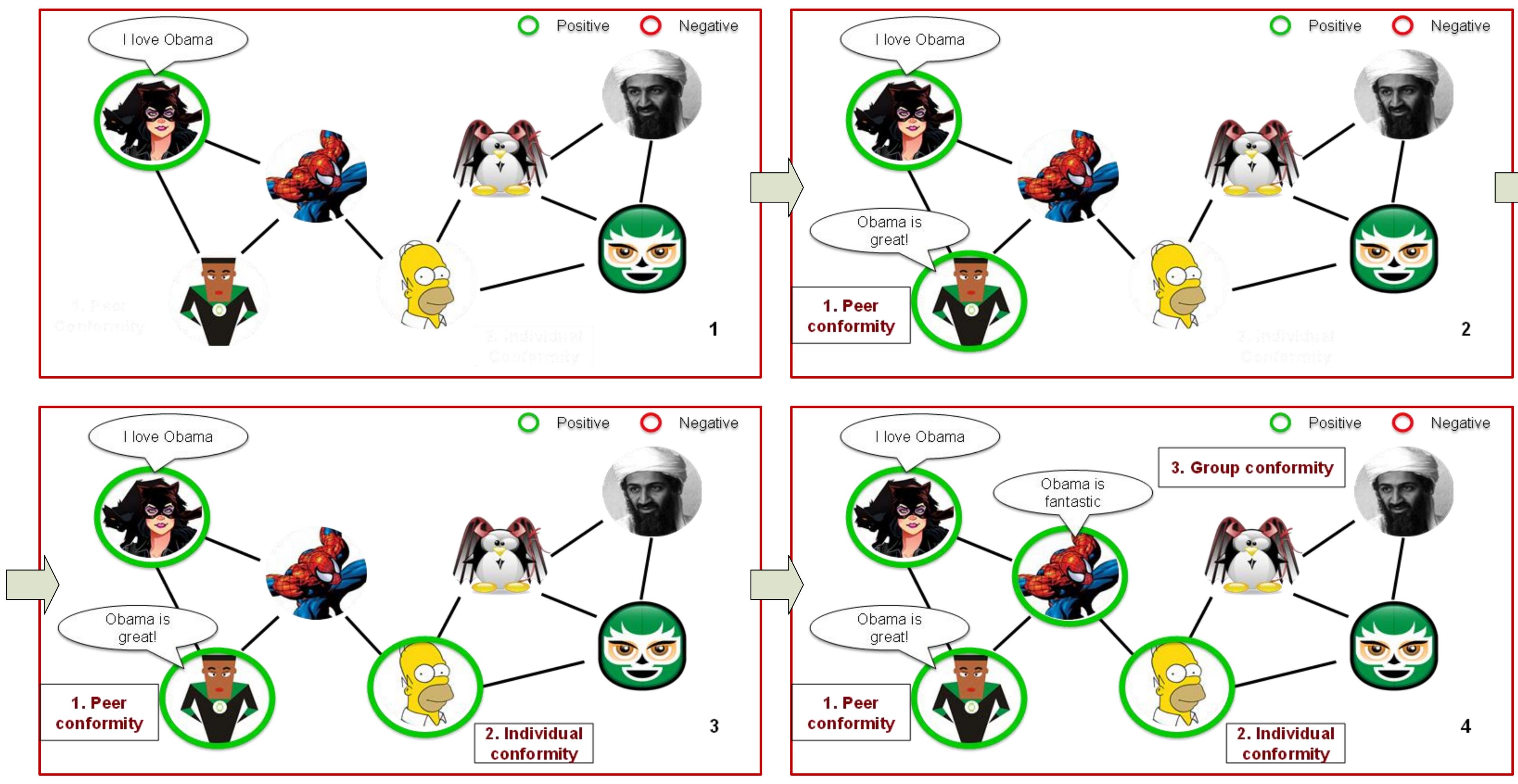
Confluence: Conformity Influence in Large Social Networks

