Confluence: Conformity Influence in Large Social Networks

Jie Tang[*], Sen Wu[*], and Jimeng Sun[+]

[*]Tsinghua University

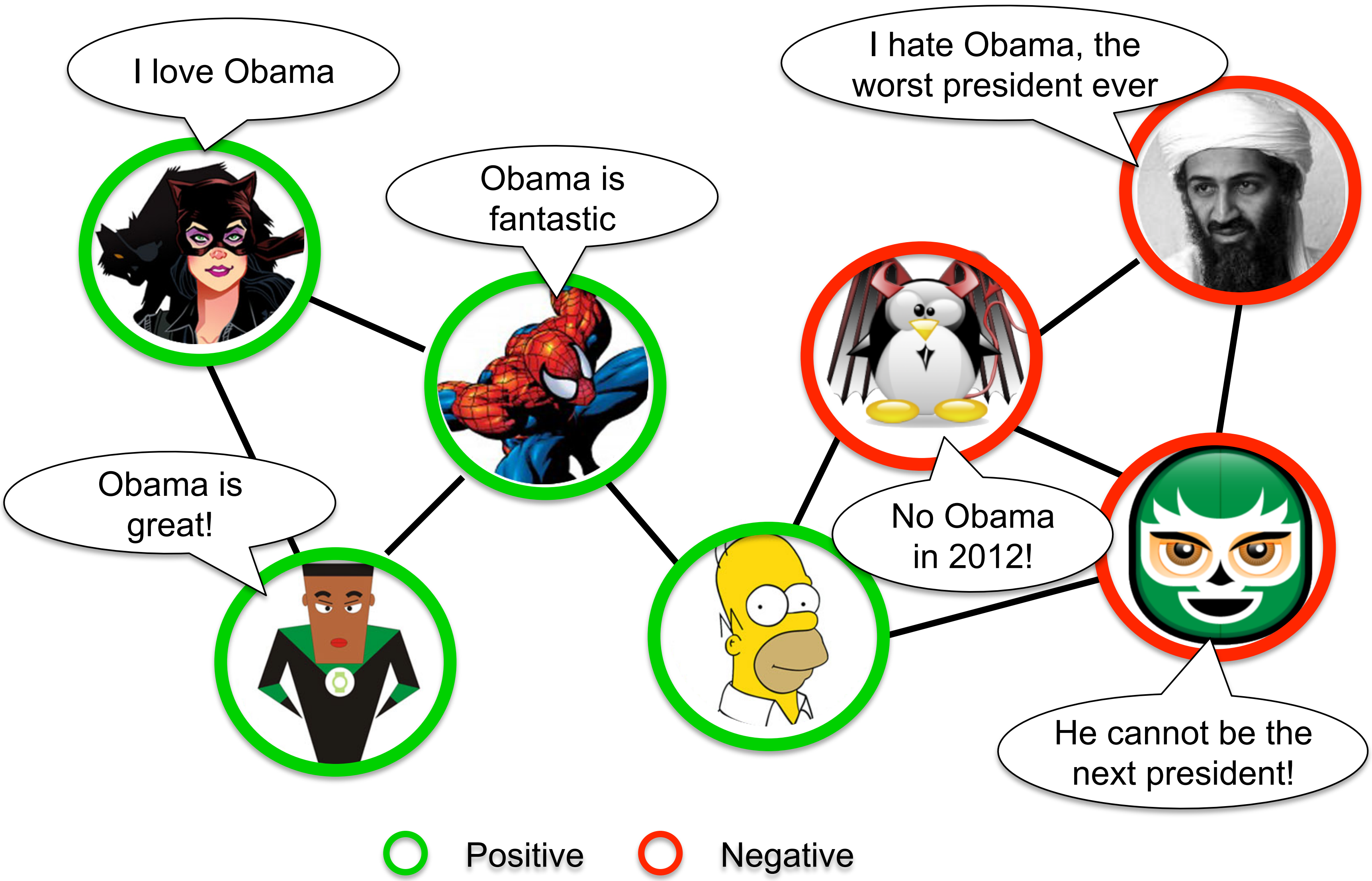
[+]IBM TJ Watson Research Center

Conformity

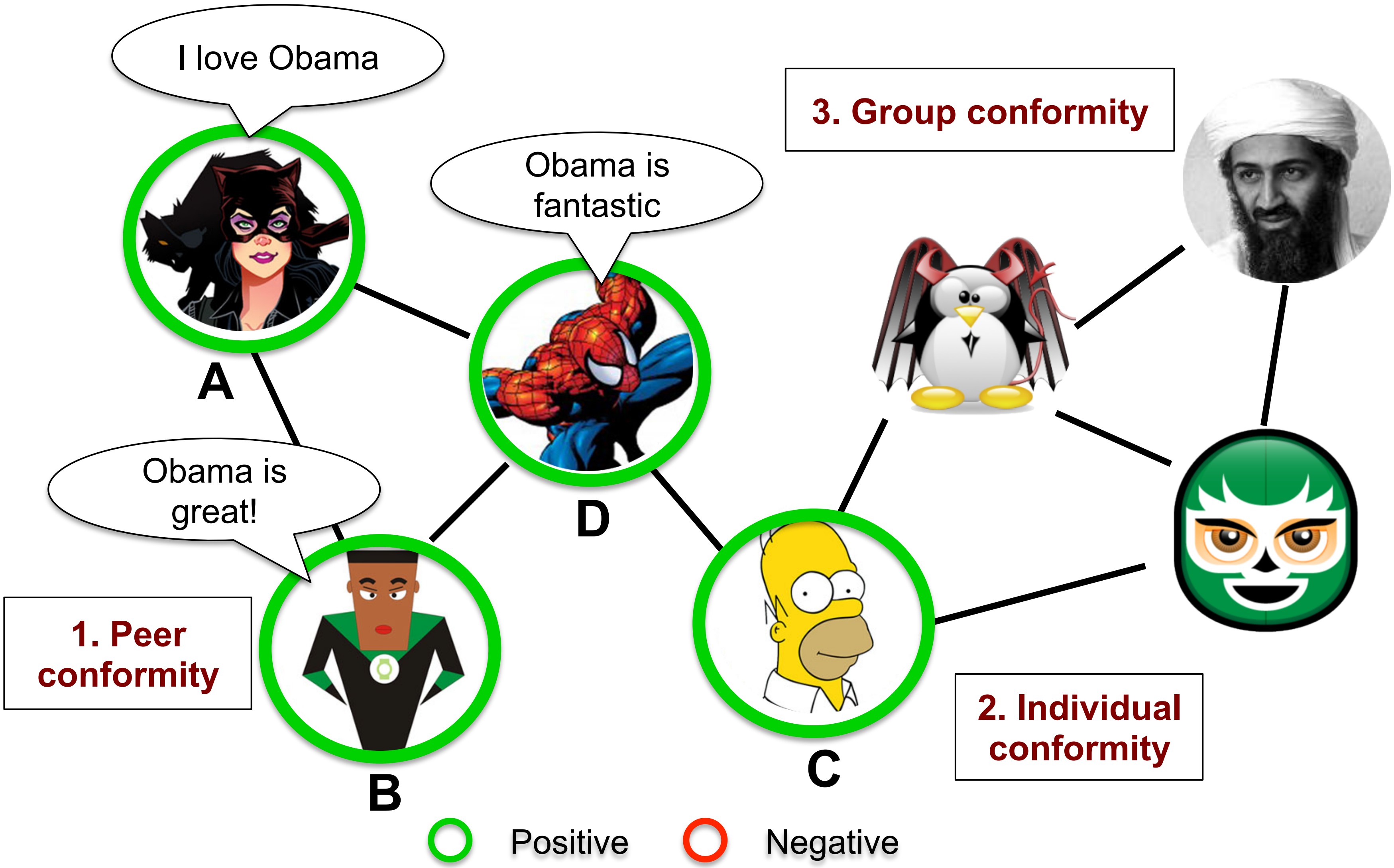
- Conformity is the act of matching attitudes, opinions, and behaviors to group norms.[1]
- Kelman identified three major types of conformity[2]
  - Compliance is public conformity, while possibly keeping one's own original beliefs for yourself.
  - Identification is conforming to someone who is liked and respected, such as a celebrity or a favorite uncle.
  - Internalization is accepting the belief or behavior, if the source is credible. It is the deepest influence on people and it will affect them for a long time.

