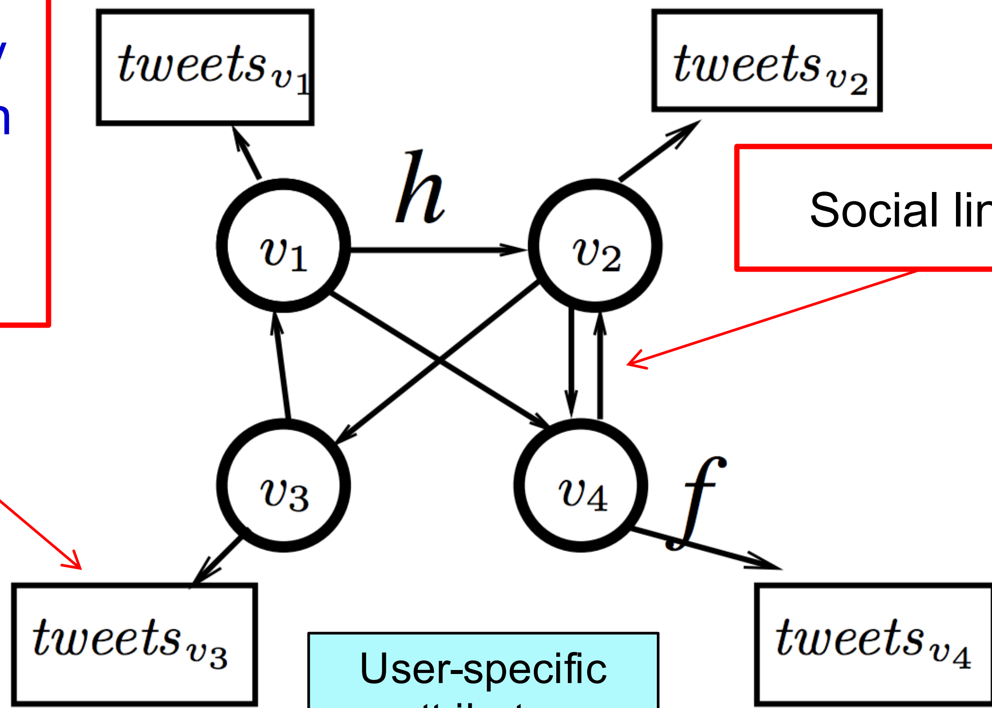


Networked Classification Model: Learning for sentiment analysis by considering the network information

Another challenge: labeled data is very limited…

Semi-supervised Factor Graph Model

Semi-FGM: learning to classify sentiments by considering both content and network structure in a semi-supervised fashion.



Social link

Tweets by user v_3

User-specific attributes

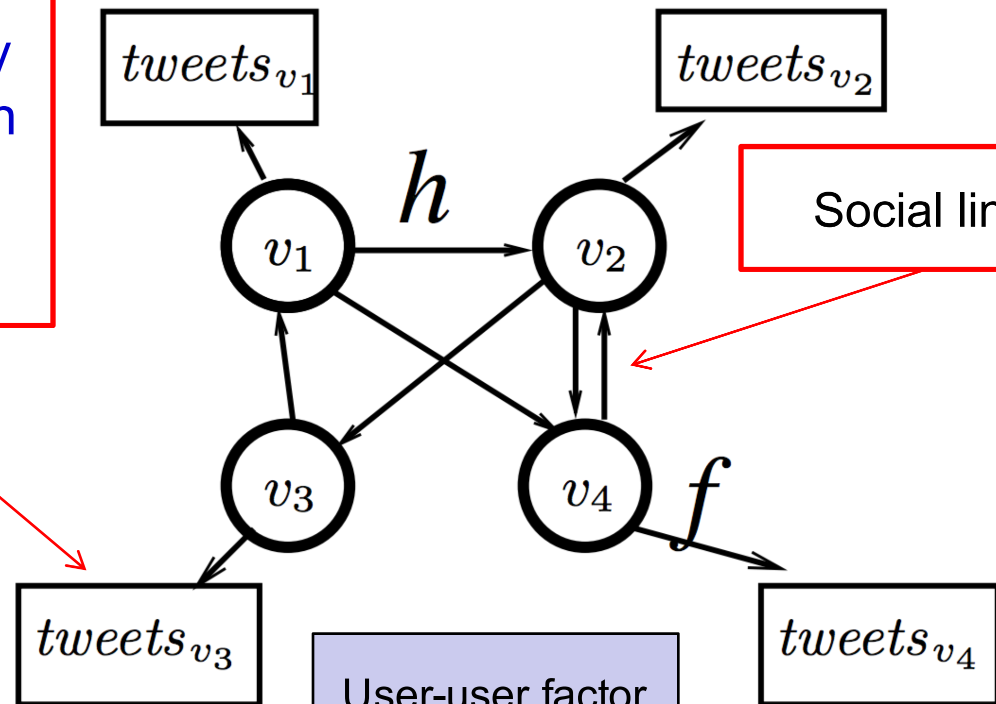
indicate our confidence level in labeled/unlabeled users

$$f_{k,\ell}(y_{v_i}, \hat{y}_t) = \begin{cases} \frac{w_{\text{labeled}}}{|tweets_{v_i}|} & y_{v_i} = k, \hat{y}_t = \ell, v_i \text{ labeled} \\ \frac{w_{\text{unlabeled}}}{|tweets_{v_i}|} & y_{v_i} = k, \hat{y}_t = \ell, v_i \text{ unlabeled} \\ 0 & \text{otherwise} \end{cases}$$

[1] Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In **KDD'11**, pages 1397–1405, 2011.

Semi-supervised Factor Graph Model

Semi-FGM: learning to classify sentiments by considering both content and network structure in a semi-supervised fashion.



Social link

Tweets by user v_3

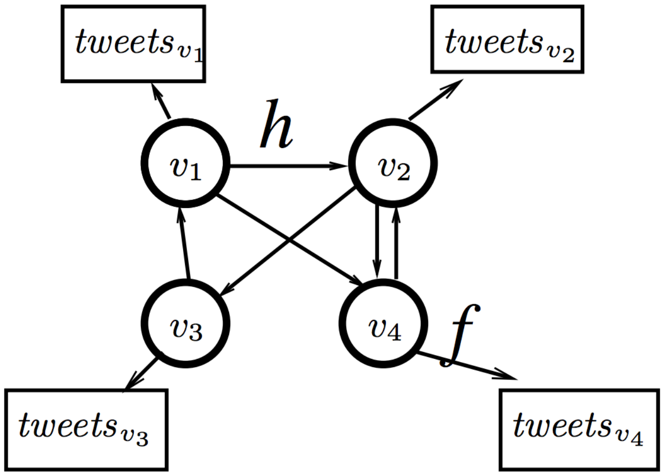
User-user factor

indicate our confidence level in network-based influence

$$h_{k,\ell}(y_{v_i}, y_{v_j}) = \begin{cases} \frac{w_{\text{relation}}}{|Neighbors_{v_i}|} & y_{v_i} = k, y_{v_j} = \ell \\ 0 & \text{otherwise} \end{cases}$$

[1] Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In **KDD’11**, pages 1397–1405, 2011.

Semi-supervised Factor Graph Model



+

$$\boxed{f_{k,\ell}(y_{v_i}, \hat{y}_t)} = \begin{cases} \frac{w_{\text{labeled}}}{|tweets_{v_i}|} & y_{v_i} = k, \hat{y}_t = \ell, v_i \text{ labeled} \\ \frac{w_{\text{unlabeled}}}{|tweets_{v_i}|} & y_{v_i} = k, \hat{y}_t = \ell, v_i \text{ unlabeled} \\ 0 & \text{otherwise} \end{cases}$$

$$\boxed{h_{k,\ell}(y_{v_i}, y_{v_j})} = \begin{cases} \frac{w_{\text{relation}}}{|Neighbors_{v_i}|} & y_{v_i} = k, y_{v_j} = \ell \\ 0 & \text{otherwise} \end{cases}$$

$$\log P(\mathbf{Y}) = \Big(\sum_{v_i \in V} \Big[\sum_{t \in tweets_{v_i}, k, \ell} \mu_{k,\ell} \boxed{f_{k,\ell}(y_{v_i}, \hat{y}_t)} + \sum_{v_j \in Neighbors_{v_i}, k, \ell} \lambda_{k,\ell} \boxed{h_{k,\ell}(y_{v_i}, y_{v_j})} \Big] \Big) - \log Z,$$

Parameter Estimation for Semi-FGM

- "NoLearning": simply use counts from the labeled subset of the data

$$\lambda_{k,\ell} := \frac{\sum_{(v_i,v_j)\in \boxed{E_{\text{labeled}}}} \boxed{I(y_{v_i} = k, y_{v_j} = \ell)}}{\sum_{(v_i,v_j)\in E_{\text{labeled}}} I(y_{v_i} = k, y_{v_j} = 1) + I(y_{v_i} = k, y_{v_j} = 0)}$$

the subset of edges in our dataset in which both endpoints are labeled

indicator function

- SampleRank ("Learning"): A sampling-based learning algorithm using Metropolis–Hastings

SampleRank ("Learning")

Input: Heterogeneous graph HG with labels on some of the user nodes, learning rate η
Output: Parameter values ϕ and full label-vector $\mathbf{Y}$

Randomly initialize $\mathbf{Y}$;
Initialize ϕ from NoLearning;
for $i := 1$ *to* Number of Steps **do**
  $\mathbf{Y}^{\mathrm{new}} := \mathsf{Sample}(\mathbf{Y})$;
  if $(\boxed{\mathsf{RelPerf}(\mathbf{Y}^{\mathrm{new}}, \mathbf{Y})} > 0$ *and* $\boxed{\mathsf{LLR}_\phi(\mathbf{Y}^{\mathrm{new}}, \mathbf{Y})} < 0)$
  //performance is better but the objective function is lower
  or $(\mathsf{RelPerf}(\mathbf{Y}^{\mathrm{new}}, \mathbf{Y}) < 0$ *and* $\mathsf{LLR}_\phi(\mathbf{Y}^{\mathrm{new}}, \mathbf{Y}) > 0)$
  //performance is worse but the objective function is higher
  then
    $\phi := \phi - \eta \nabla_\phi \mathsf{LLR}_\phi(\mathbf{Y}^{\mathrm{new}}, \mathbf{Y})$;
  end
  if *convergence* **then**
    break;
  end
  if $\mathsf{RelPerf}(\mathbf{Y}^{\mathrm{new}}, \mathbf{Y}) > 0$ **then**
    $\mathbf{Y} := \mathbf{Y}^{\mathrm{new}}$;
  end
end