Jie Tang
Computer Science
Tsinghua University
jietang@tsinghua.edu.cn

Sen Wu
Computer Science
Tsinghua University
ronaldosen@gmail.com

Jimeng Sun
IBM TJ Watson
Research Center
jimeng@us.ibm.com

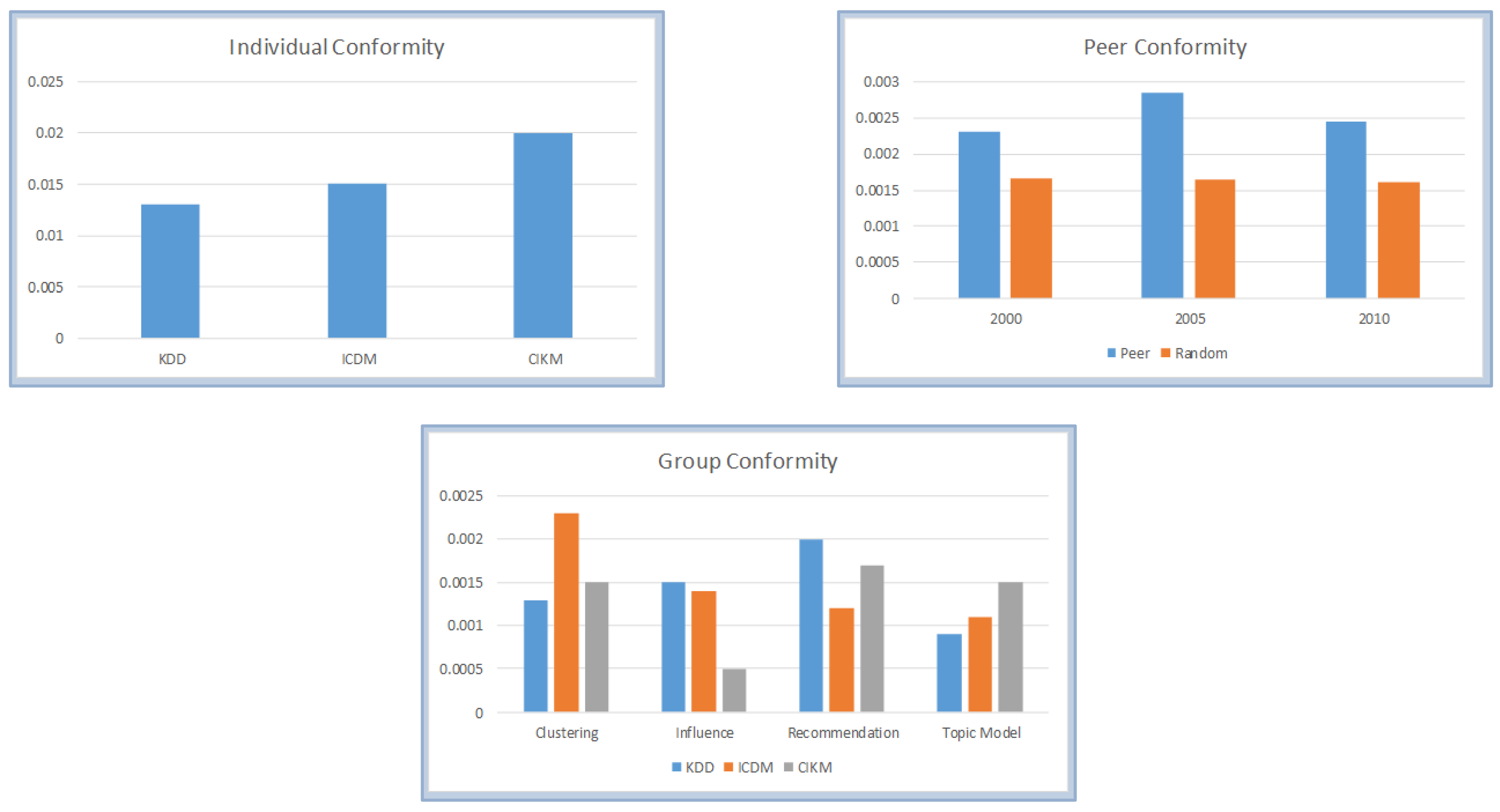
Conformity is the act of matching attitudes, beliefs, and behaviors to group norms



Challenges:

- How to formally define and differentiate different types of conformities?
- How to construct a computational model to learn the different conformity factors?
- How to validate the proposed model in real large networks?