[1] R.B. Cialdini, & N.J. Goldstein. Social influence: Compliance and conformity. Annual Review of Psych., 2004, 55, 591–621.
[2] H.C. Kelman. Compliance, Identification, and Internalization: Three Processes of Attitude Change. Journal of Conflict Resolution, 1958, 2 (1): 51–60.

“Love Obama”


I love Obama
Obama is fantastic
I hate Obama, the worst president ever
Obama is great!
No Obama in 2012!
He cannot be the next president!
Positive
Negative

Conformity Influence Analysis



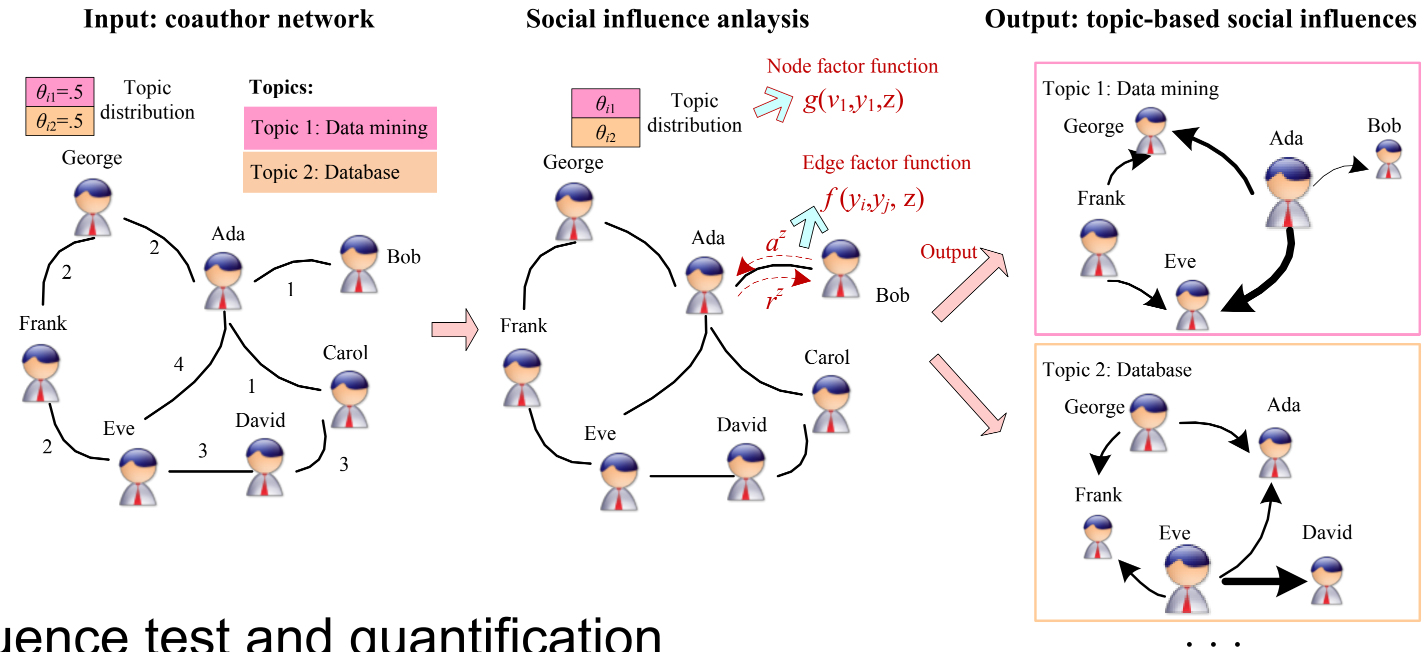
Related Work—Conformity

- Conformity theory
  - Compliance, identification, and internalization [Kelman 1958]
  - A theory of conformity based on game theory [Bernheim 1994]
- Influence and conformity
  - Conformity-aware influence analysis [Li-Bhowmick-Sun 2011]
- Applications
  - Social influence in social advertising [Bakshy-el-al 2012]

"All those in favour say 'Aye'."

"Aye." "Aye." "Aye." "Aye." "Aye."

Related Work—social influence



- Influence test and quantification
  - Influence and correlation [Anagnostopoulos-et-al 2008]
    Distinguish influence and homophily [Aral-et-al 2009, La Fond-Nevill 2010]
  - Topic-based influence measure [Tang-Sun-Wang-Yang 2009, Liu-et-al 2012]
    Learning influence probability [Goyal-Bonchi-Lakshmanan 2010]
- Influence diffusion model
  - Linear threshold and cascaded model [Kempe-Kleinberg-Tardos 2003]
  - Efficient algorithm [Chen-Wang-Yang 2009]

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?