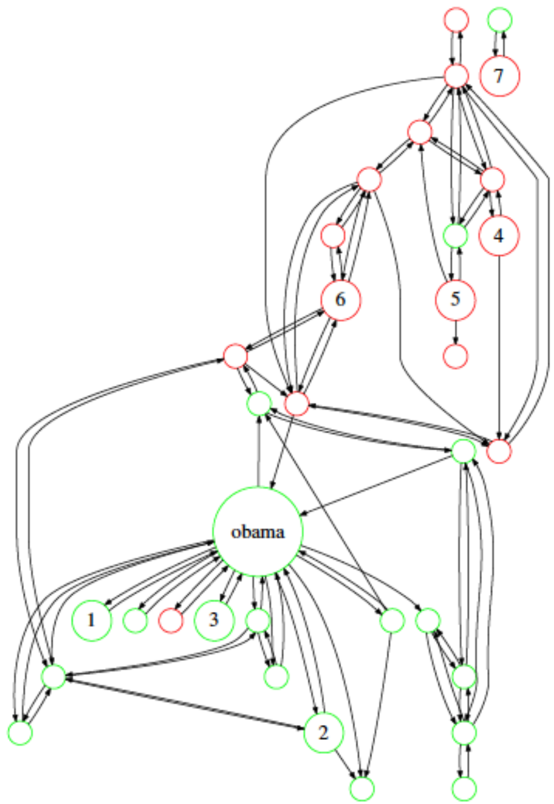
likelihood ratio of new sample Y^{new} and previous label Y for all users

Update model parameters when two results are inconsistent

Relative performance between new sample Y^{new} and previous label Y on *labeled user only*.

Results of network sentiment analysis

- Twitter
  - 1,414,340 users and 480,435,500 tweets
  - 274,644,047 t-follow edges and 58,387,964 @ edges
- Methods
  - SVM Vote
  - Semi-FGM (NoLearning)
  - Semi-FGM (SampleRank)
- Measures
  - Accuracy and Macro F1

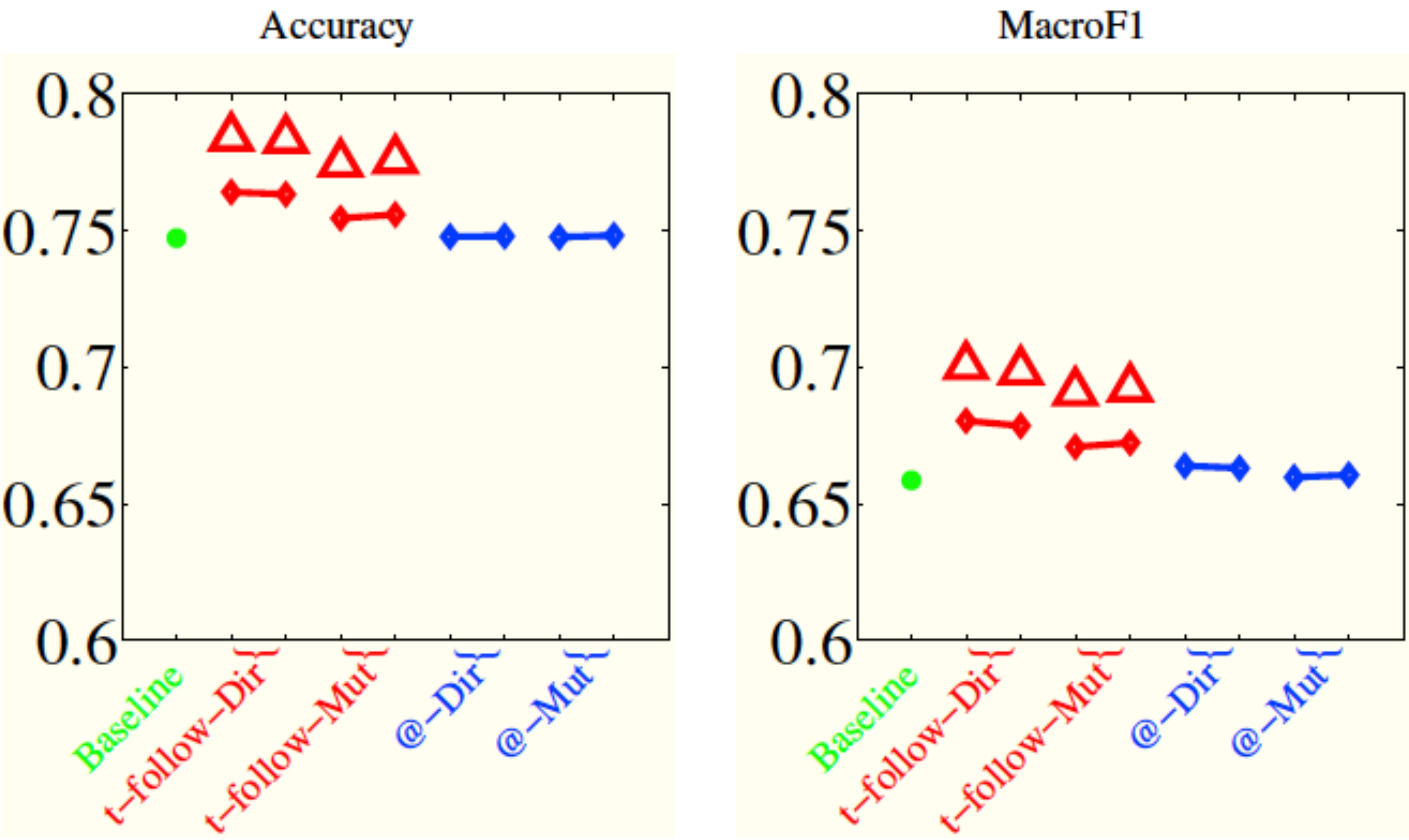


(a) Ground Truth (b) Text-Only Approach (c) Our algorithm

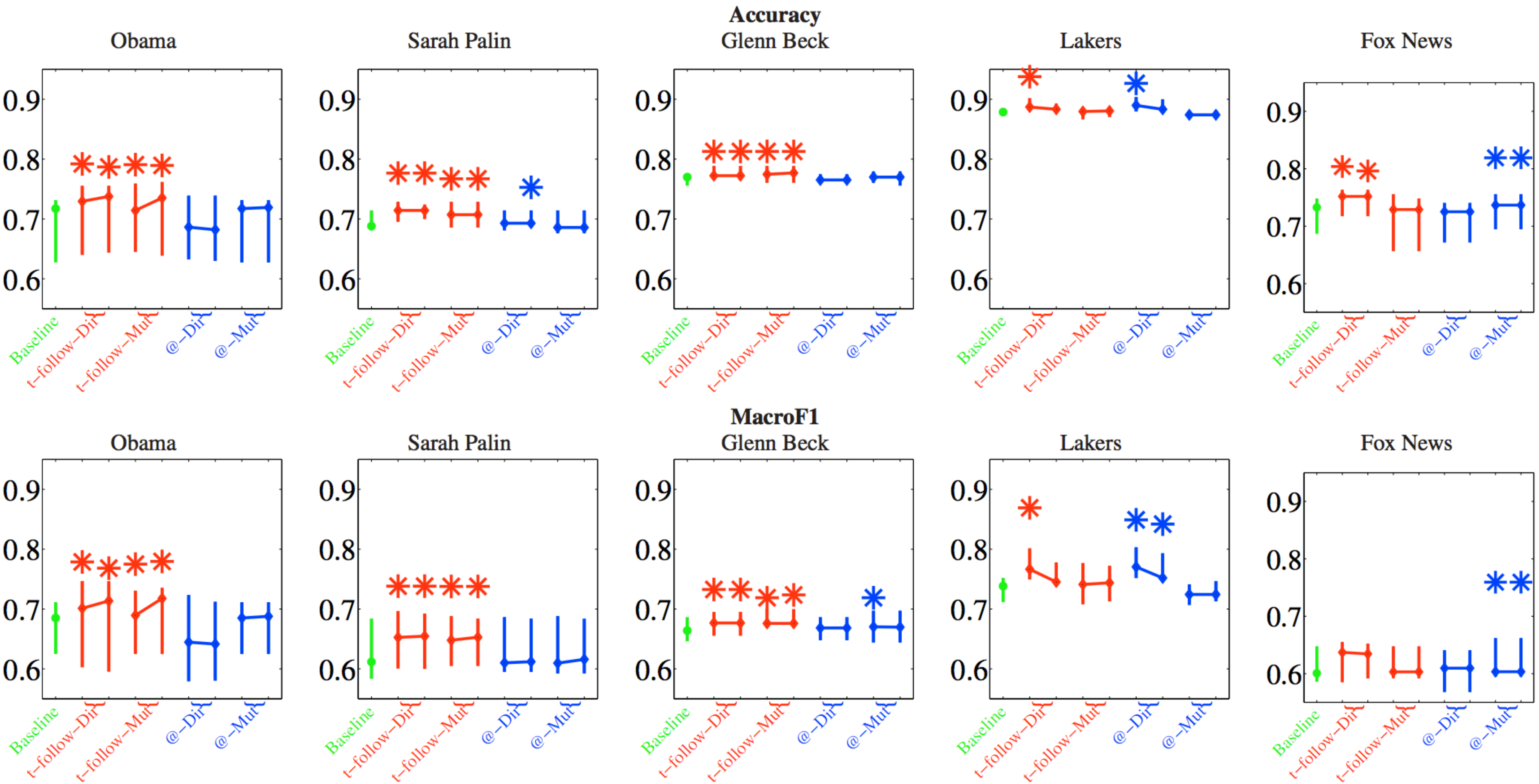
Sample tweets of users classified correctly only when network information is incorporated

User ID	SVM Vote	HGM	True	Tweet
1	NEG	POS	POS	Obama is making the repubs look silly and petty. #hrc
2	NEG	POS	POS	Is happy Obama is President
				Obama collectable http://tinyurl.com/c5u7jf
3	NEG	POS	POS	I am praying that the government is able to get health care reformed this year! President Obama seems like the ONE to get it worked out!!
				Watching House on TV. I will be turning to watch Rachel M. next. I am hoping Pres. Obama gets his budget passed. Especially Health Care!
4	POS	NEG	NEG	RT @TeaPartyProtest Only thing we have 2 fear is Obama himself & Pelosi & Cong & liberal news & Dems &... http://ow.ly/15M9Xv
				RT @GlennBeckClips: Barack Obama can no more disown ACORN than he could disown his own grandmother. #TCOT
5	POS	NEG	NEG	RT @JosephAGallant Twitlonger: Suppose I wanted to Immigrant to Mexico? A Letter to President Obama.. http://tl.gd/1kr5rh
				George Bush was and acted like a war time President. Obama is on a four year power grab and photo op. #tcot
6	POS	NEG	NEG	ObamaCare forces Americans to buy or face a fine! It is UNCONSTITUITIONAL to force us to buy obamacare. Marxist Govt. taking our Freedoms!
				Look up Chicago Climate Exchange,an organization formed years ago by Obama & his Marxist-Commie Cronies to form a profit off cap & trade.