Conformtiy in Co-Auhtor Network



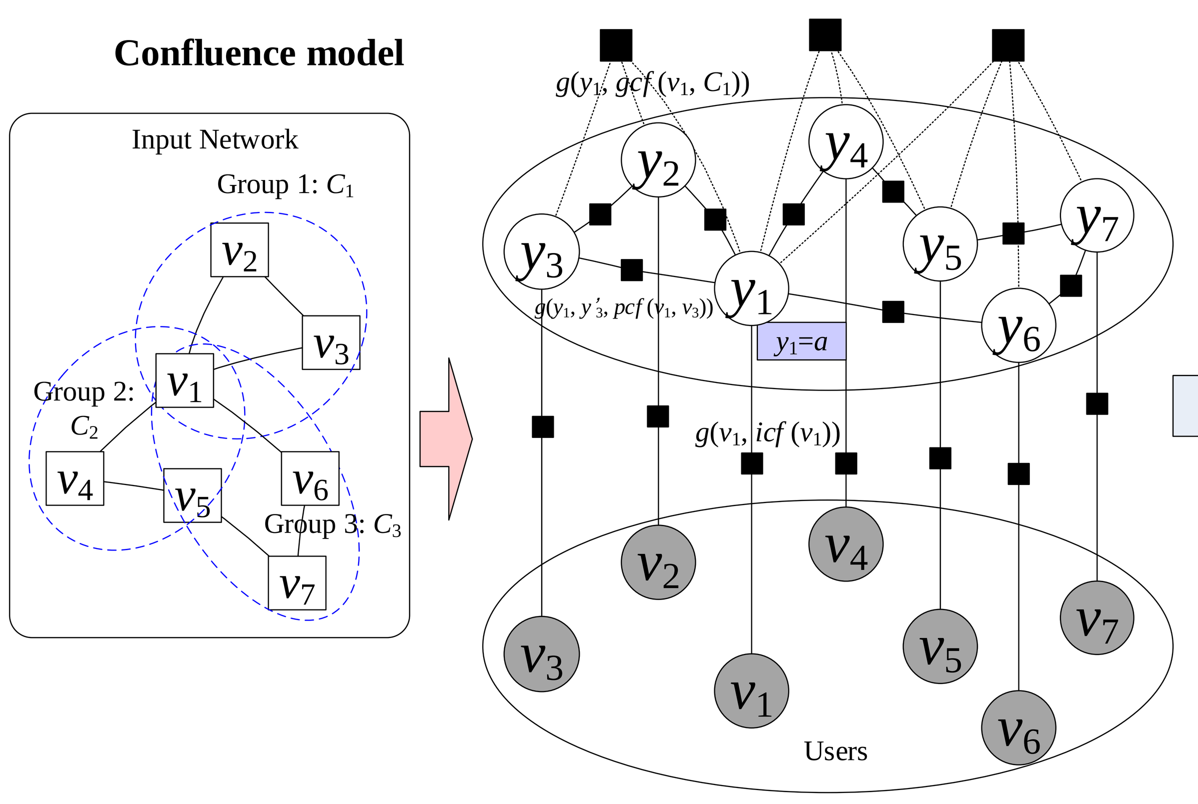
Individual Conformity: $icf(v) = \frac{|(a,v,t) \in A_v | \exists (a,v',t') : e_{vv'} \in E \wedge \epsilon \geq t - t' \geq 0|}{|A_v|}$