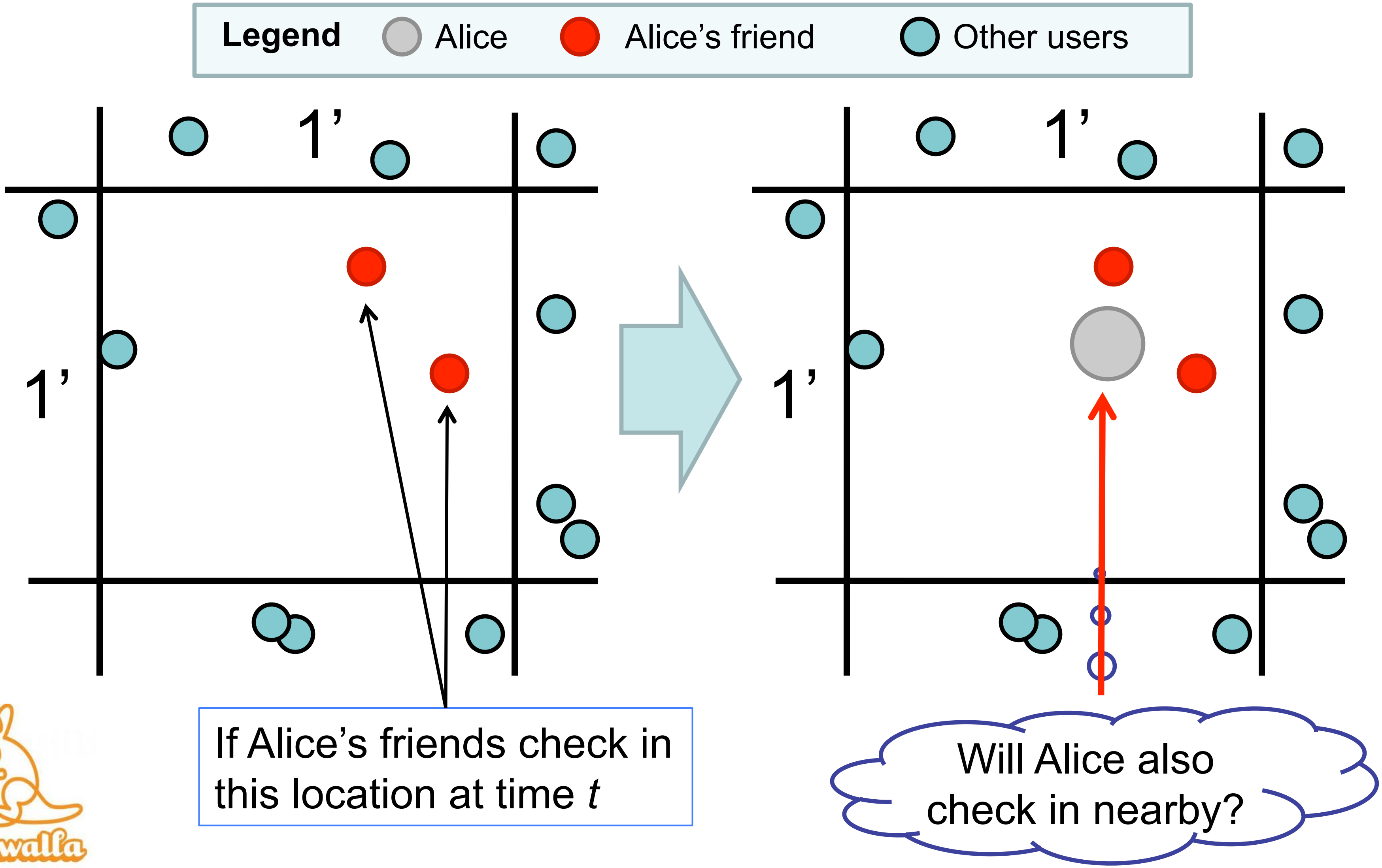
Problem Formulation and Methodologies

Four Datasets

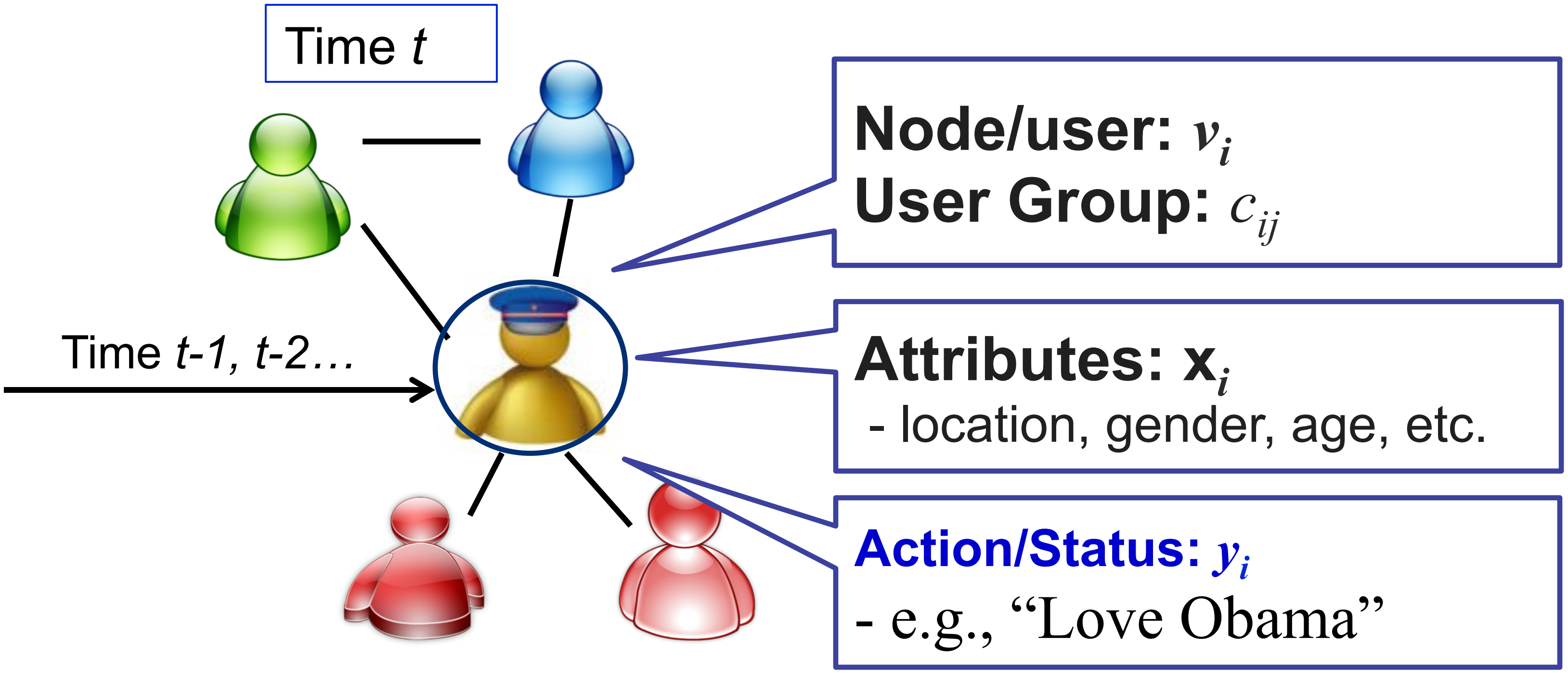
Network	#Nodes	#Edges	Behavior	#Actions
Weibo	1,776,950	308,489,739	Tweet on popular topics	6,761,186
Flickr	1,991,509	208,118,719	Comment on a popular photo	3,531,801
Gowalla	196,591	950,327	Check-in some location	6,442,890
ArnetMiner	737,690	2,416,472	Publish in a specific domain	1,974,466

All the datasets are publicly available for research.