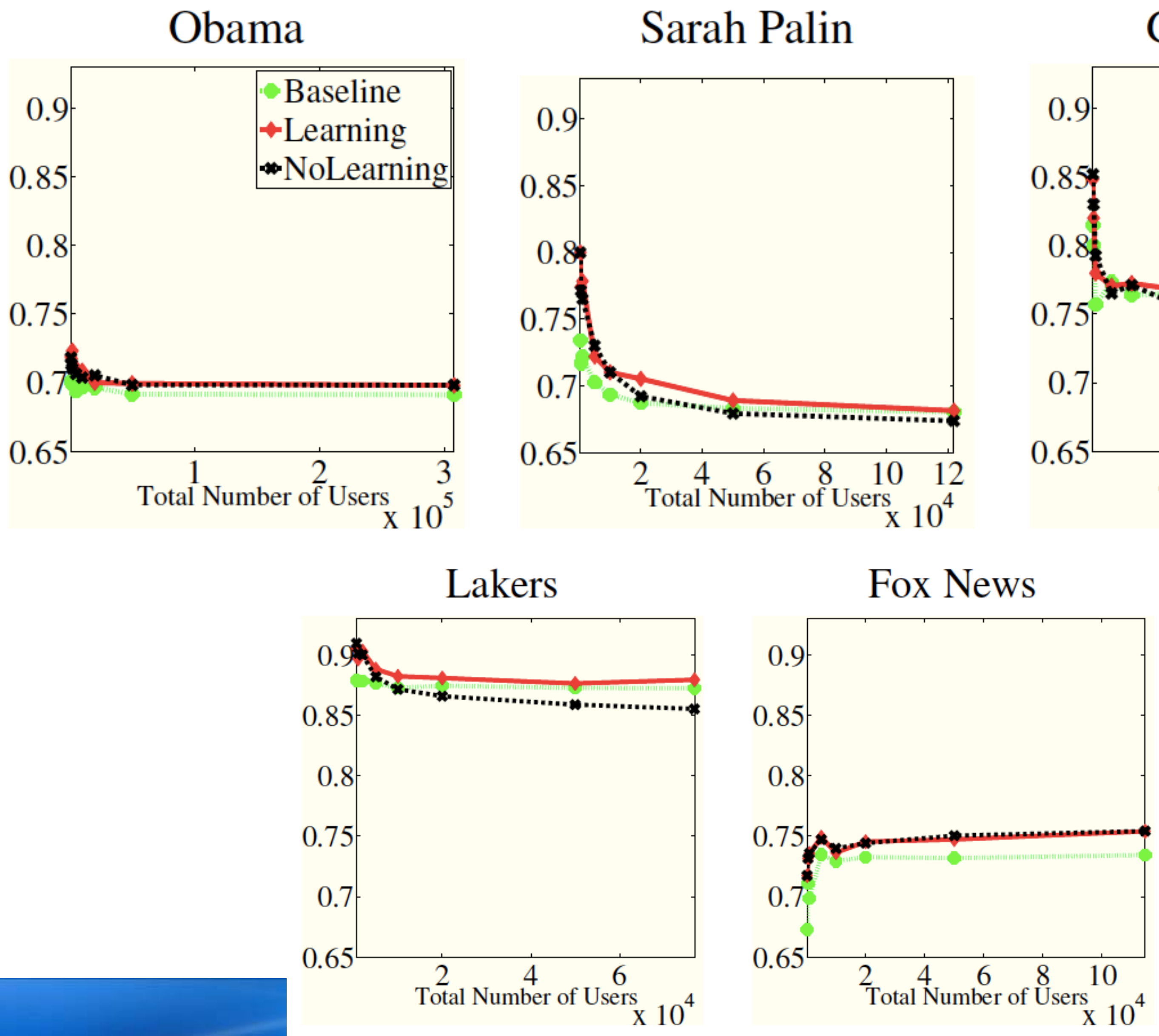
Performance



Performance Analysis in Different Topics

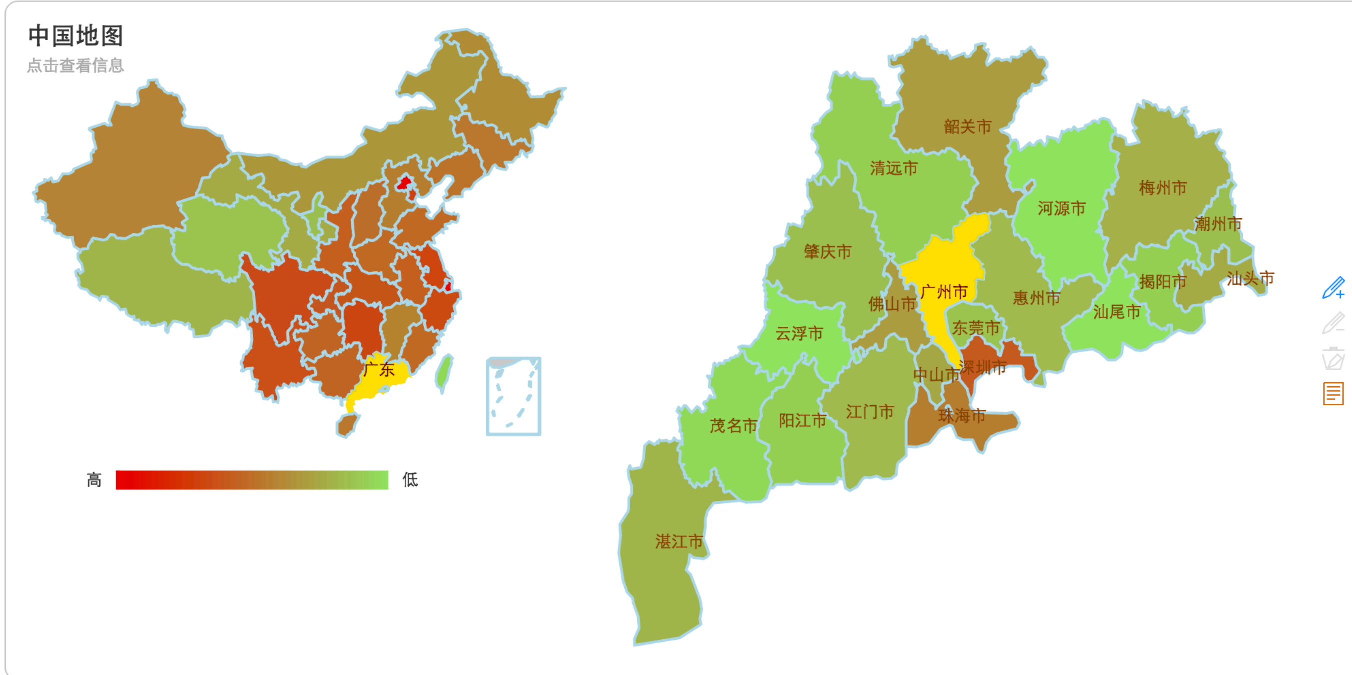


Results of Different Learning Algorithms



Twitter to Weibo

Activity Net 要搜索的活动 搜索 登录 菜单



广州市 幸福指数
男性 29.03 | 女性 31.32

省份排名 积极词 消极词

省份	幸福指数
上海	40.76491
北京	40.27117
天津	33.05676
江苏	32.77226
湖南	32.52948
浙江	32.24124
四川	31.95578

用这张相最好//@我爱黄世仁-奎:好。

转发微博

08月04日 15:36 转发(0) | 评论(3)

我知道是我自己不够努力，所以才会是这样的结果，可是为什么上天不给我点点的奖励呢，我一直好好地做人，做好好的人，为什么不能让我好运点呢?

03月13日 21:49 转发(0) | 评论(0)

我参与了@和谐宿舍-快乐心理 发起的投票【最佳展示宿舍】，我投给了"生技春霖2519 神迹519"这个选项。你也快来表态吧：http://t.cn/zOuIWIq

05月21日 23:42 转发(0) | 评论(1)

We have a picture of sentiment analysis in social networks…

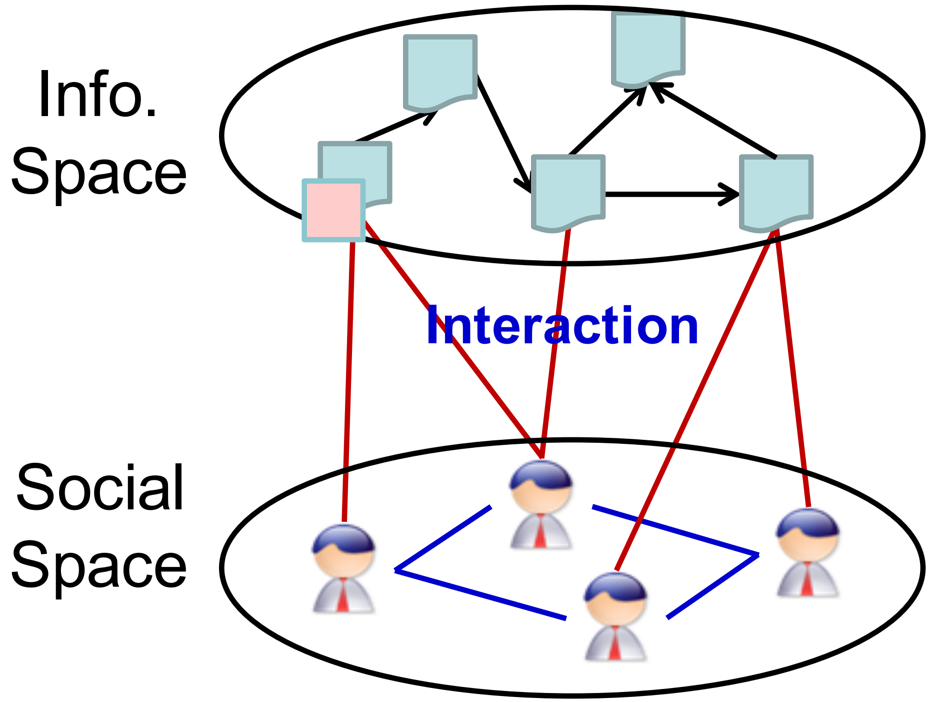
- From text sentiment to user sentiment
- From user sentiment to network sentiment
- Challenges:
  - Short text and noisy data
  - Limited labeled data
  - Networked user sentiments
- Proposal of a Semi-supervised Factor Graph Model (Semi-FGM) to learn to classify sentiments by considering both content and network structure

Now, let us think…

- What are the fundamental factors behind
  - What is behind the network of social users?
  - What is behind the sentiment of social users?