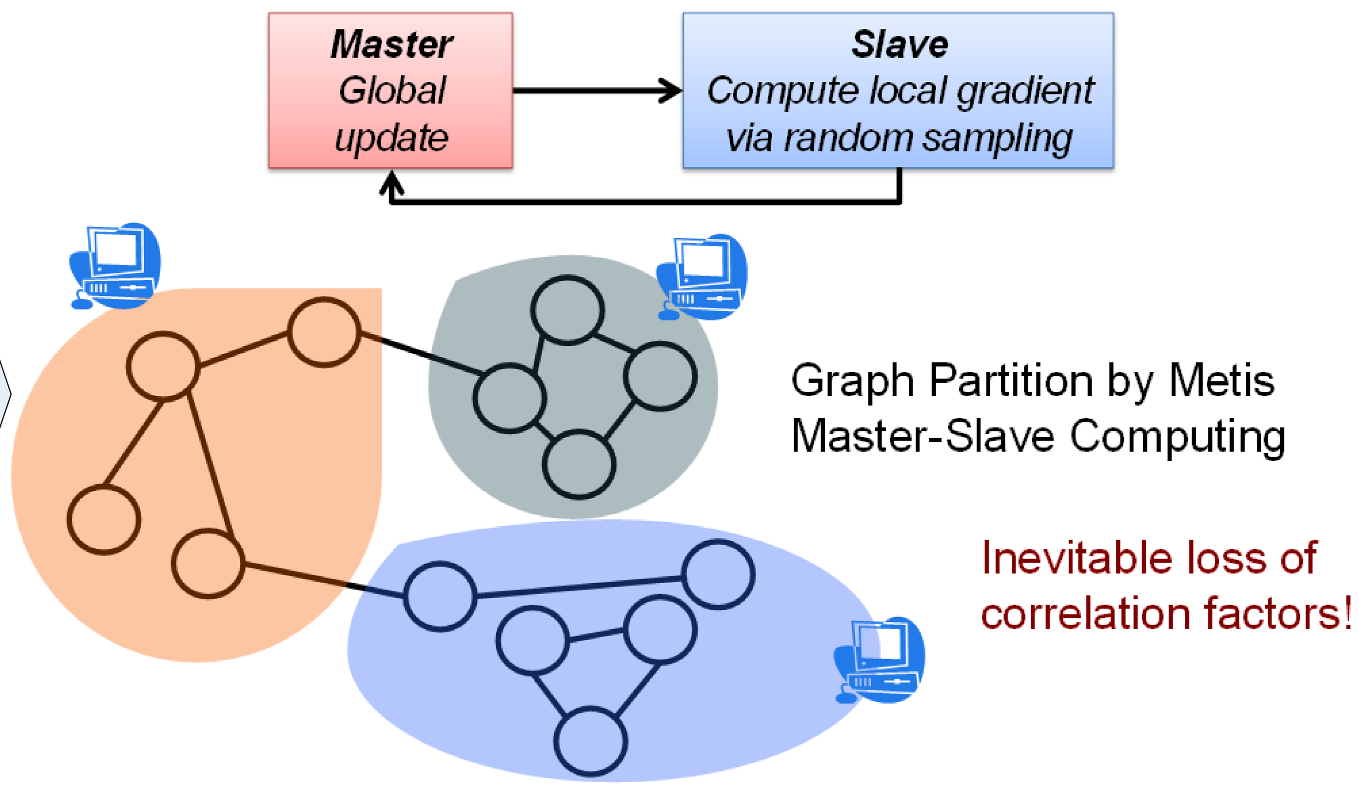
Peer Conformity: $pcf(v,v') = \frac{|(a,v',t') \in A_{v'} | \exists (a,v,t) : e_{vv'} \in E \wedge \epsilon \geq t - t' \geq 0|}{|A_{v'}|}$

Group Conformity: $gcf^{\tau}(v, C_{vk}) = \frac{|(a,v',t') \in A^{\tau}_{C_k} | \exists (a,v,t) : \mathbb{I}[c_{ik}] \wedge \epsilon \geq t - t' \geq 0|}{|A^{\tau}_{C_k}|}$

(a, v_i, t) to represent user v_i performed action $a \in \mathcal{A}$ at time t

Conformity-aware Factor Graph Model (ConFG)





Input: network G, action history A, and learning rate η;
Output: learned parameters $\theta = (\{\alpha\}, \{\beta\}, \{\gamma\}, \{\mu\})$;