A concrete example in Gowalla



Notations



$$G = (V, E, C, X)$$

$$\mathbf{A} = \{(a, v_i, t)\}_{a,i,t}$$

— each (a, v_i, t) represents user v_i performed action a at time t

Conformity Definition

- Three levels of conformities
  - Individual conformity
  - Peer conformity
  - Group conformity

Individual Conformity

- The individual conformity represents how easily user v's behavior conforms to her friends

A specific action performed by user v at time t

Exists a friend v' who performed the same action at time t"

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

All actions by user v

Peer Conformity

- The peer conformity represents how likely the user v's behavior is influenced by one particular friend v'

A specific action performed by user v' at time t'

User v follows v' to perform the action a at time t

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

All actions by user v'

Group Conformity

- The group conformity represents the conformity of user v's behavior to groups that the user belongs to.

***τ-group action:* an action performed by more than a percentage τ of all users in the group C_k**

A specific τ-group action

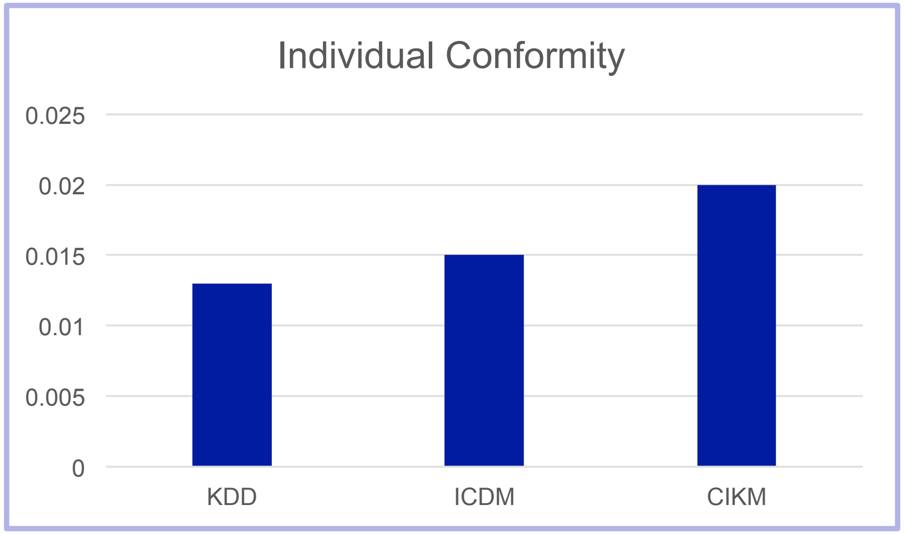
User v conforms to the group to perform the action a at time t

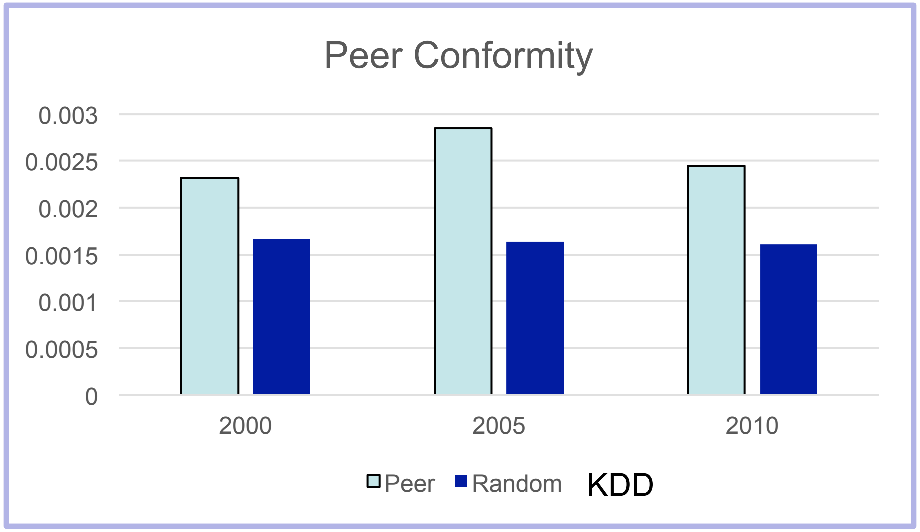
$$gcf^{\tau}(v, C_{vk}) = \frac{\left| \boxed{(a, v', t')} \in A^{\tau}_{C_k} \boxed{\left| \exists (a, v, t) : \mathbb{I}[c_{ik}] \wedge \epsilon \geq t - t' \geq 0 \right|} \right.}{\boxed{\left| A^{\tau}_{C_k} \right|}}$$

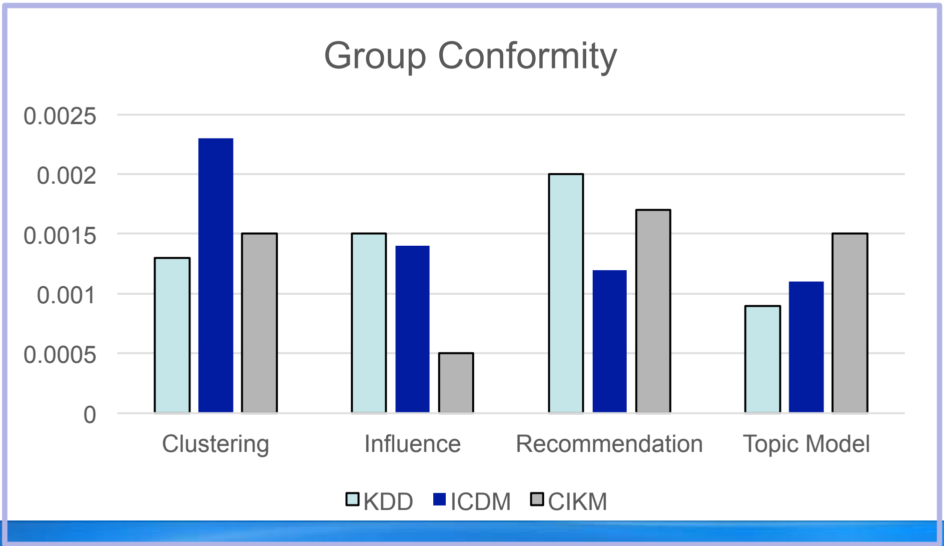
All τ-group actions performed by users in the group C_k