Well, what is the fundamental factor…

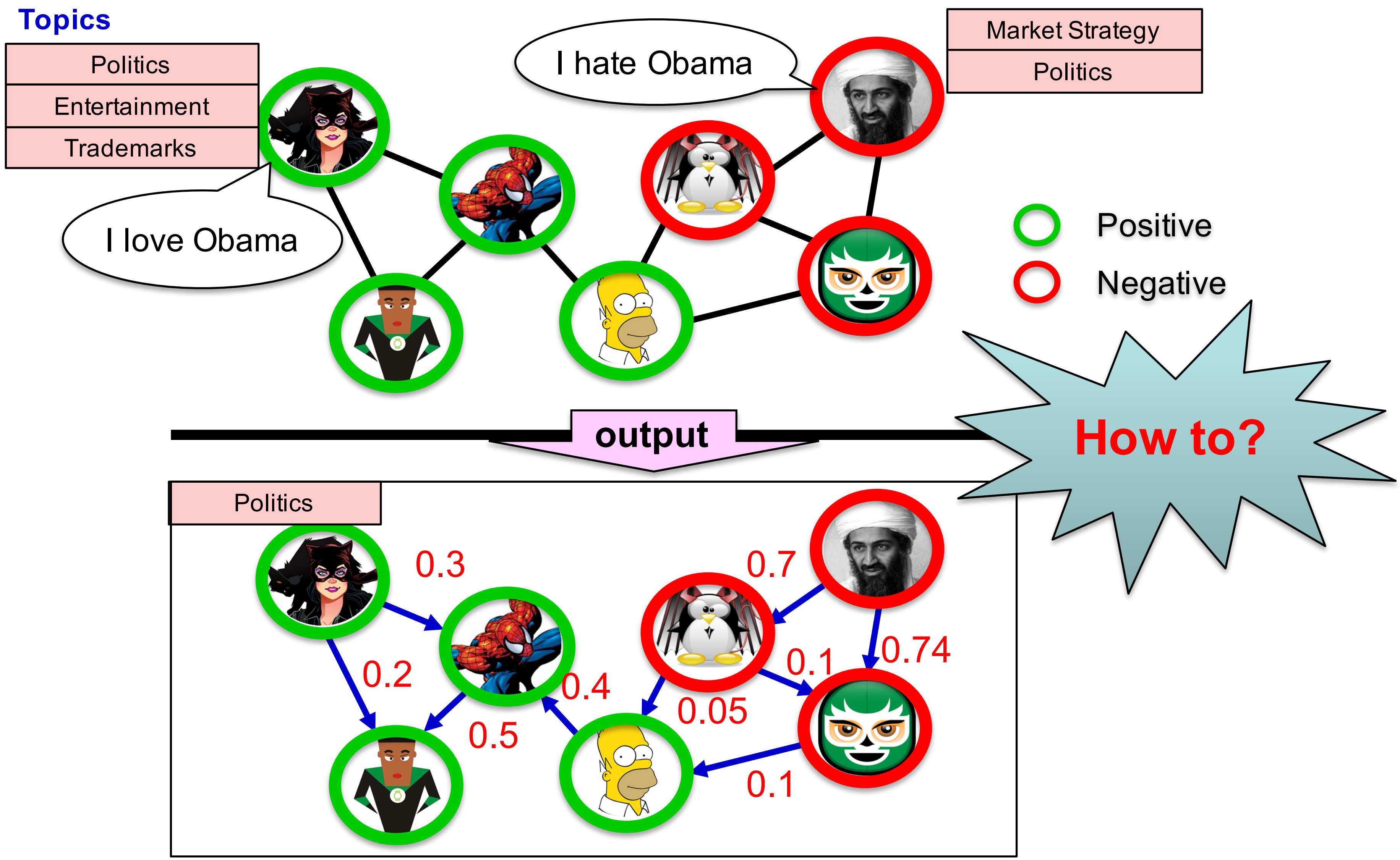
Info. Space vs. Social Space



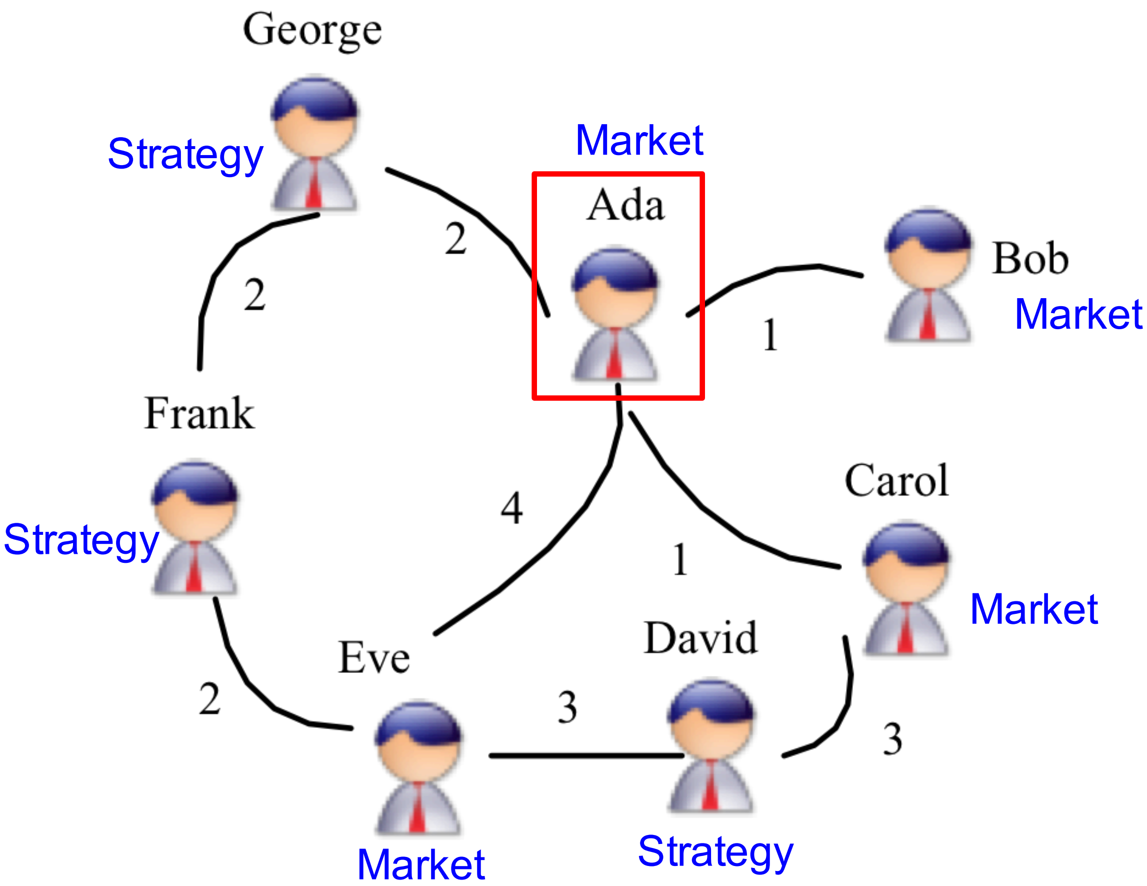
From the social network research perspective, what are the fundamental factors behind?

Understanding the
mechanism of interaction dynamics

Topic-based Social Influence Analysis



The Solution: Topical Affinity Propagation



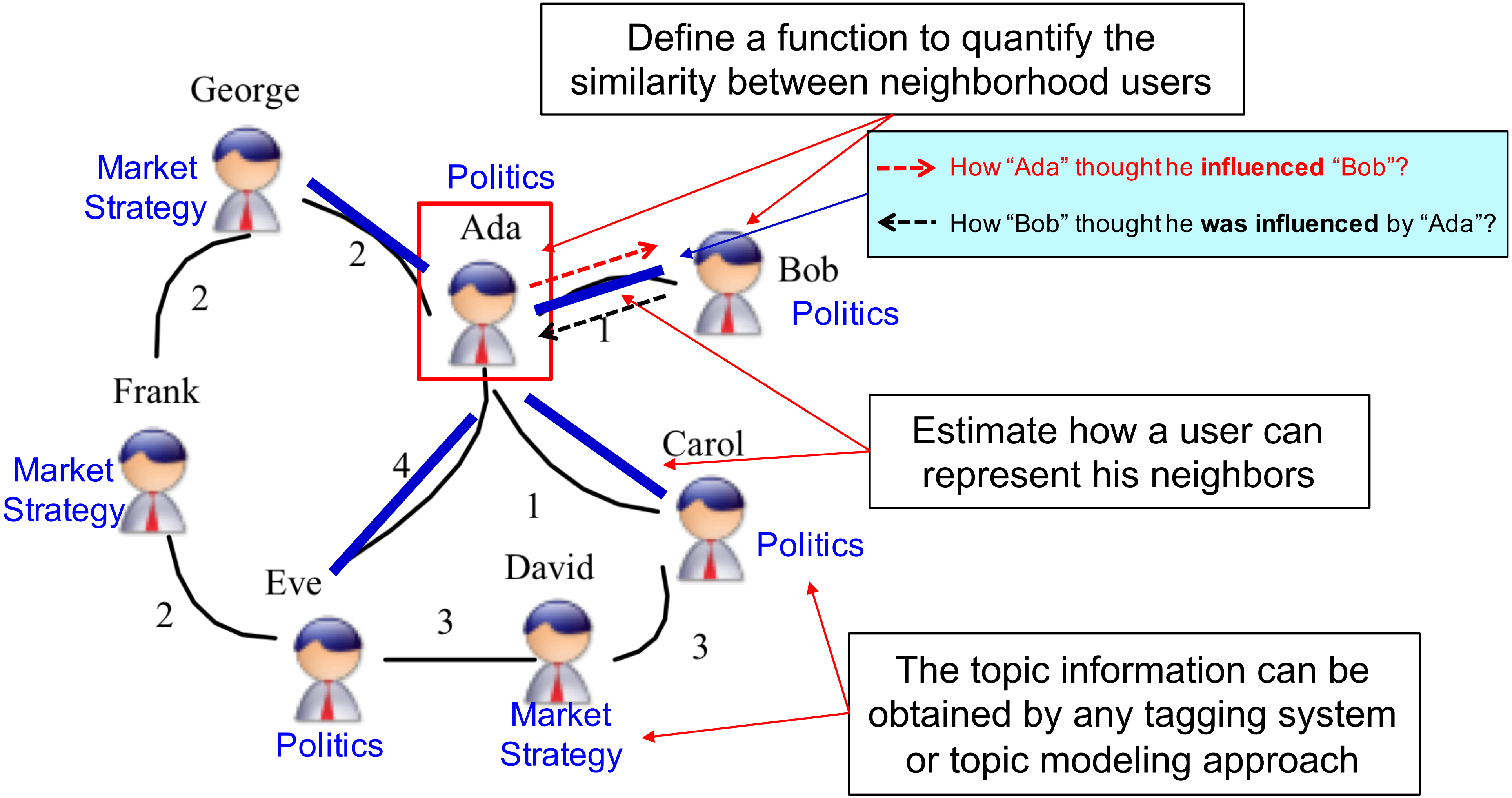
Basic Idea:

If a user is located in the center of a "Market" community, and is "similar" to the other users, then she/he would have a strong influence on the other users.

—Homophily theory

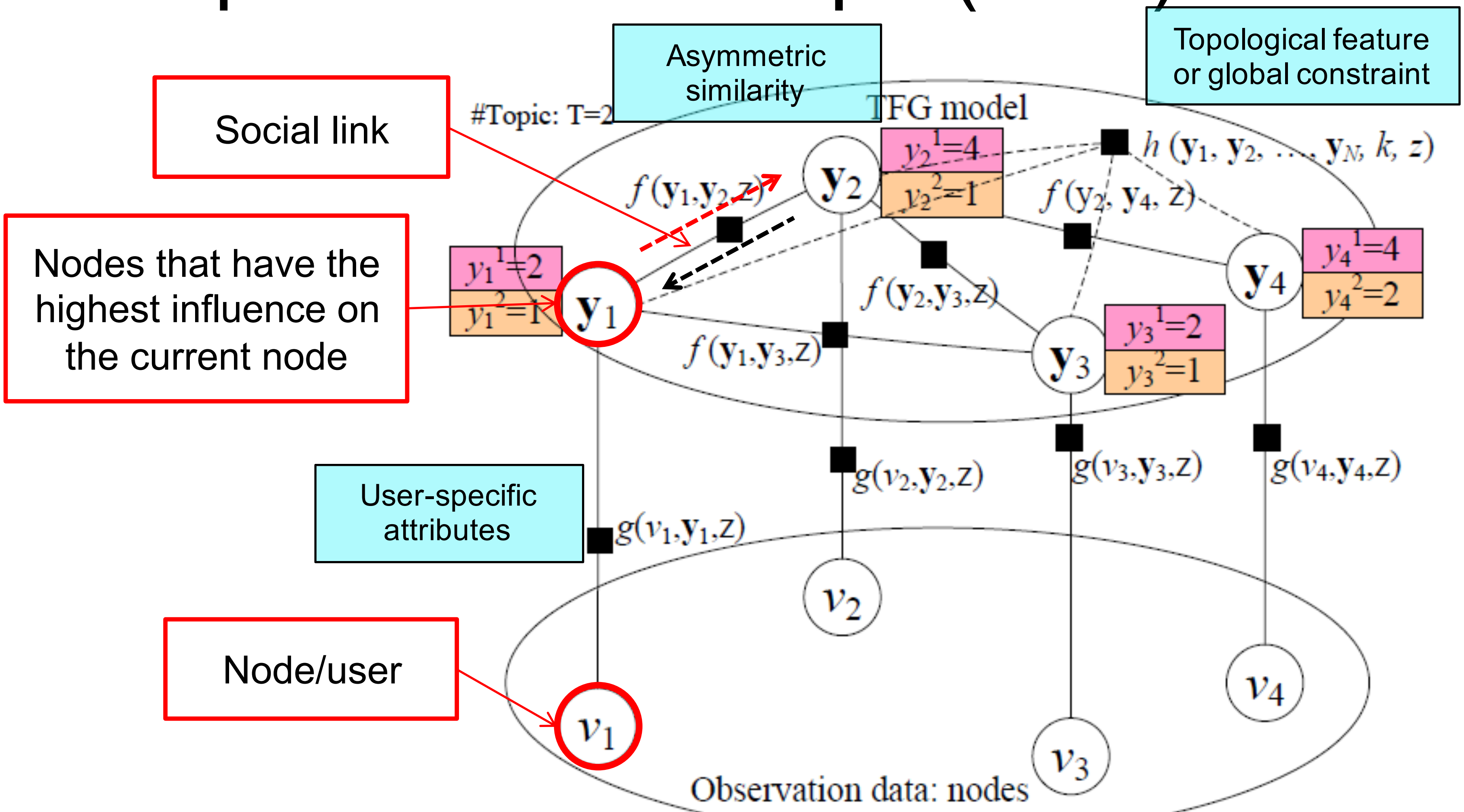
[1] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In KDD, pages 807-816, 2009.

The Solution: Topical Affinity Propagation



[1] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In KDD, pages 807-816, 2009.

Topical Factor Graph (TFG) Model



The problem is cast as identifying which node has the **highest probability** to **influence** another node on a **specific topic** along with the edge.

Topical Factor Graph (TFG)

Objective function:

$$P(\mathbf{v}, \mathbf{Y}) = \frac{1}{Z} \prod_{k=1}^{N} \prod_{z=1}^{T} \boxed{h(\mathbf{y}_1, \ldots, \mathbf{y}_N, k, z)}$$

$$\prod_{i=1}^{N} \prod_{z=1}^{T} \boxed{g(v_i, \mathbf{y}_i, z)} \prod_{e_{kl} \in E} \prod_{z=1}^{T} \boxed{f(\mathbf{y}_k, \mathbf{y}_l, z)}$$

1. How to define?

2. How to optimize?

- The learning task is to find a configuration for all $\{y_i\}$ to maximize the joint probability.

How to define (topical) feature functions?

– Node feature function

similarity

$$g(v_i, \mathbf{y}_i, z) = \begin{cases} \frac{\boxed{w^z_{iy^z_i}}}{\sum_{j \in NB(i)} (w^z_{ij} + w^z_{ji})} & y^z_i \neq i \\ \frac{\sum_{j \in NB(i)} w^z_{ji}}{\sum_{j \in NB(i)} (w^z_{ij} + w^z_{ji})} & y^z_i = i \end{cases}$$

– Edge feature function

$$f(y_i, y_j) = \begin{cases} w[v_i \sim v_j] & y_i = y_j \\ 1 - w[v_i \sim v_j] & y_i \neq y_j \end{cases}$$

or simply binary

– Global feature function

$$h(\mathbf{y}_1, \ldots, \mathbf{y}_N, k, z) = \begin{cases} 0 & \text{if } y^z_k = k \text{ and } y^z_i \neq k \text{ for all } i \neq k \\ 1 & \text{otherwise.} \end{cases}$$

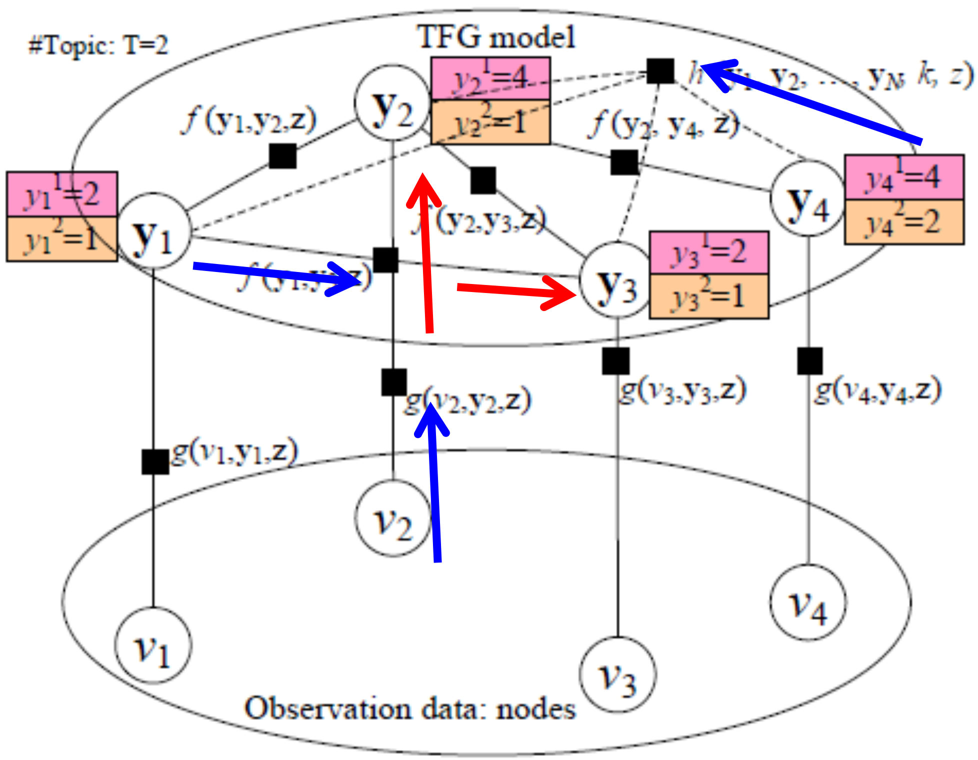
Model Learning Algorithm

Sum-product:

$$\boxed{m_{y\to f}(y,z)} = \prod_{f'\sim y\setminus f} m_{f'\to y}(y,z) \prod_{z'\neq z}\prod_{f'\sim y\setminus f} m_{f'\to y}(y,z')^{(\tau_{z'z})}$$

$$\boxed{m_{f\to y}(y,z)} = \sum_{\sim\{y\}}\left(f(Y,z)\prod_{y'\sim f\setminus y} m_{y'\to f}(y',z)\right)$$

$$+ \sum_{z'\neq z}\tau_{z'z}\sum_{\sim\{y\}}\left(f(Y,z')\prod_{y'\sim f\setminus y} m_{y'\to f}(y',z')\right) \quad (4)$$



- Low efficiency!
- Not easy for distributed learning!