Initialize $\alpha, \beta, \gamma, \mu$;
Construct the graphical structure G in the Confluence model;
Partition the graph G into M subgraphs $[G_1, \cdots, G_M]$;
repeat
  %Distribute the parameter to calculate local belief ;
  Master broadcasts θ to all Slaves;
  for $l = 1$ *to* M **do**
    Each Slave calculates local belief for each marginal probability according to Eqs. 6 and 7 on subgraph G_l;
    Slave send back the obtained local belief;
  end
  %Calculate the marginals and update all parameters ;
  Master calculates the marginal according to Eq. 8;
  Master calculates the gradient for each parameter (e.g., by Eq. 5);
  Master updates all parameters, e.g. for α_j,

$$\alpha_j^{new} = \alpha_j^{old} + \eta \frac{\mathcal{O}(\theta)}{\alpha_j}$$

until *convergence*;

Algorithm 1: Distributed model learning.

Conformity Factors

$$g(y_i, y_j', pcf(v_i, v_j)) = (\frac{1}{2})^{\frac{t-t'}{\lambda}} pcf(v_i, v_j)$$

$$g(y_i, gcf^{\tau}(v_i, C_k)) = (\frac{1}{2})^{\frac{t-t'}{\lambda}} gcf^{\tau}(v_i, C_k)$$

$$g(y_i, icf(v_i)) = \frac{\sum_{k=1}^{|A_{v_i}|} (\frac{1}{2})^{\frac{t-t'}{\lambda}} \mathbb{I}[y_j' \wedge e_{ij} \in E]}{|A_v|}$$

$A_v \subset A$ denotes the action history of user v
$\mathbb{I}[c_{ik}]$ is an indicator function

Objective Function

$$\begin{aligned} \mathcal{O}(\theta) &= \log P_\theta(Y|G, A) \\ &= \sum_{i=1}^{N} \left[\sum_{j=1}^{d} \alpha_j f(y_i, x_{ij}) + \beta_i g(y_i, icf(v_i)) \right] \\ &\quad + \sum_{e_{ij} \in E} \mathbb{I}[y'] \gamma_{ij} g(y_i, y_j', pcf(v_i, v_j)) \\ &\quad + \sum_{i=1}^{N} \sum_{k=1}^{m} \mathbb{I}[c_{ik}] \mu_{ik} g(y_i, gcf(v_i, C_k)) - \log Z \end{aligned}$$

learned parameters $\theta = (\{\alpha\}, \{\beta\}, \{\gamma\}, \{\mu\})$

Marginal probability

$$\begin{aligned} m_{ij}^l(y_i) &= \sigma \sum_{y_i} \psi_{ij}^l(y_i, y_j) \psi_i^l(y_i) \prod_{k \in NB(i) \backslash j} m_{ki}^l(y_i) \\ b_i^l(y_i) &= \psi_i^l(y_i) \prod_{k \in NB(i)} m_{ki}^l(y_i) \\ P(y_i|.) &= \sigma \sum_{l=1}^{M} b_i^l(y_i) \end{aligned}$$

$m_{ij}^l(y_i)$ is the "belief" propagated from node y_j to node y_i
$\psi_i^l(y_i)$ denotes all factor functions related to y_i in the subgraph G_l

Empirical Analysis

Datasets:

Flickr — Photo sharing network (04/01/2012 – 06/16/2012).
- Users share photos and add comments to other photos.

Gowalla — Location-based social network (07/10/2010 – 07/29/2010).
- Users share their locations by checking-in.

Weibo — Most popular microblogging service in China (09/28/2012 – 10/29/2012).
- Collect a complete network and all the tweets.

Co-Author — Co-author network (1975 – 2012).
- Computer science authors and relationships from Arnetminer.org.