For an example

Conformity in the Co-Author Network







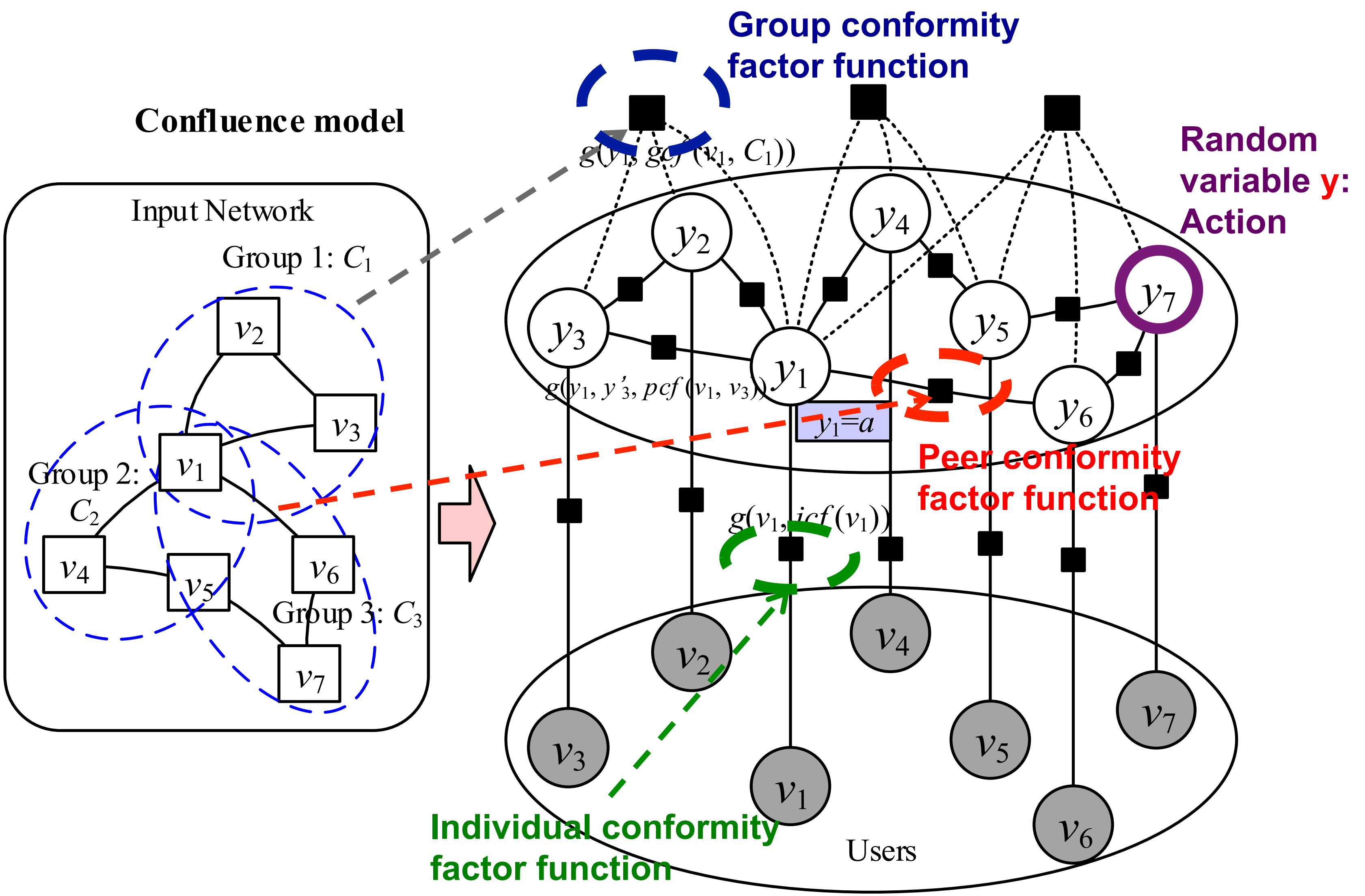
Now our problem becomes

- How to incorporate the different types of conformities into a unified model?

Input:
$G=(V, E, C, X), A$

Output:
$F: f(G, A) \rightarrow Y^{(t+1)}$

Confluence
—A conformity-aware factor graph model



Model Instantiation

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

Individual conformity
factor function

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

Peer conformity
factor function

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

Group conformity
factor function

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

General Social Features

- Opinion leader[1]
  - Whether the user is an opinion leader or not
- Structural hole[2]
  - Whether the user is a structural hole spanner
- Social ties[3]
  - Whether a tie between two users is a strong or weak tie
- Social balance[4]
  - People in a social network tend to form balanced (triad) structures (like “my friend’s friend is also my friend”).

[1] X. Song, Y. Chi, K. Hino, and B. L. Tseng. Identifying opinion leaders in the blogosphere. In **CIKM’06**, pages 971–974, 2007.
[2] T. Lou and J Tang. Mining Structural Hole Spanners Through Information Diffusion in Social Networks. In **WWW'13**. pp. 837-848.
[3] M. Granovetter. The strength of weak ties. American Journal of Sociology, 78(6):1360–1380, 1973.
[4] D. Easley and J. Kleinberg. Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press, 2010.

Distributed Model Learning

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.