New TAP Learning Algorithm

1. Introduce two new variables *r* and *a, to* replace the original message *m*.

2. Design new update rules:

How user *i* thought he **influenced** user *j*?

m_{ij}

$$r_{ij}^z = b_{ij}^z - \max_{k \in NB(j)} \{b_{ik}^z + a_{ik}^z\}$$

$$a_{jj}^z = \max_{k \in NB(j)} \min\{r_{kj}^z, 0\}$$

$$a_{ij}^z = \min(\max\{r_{jj}^z, 0\}, -\min\{r_{jj}^z, 0\} - \max_{k \in NB(j)\setminus\{i\}} \min\{r_{kj}^z, 0\}), i \in NB(j)$$

How user *j* thought he **was influenced** by user *i*?

[1] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In KDD, pages 807-816, 2009.

The TAP Learning Algorithm

Input: $G = (V, E)$ and topic distributions $\{\theta_v\}_{v \in V}$
Output: topic-level social influence graphs $\{G_z = (V_z, E_z)\}_{z=1}^{T}$

Calculate the node feature function $g(v_i, \mathbf{y}_i, z)$;
Calculate b_{ij}^z according to Eq. 8;
Initialize all $\{r_{ij}^z\} \leftarrow 0$;
repeat
  foreach *edge-topic pair* (e_{ij}, z) **do**
    Update r_{ij}^z according to Eq. 5;
  end
  foreach *node-topic pair* (v_j, z) **do**
    Update a_{jj}^z according to Eq. 6;
  end
  foreach *edge-topic pair* (e_{ij}, z) **do**
    Update a_{ij}^z according to Eq. 7;
  end
until *convergence*;
foreach *node* v_t **do**
  foreach *neighboring node* $s \in NB(t) \cup \{t\}$ **do**
    Compute μ_{st}^z according to Eq. 9;
  end
end
Generate $G_z = (V_z, E_z)$ for every topic z according to $\{\mu_{st}^z\}$;

$$b_{ij}^z = \log \frac{g(v_i, \mathbf{y}_i, z)|_{y_i^z = j}}{\sum_{k \in NB(i) \cup \{i\}} g(v_i, \mathbf{y}_i, z)|_{y_i^z = k}}$$

$$r_{ij}^z = b_{ij}^z - \max_{k \in NB(j)} \{b_{ik}^z + a_{ik}^z\}$$

$$a_{jj}^z = \max_{k \in NB(j)} \min\{r_{kj}^z, 0\}$$

$$a_{ij}^z = \min(\max\{r_{jj}^z, 0\}, -\min\{r_{jj}^z, 0\} - \max_{k \in NB(j) \backslash \{i\}} \min\{r_{kj}^z, 0\}), i \in NB(j)$$

$$\mu_{st}^z = \frac{1}{1 + e^{-(r_{ts}^z + a_{ts}^z)}}$$

Experiments

- Data&Codes: (http://arnetminer.org/lab-datasets/soinf/)

Data set	#Nodes	#Edges
Coauthor	640,134	1,554,643
Citation	2,329,760	12,710,347
Film (Wikipedia)	18,518 films 7,211 directors 10,128 actors 9,784 writers	142,426

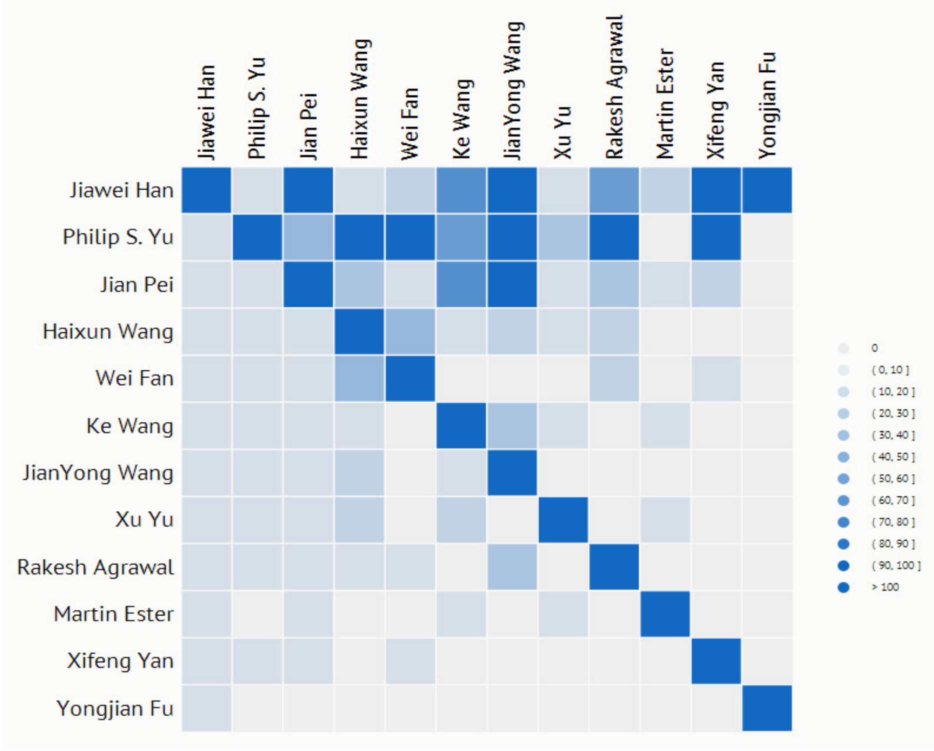
- Evaluation measures
  - CPU time
  - Case study
  - Application

Social Influence Sub-graph on “Data mining”

Table 4: Dynamic influence analysis for Dr. Jian Pei during 2000-2009. Due to space limitation, we only list coauthors who most influence on/by Dr. Pei in each time window.

Year	Pairwise	Influence
2000 - 2001	Influence on Dr. Pei	Jiawei Han (0.4961)
	Influenced by Dr. Pei	Jiawei Han (0.0082)
2002 - 2003	Influence on Dr. Pei	Jiawei Han (0.4045), Ke Wang (0.0418), Jianyong Wang (0.019), Xifeng Yan (0.007), Shiwei Tang (0.0052)
	Influenced by Dr. Pei	Shiwei Tang (0.436), Hasan M.Jamil (0.4289), Xifeng Yan (0.2192), Jianyong Wang (0.1667), Ke Wang (0.0687)
2004 - 2005	Influence on Dr. Pei	Jiawei Han (0.2364), Ke Wang (0.0328), Wei Wang (0.0294), Jianyong Wang (0.0248), Philip S. Yu (0.0156)
	Influenced by Dr. Pei	Chun Tang (0.5929), Shiwei Tang (0.5426), Hasan M.Jamil (0.3318), Jianyong Wang (0.1609), Xifeng Yan (0.1458), Yan Huang (0.1054)
2006 - 2007	Influence on Dr. Pei	Jiawei Han (0.1201), Ke Wang (0.0351), Wei Wang (0.0226), Jianyong Wang (0.018), Ada Wai-Chee Fu (0.0125)
	Influenced by Jian Pei	Chun Tang (0.6095), Shiwei Tang (0.6067), Byung-Won On (0.4599), Hasan M.Jamil (0.3433), Jaewoo Kang (0.3386)
2008 - 2009	Influence on Dr. Pei	Jiawei Han (0.2202), Ke Wang (0.0234), Ada Wai-Chee Fu (0.0208), Wei Wang (0.011), Jianyong Wang (0.0095)
	Influenced by Dr. Pei	ZhaoHui Tang (0.654), Chun Tang (0.6494), Shiwei Tang (0.5923), Zhengzheng Xing (0.5549), Hasan M.Jamil (0.3333), Jaewoo Kang (0.3057)

On “Data Mining” in 2009



Now, let us think…

- What are the fundamental factors behind
  - What is behind the network of social users?
  - What is behind the sentiment of social users?

What drives users' sentiments?

Sentiment vs. Emotion

Emotion is the driving force of user's sentiments…

Charles Darwin:

– Emotion serves as a purpose for humans in aiding their survival during the evolution.[1]

Emotion stimulates the mind 3000 times quicker than rational thought!