Statistics of the four networks

Dataset	Flickr	Gowalla	Weibo	Co-Author
#nodes	1,991,509	196,591	1,776,950	737,690
#edges	208,118,719	950,327	308,489,739	2,416,472
#groups	460,888	N/A	N/A	60
#actions	3,531,801	6,442,890	6,761,186	1,974,466

Baselines:

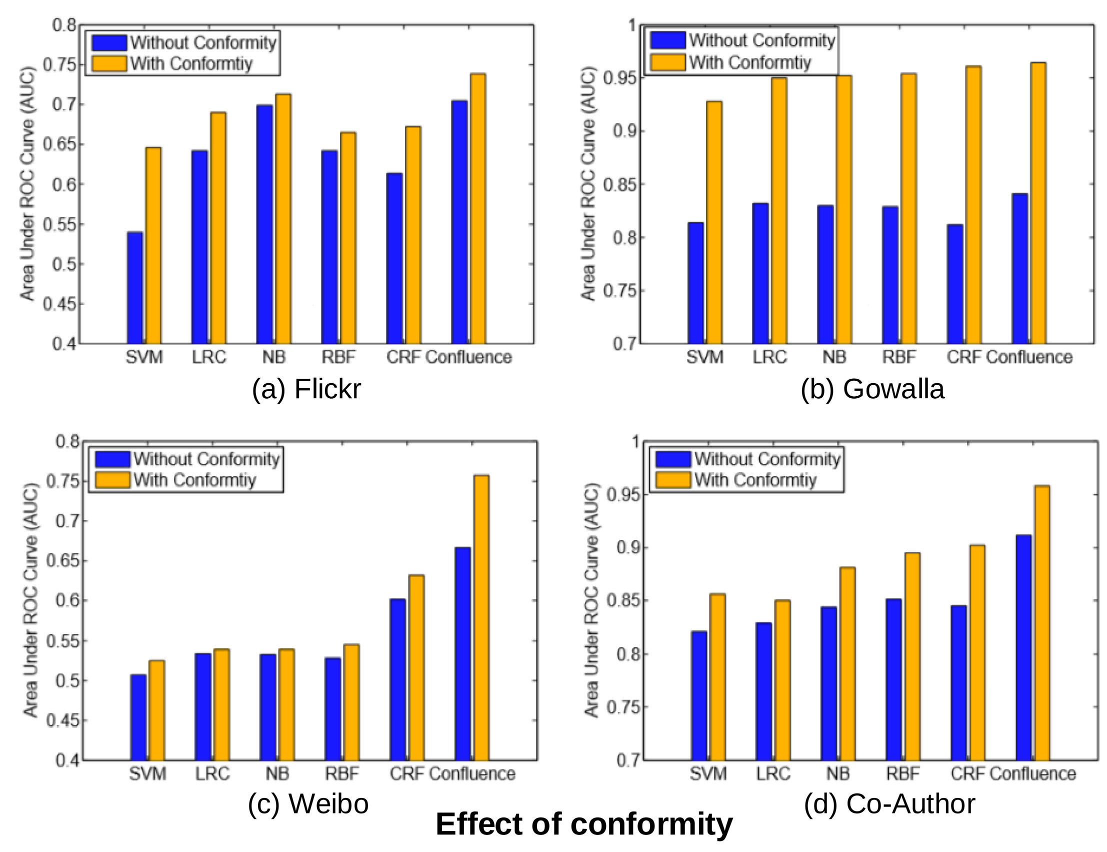
Support Vector Machine (SVM)—Features associated with users
Logistic Regression (LR)—Features associated with users
Naive Bayes (NB)—Features associated with users
Gaussian Radial Basis Function Neural Network (RBF)—Features associated with users
Conditional Random Field (CRF)—Use attributes-based, social-based, and individual conformity features

Evaluation metrics:

- *Prediction accuracy:* evaluate its performance in terms of Precision, Recall, F1-Measure, and Area Under Curve (AUC).
- *Scalability performance:* Evaluate the computational time as the efficiency metric.
- *Qualitative case study:* Use a case study to demonstrate the effectiveness of model