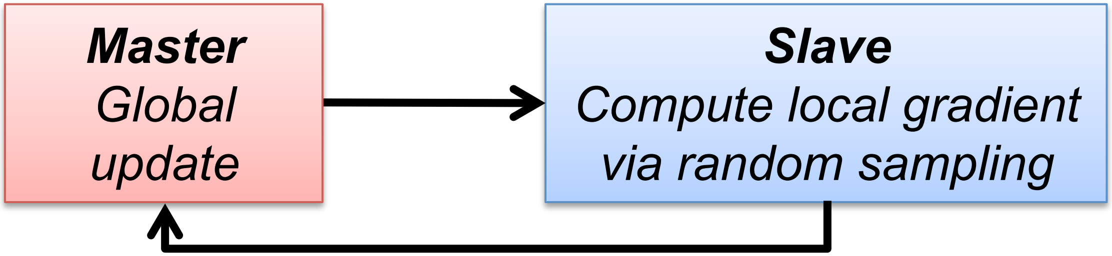
Unknown parameters to estimate

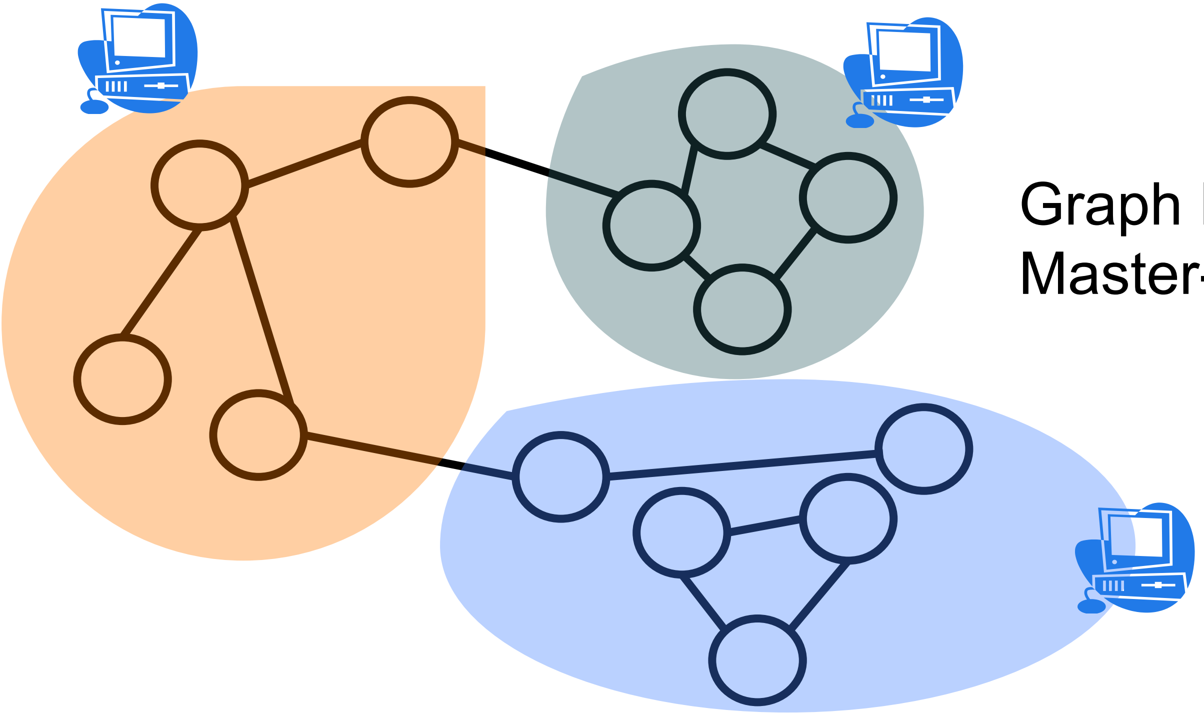
(1) Master

(2) Slave

(3) Master

Distributed Learning





Graph Partition by Metis
Master-Slave Computing

Inevitable loss of correlation factors!

Experiments

Data Set and Baselines

Network	#Nodes	#Edges	Behavior	#Actions
Weibo	1,776,950	308,489,739	Post a tweet	6,761,186
Flickr	1,991,509	208,118,719	Add comment	3,531,801
Gowalla	196,591	950,327	Check-in	6,442,890
ArnetMiner	737,690	2,416,472	Publish paper	1,974,466

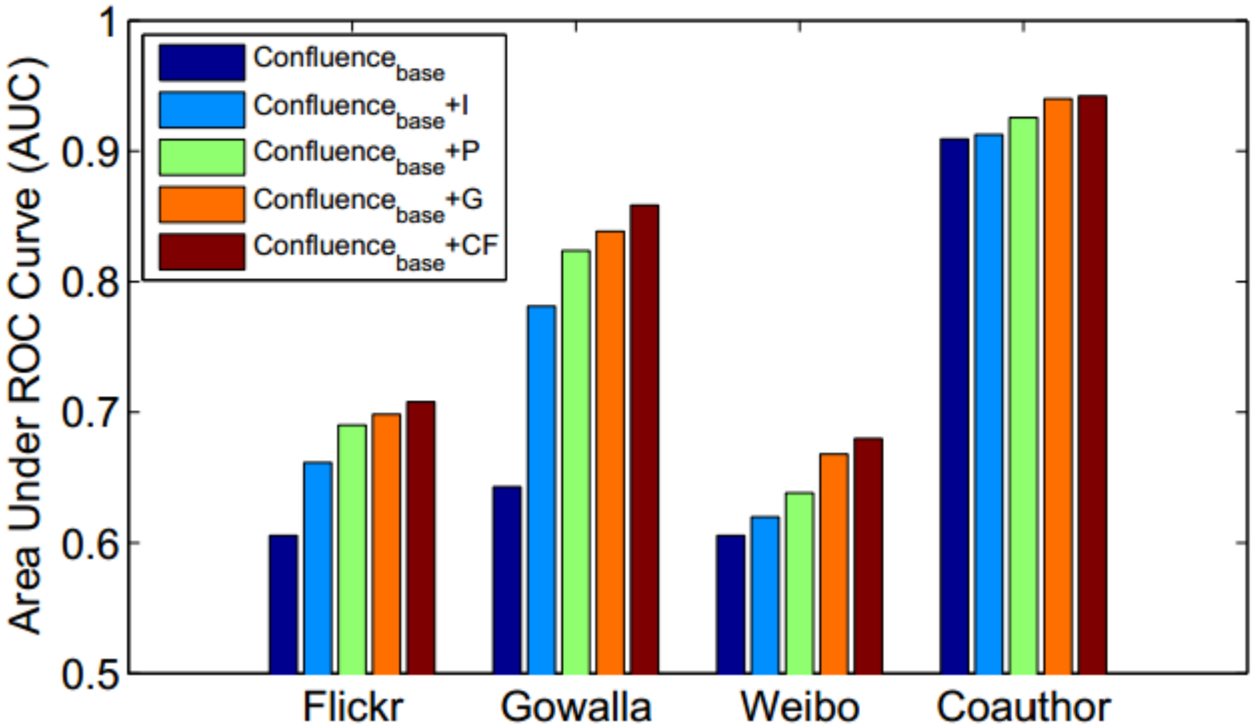
- Baselines
  - *Support Vector Machine (SVM)*
  - *Logistic Regression (LR)*
  - *Naive Bayes (NB)*
  - *Gaussian Radial Basis Function Neural Network (RBF)*
  - *Conditional Random Field (CRF)*
- Evaluation metrics
  - *Precision, Recall, F1, and Area Under Curve (AUC)*