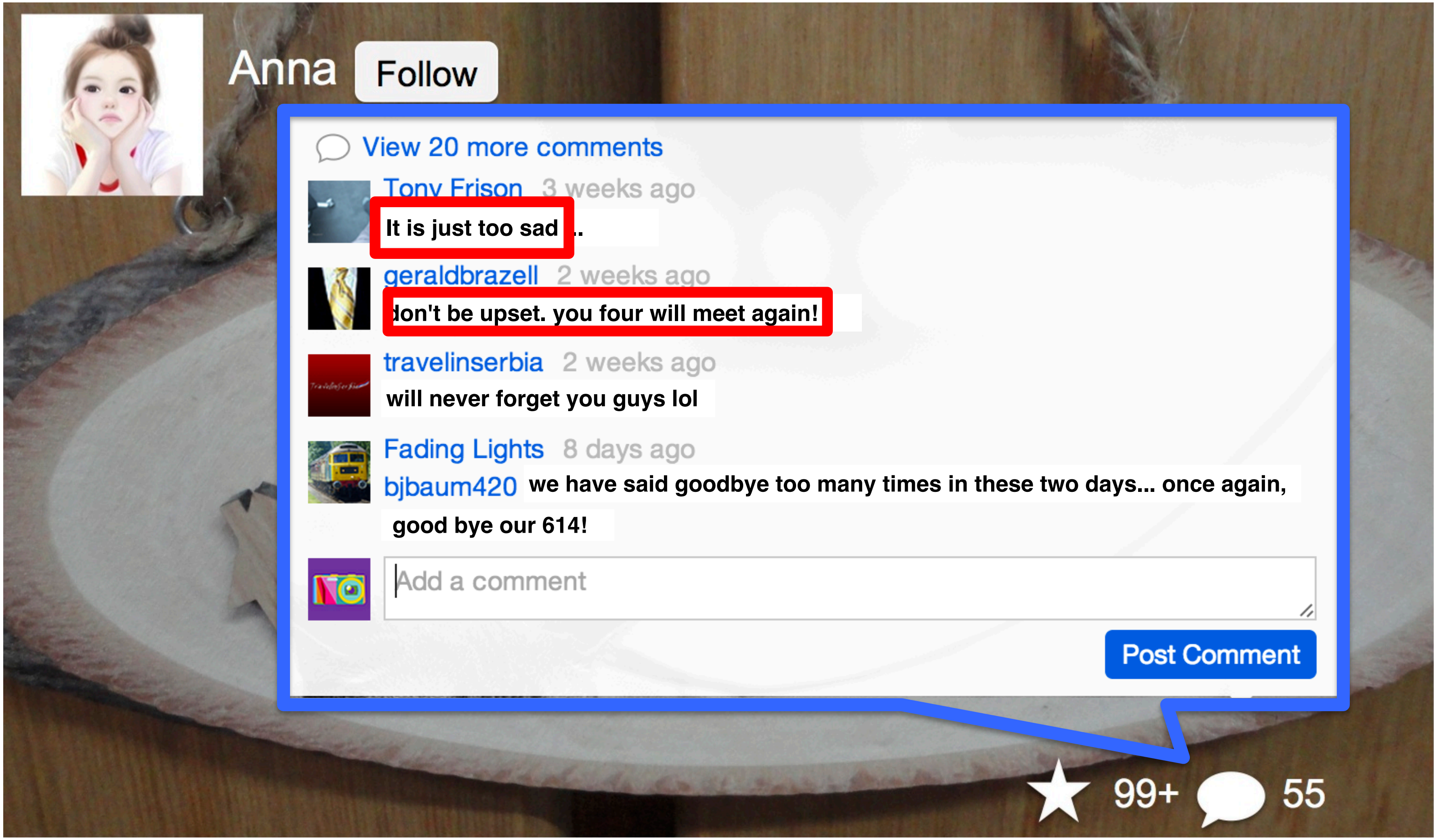
[1] Charles Darwin. *The Expression of Emotions in Man and Animals*. John Murray, 1872.

Potential Directions

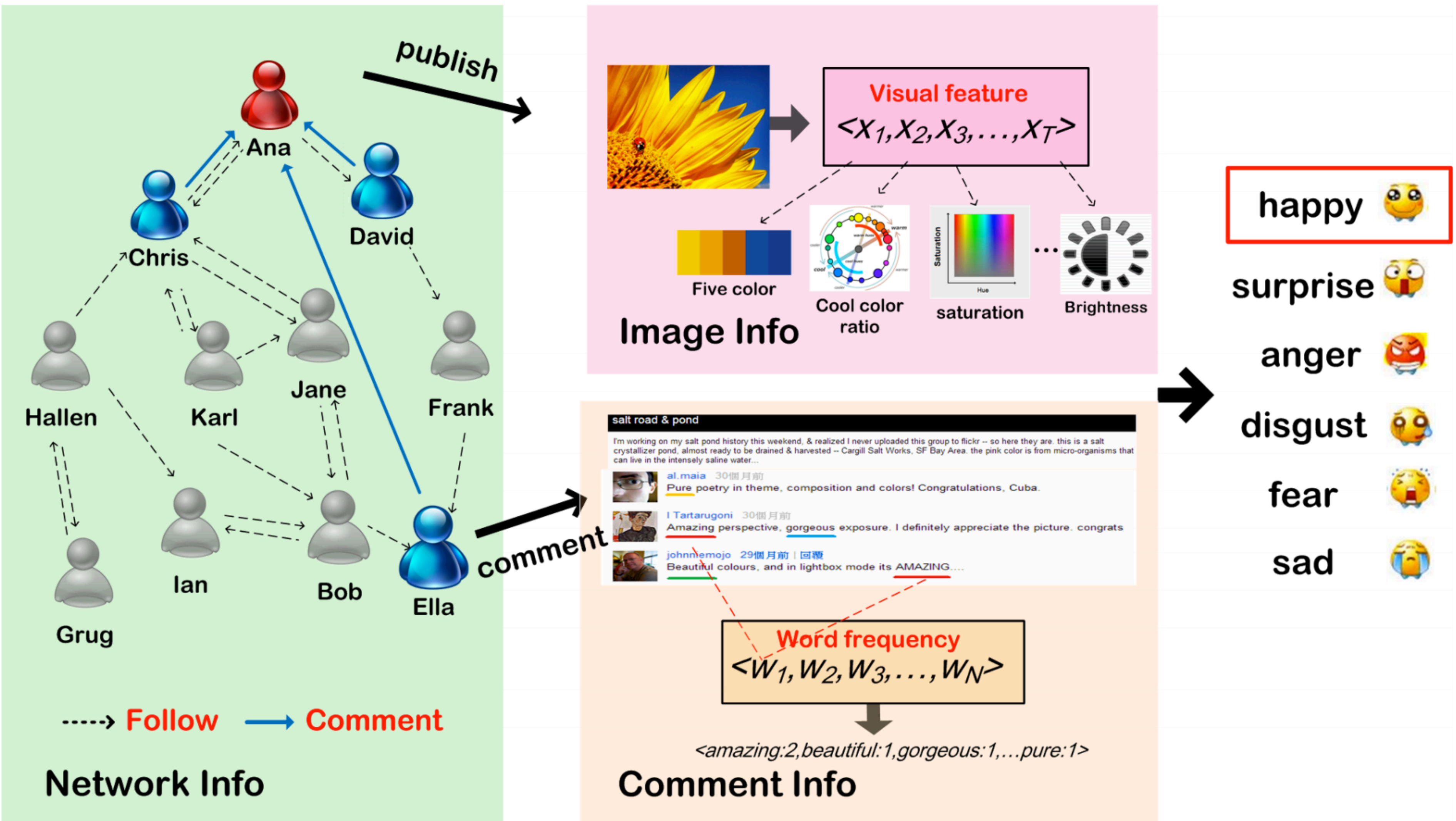
- From sentiment to **emotion analysis**?
- Add **social theories** into emotion analysis?
- Sentiment/emotion analysis for "**Social Good**"?

Was Anna Happy When She Published This Photo On Flickr?

Was Anna Happy When She Published This Photo On Flickr?

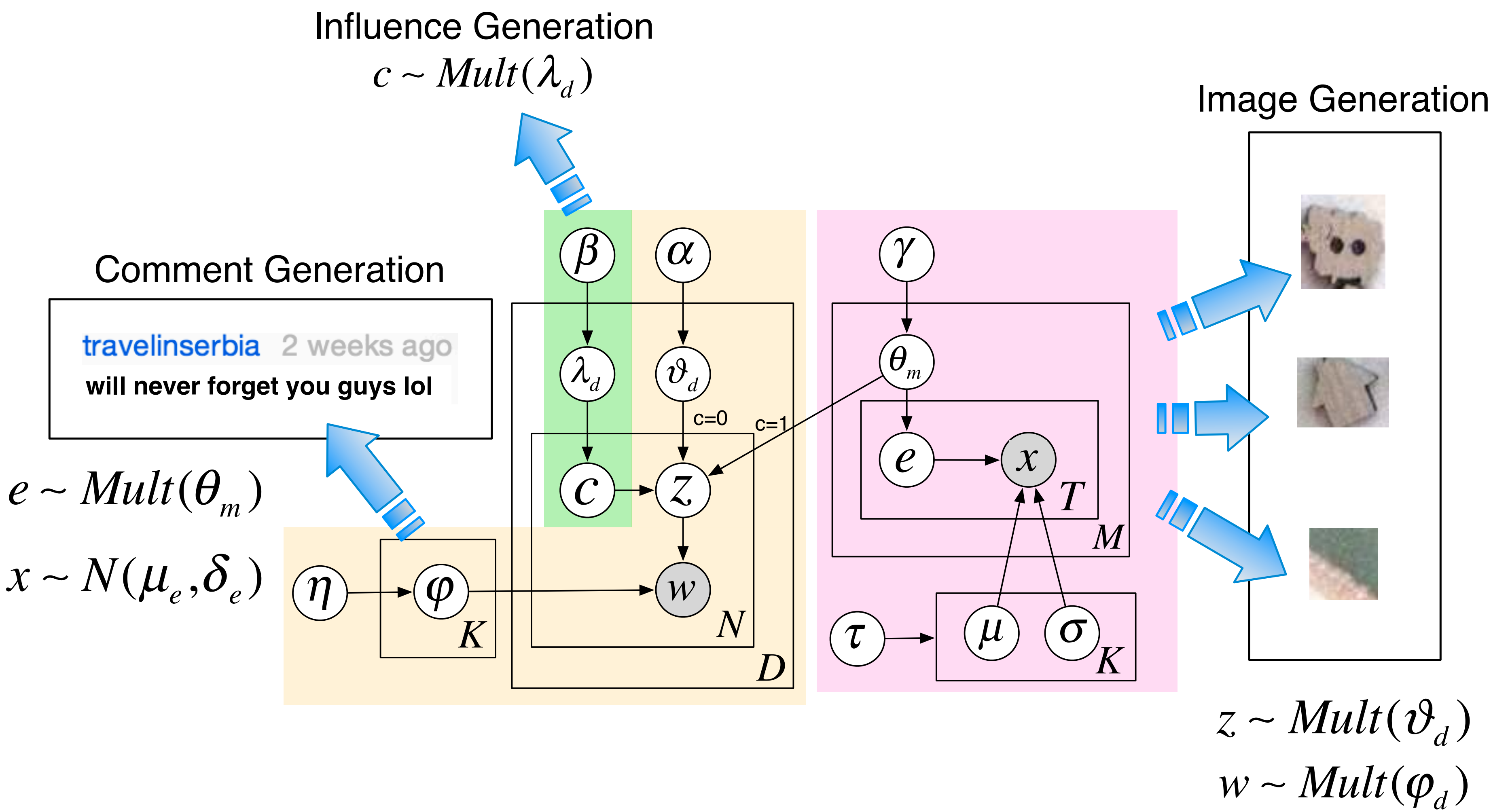


Problem



[1] Yang Yang, Jia Jia, Shumei Zhang, Boya Wu, Qicong Chen, Juanzi Li, Chunxiao Xing, and Jie Tang. How Do Your Friends on Social Media Disclose Your Emotions? In **AAAI'14**. pp. 306-312.