Average prediction performance of different methods on the four data sets

Data	Method	Precision	Recall	F1-Measure	AUC
Flickr	SVM	0.5921 (±0.0036)	0.5905 (±0.0031)	0.5802 (±0.0012)	0.6473 (±0.0004)
	LR	0.6010 (±0.0052)	0.5900 (±0.0057)	0.5770 (±0.0018)	0.6510 (±0.0008)
	NB	0.6170 (±0.0071)	0.6040 (±0.0083)	0.5920 (±0.0031)	0.6520 (±0.0019)
	RBF	**0.6250 (±0.0039)**	0.5960 (±0.0010)	0.5720 (±0.0024)	0.6700 (±0.0010)
	CRF	0.5474 (±0.0030)	**0.8002 (±0.0009)**	0.6239 (±0.0016)	0.6722 (±0.0010)
	Confluence	0.5472 (±0.0025)	0.7770(±0.0010)	**0.6342 (±0.0010)**	**0.7383 (±0.0006)**
Gowalla	SVM	0.9290 (±0.0212)	0.9310 (±0.0121)	0.9295 (±0.0105)	0.9280 (±0.0042)
	LR	0.9320 (±0.0234)	0.9290 (±0.0234)	0.9310 (±0.0155)	0.9500 (±0.0054)
	NB	0.9310 (±0.0197)	0.9290 (±0.0335)	0.9300 (±0.0223)	0.9520 (±0.0030)
	RBF	0.9320 (±0.0254)	0.9280 (±0.0284)	0.9300 (±0.0182)	0.9540 (±0.0022)
	CRF	0.9330 (±0.0100)	0.9320 (±0.0291)	0.9330 (±0.0164)	0.9610 (±0.0019)
	Confluence	**0.9372 (±0.0097)**	**0.9333 (±0.0173)**	**0.9352 (±0.0101)**	**0.9644 (±0.0140)**
Weibo	SVM	0.5060 (±0.0381)	0.5060 (±0.0181)	0.5060 (±0.0157)	0.5070 (±0.0053)
	LR	0.5190 (±0.0461)	0.6450 (±0.0104)	0.5750 (±0.0281)	0.5390 (±0.0133)
	NB	0.5120 (±0.0296)	0.6700 (±0.0085)	0.5810 (±0.0165)	0.5390 (±0.0132)
	RBF	**0.5240 (±0.0248)**	0.5690 (±0.0098)	0.5460 (±0.0159)	0.5450 (±0.0103)
	CRF	0.5150 (±0.0353)	0.6310 (±0.0121)	0.5720 (±0.0209)	0.6320 (±0.0139)
	Confluence	0.5185 (±0.0296)	**0.9967 (±0.0085)**	**0.6816 (±0.0156)**	**0.7572 (±0.0077)**
Co-Author	SVM	0.7672 (±0.0338)	0.8671 (±0.0145)	0.8256 (±0.0129)	0.8562 (±0.0115)
	LR	0.8700 (±0.0261)	0.7640 (±0.0346)	0.8140 (±0.0221)	0.8500 (±0.0030)
	NB	0.7640 (±0.0177)	0.8510 (±0.0185)	0.8050 (±0.0048)	0.8720 (±0.0074)
	RBF	0.7720 (±0.0182)	0.8830 (±0.0191)	0.8240 (±0.0145)	0.8790 (±0.0031)
	CRF	0.8081 (±0.0252)	0.8771 (±0.0249)	0.8360 (±0.0087)	0.9025 (±0.0025)
	Confluence	**0.8818 (±0.0105)**	**0.9089 (±0.0130)**	**0.8818 (±0.0084)**	**0.9579 (±0.0022)**



(a) Flickr (b) Gowalla (c) Weibo (d) Co-Author

Effect of conformity

Running time of the proposed algorithm (hour).

Data Set	Flickr	Gowalla	Weibo	Co-Author
Confluence	1.602	0.245	1.083	0.512
Confluence (single)	19.637	2.395	11.229	6.464
CRF	3.864	0.387	2.547	1.823

Paper ID: 1136

References

Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In SIGKDD'2009. pp. 807-816.

B. D. Bernheim. A theory of conformity. Journal of Political Economy, 1027(5):841–877, 1994.

R. B. Cialdini and N. J. Goldstein. Social influence: Compliance and conformity. Annual Review of Psychology, 55:591–621, 2004.

T. Lou and J. Tang. Mining structural hole spanners through information diffusion in social networks. In WWW'13, pages 837–848, 2013.