Prediction Accuracy

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)**

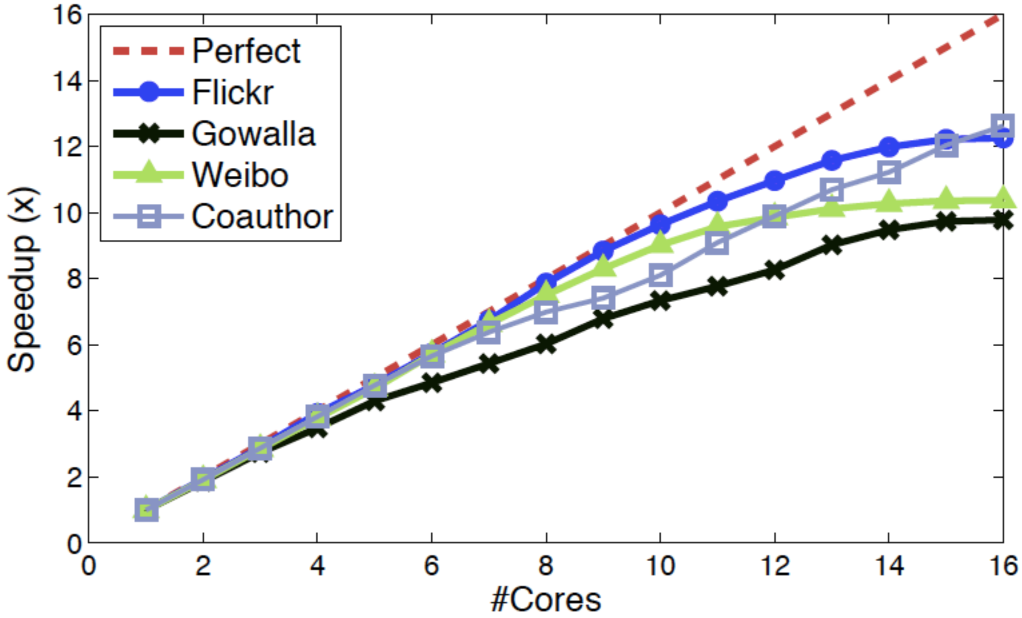
t-test, p<<0.01

Effect of Conformity



Confluence$_{base}$ stands for the Confluence method without any social based features
Confluence$_{base}$+I stands for the Confluence$_{base}$ method plus only individual conformity features
Confluence$_{base}$+P stands for the Confluence$_{base}$ method plus only peer conformity features
Confluence$_{base}$+G stands for the Confluence$_{base}$ method plus only group conformity

Scalability performance



Achieve ~ 9×speedup with 16 cores

Table 4: 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

Conclusion

- Study a novel problem of conformity influence analysis in large social networks
- Formally define three conformity functions to capture the different levels of conformities
- Propose a Confluence model to model users' actions and conformity
- Our experiments on four datasets verify the effectiveness and efficiency of the proposed model

Future work

- Connect the conformity phenomena with other social theories
  - e.g., social balance, status, and structural hole
- Study the interplay between conformity and reactance
- Better model the conformity phenomena with other methodologies (e.g., causality)

Confluence: Conformity Influence in Large Social Networks

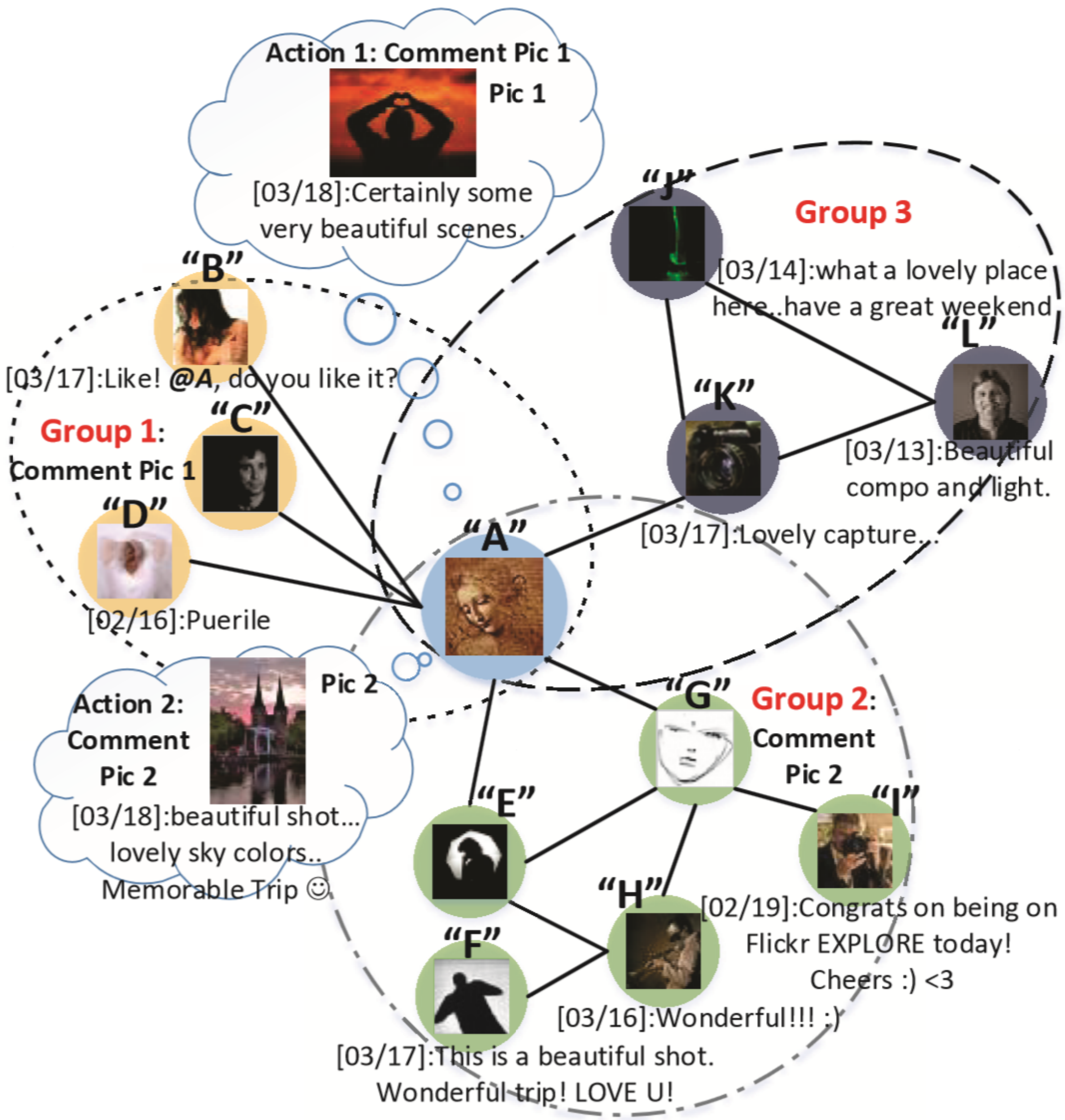
Jie Tang[*], Sen Wu[*], and Jimeng Sun[+]