Emotion Learning Method



[1] Yang Yang, Jia Jia, Shumei Zhang, Boya Wu, Qicong Chen, Juanzi Li, Chunxiao Xing, and Jie Tang. How Do Your Friends on Social Media Disclose Your Emotions? In **AAAI'14**. pp. 306-312.

Flickr Data

- 354,192 images posted by 4,807 users
  - For each image, we also collect its tags and all comments.
  - Thus we get 557,177 comments posted by 6,735 users in total
- Infer emotion of users by considering both image and tag/comments

Emotion Inference

Averagely **+37.4%** in terms of F1

Table 2: Performance of emotion inference.

Emotion	Method	Precision	Recall	F1-score	Emotion	Method	Precision	Recall	F1-score
Happiness	SVM	0.242	0.279	0.259	Disgust	SVM	0.192	0.236	0.212
	PFG	0.337	0.312	0.324		PFG	0.309	0.374	0.339
	LDA+SVM	0.333	**0.727**	**0.457**		LDA+SVM	0.223	0.223	0.223
	EL+SVM	**0.367**	0.410	0.388		EL+SVM	**0.331**	**0.432**	**0.374**
Surprise	SVM	0.197	0.037	0.063	Fear	SVM	0.204	0.264	0.230
	PFG	0.349	0.340	0.345		PFG	0.301	**0.408**	0.347
	LDA+SVM	0.218	0.048	0.078		LDA+SVM	0.211	0.225	0.217
	EL+SVM	**0.425**	**0.516**	**0.466**		EL+SVM	**0.371**	0.343	**0.356**
Anger	SVM	0.188	0.105	0.135	Sadness	SVM	0.225	0.365	0.278
	PFG	0.191	0.142	0.163		PFG	0.357	0.286	0.317
	LDA+SVM	0.222	0.109	0.146		LDA+SVM	0.257	0.278	0.267
	EL+SVM	**0.390**	**0.370**	**0.380**		EL+SVM	**0.561**	**0.617**	**0.588**

SVM: regards the visual features of images as inputs and uses a SVM as a classifier.

PFG: considers both color features and social correlations among images.

LDA+SVM: first uses LDA to extract latent topics from comments, then uses visual features, topic distributions, and social ties as features to train a SVM.

To What Extend Your Friends Can Disclose Your Emotions?

-Comments stands for the proposed method ignoring comment information

-Tie ignores social tie information

Fear images have similar visual features with Sadness and Anger.

Homophily suggests that friends with similar interests tend to have similar understanding of disgust

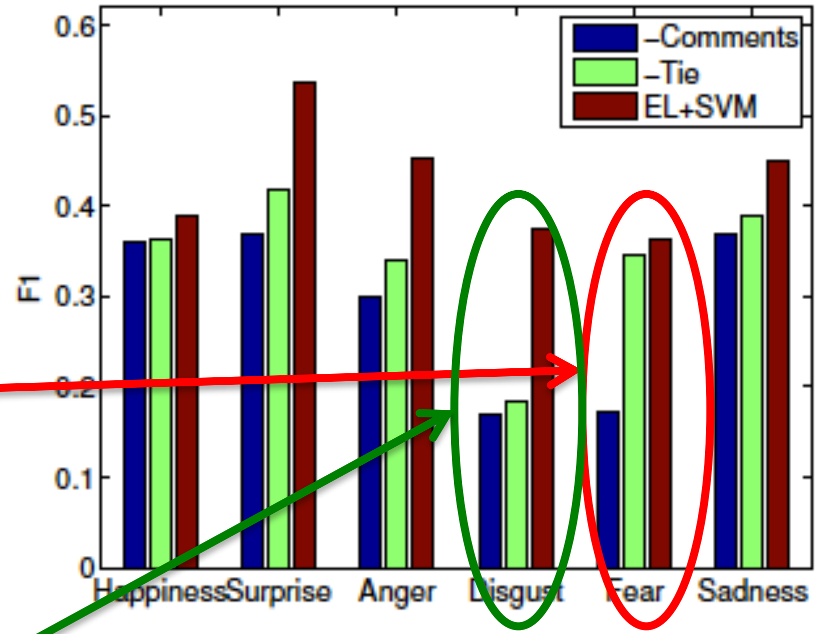
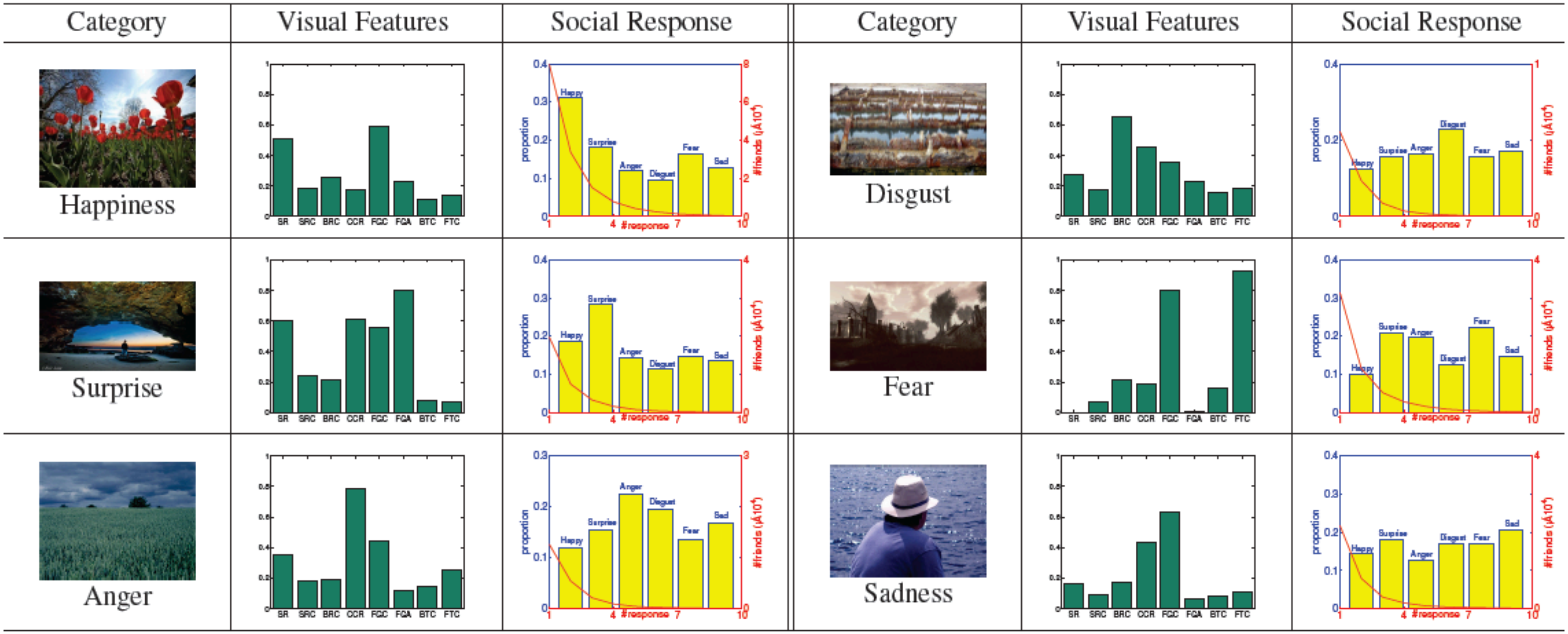


Image Interpretations



- Our model demonstrates how visual features distribute over different emotions. (e.g., images representing Happiness have high saturation)
- Positive emotions attract more response (***+4.4*** times) and more easily to influence others compared with negative emotions.

Potential Directions

- From sentiment to **emotion analysis**?
- Add **social theories** into emotion analysis?
- Sentiment/emotion analysis for “**Social Good**”?

Summary

- From text sentiment to user sentiment
- From user sentiment to network sentiment
- From sentiment analysis to emotion analysis
- From network interaction to social influence

Related Publications

- Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In **KDD'09**, pages 807-816, 2009.
- Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In **KDD’11**, pages 1397–1405, 2011.
- Jie Tang, Sen Wu, and Jimeng Sun. Confluence: Conformity Influence in Large Social Networks. In **KDD'13**, pages 347-355, 2013.
- Yang Yang, Jia Jia, Shumei Zhang, Boya Wu, Qicong Chen, Juanzi Li, Chunxiao Xing, and Jie Tang. How Do Your Friends on Social Media Disclose Your Emotions? In **AAAI'14**. pp. 306-312.
- Jie Tang, Yuan Zhang, Jimeng Sun, Jinghai Rao, Wenjing Yu, Yiran Chen, and ACM Fong. Quantitative Study of Individual Emotional States in Social Networks. IEEE Transactions on Affective Computing (**TAC**), 2012, Volume 3, Issue 2, Pages 132-144. (Selected as the Spotlight Paper)
- Xiaohui Wang, Jia Jia, Jie Tang, Boya Wu, Lianhong Cai, and Lexing Xie. Modeling Emotion Influence in Image Social Networks. IEEE Transactions on Affective Computing (**TAC**), Volume 6, Issue 3, 2015, Pages 286-297.
- Yuan Zhang, Jie Tang, Jimeng Sun, Yiran Chen, and Jinghai Rao. MoodCast: Emotion Prediction via Dynamic Continuous Factor Graph Model. In **ICDM’10**. pp. 1193-1198.
- Jia Jia, Sen Wu, Xiaohui Wang, Peiyun Hu, Lianhong Cai, and Jie Tang. Can We Understand van Gogh’s Mood? Learning to Infer Affects from Images in Social Networks. In **ACM MM**, pages 857-860, 2012.
- Xiaohui Wang, Jia Jia, Peiyun Hu, Sen Wu, Lianhong Cai, and Jie Tang. Understanding the Emotional Impact of Images. **(Grand Challenge)** In **ACM MM**. pp. 1369-1370. (Grand Challenge 2nd Prize Award)

Thank you！

Collaborators: Lillian Lee, Chenhao Tan (**Cornell**)

Jinghai Rao (**Nokia**) Jimeng Sun (**IBM/GIT**)

Ming Zhou, Long Jiang (**Microsoft**)

Yuan Zhang, Jia Jia, Yang Yang, Boya Wu, Xiaohui Wang (**THU**)

Jie Tang, KEG, Tsinghua U,	http://keg.cs.tsinghua.edu.cn/jietang
Download all data & Codes,	http://aminer.org/data http://aminer.org/data-sna