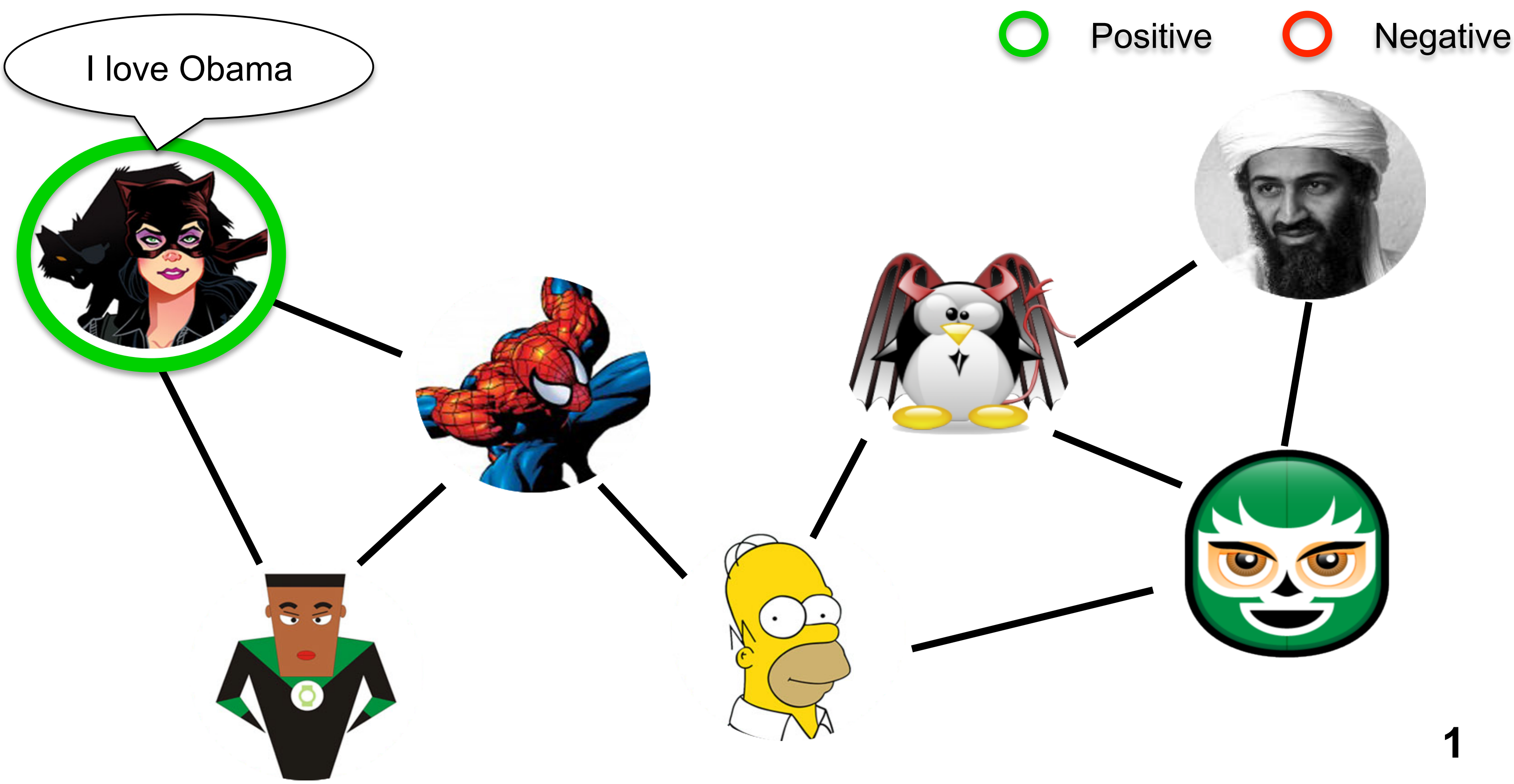
[*]Tsinghua University

[+]IBM TJ Watson Research Center

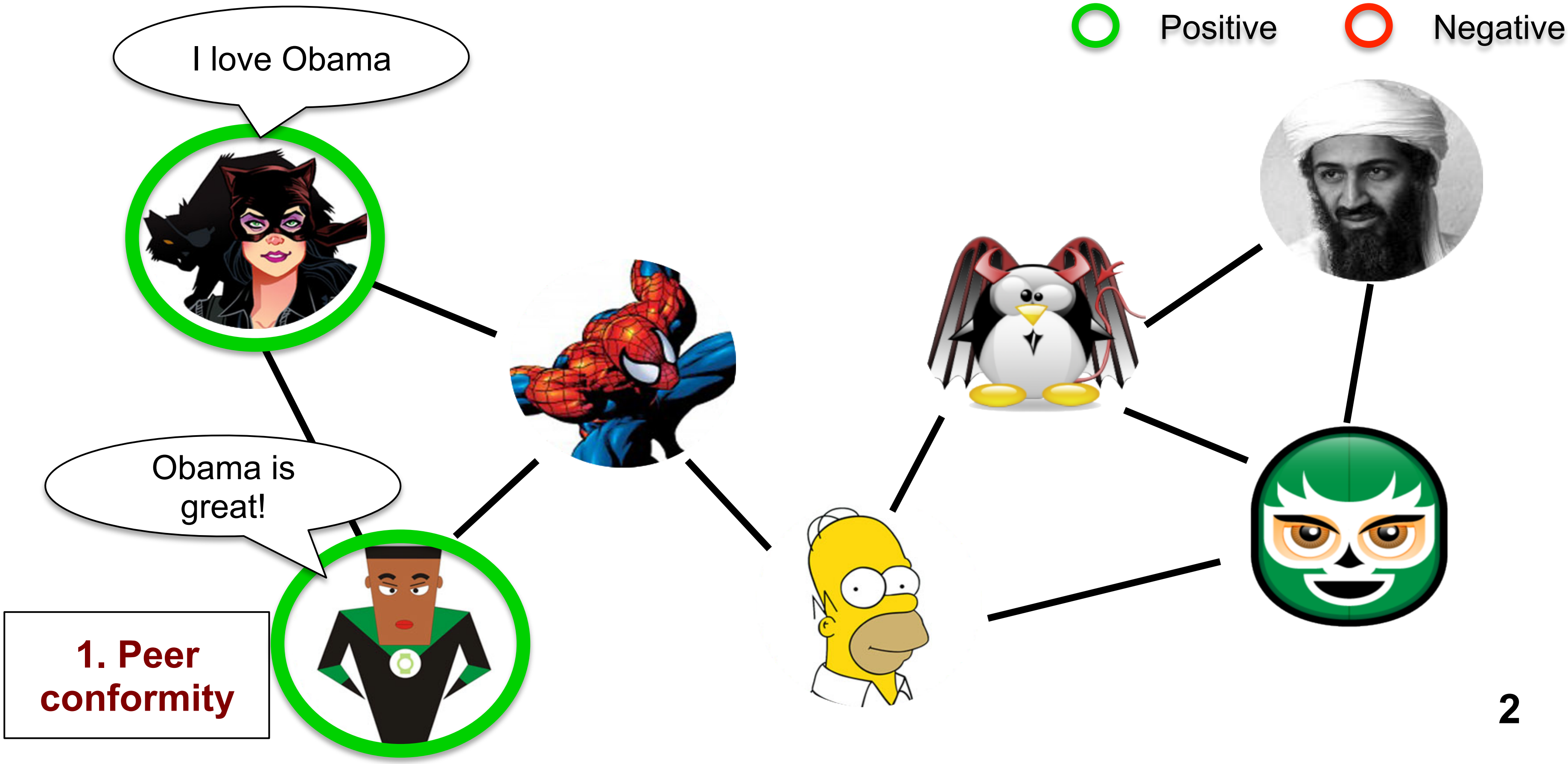
Data and codes are available at: http://arnetminer.org/conformity/

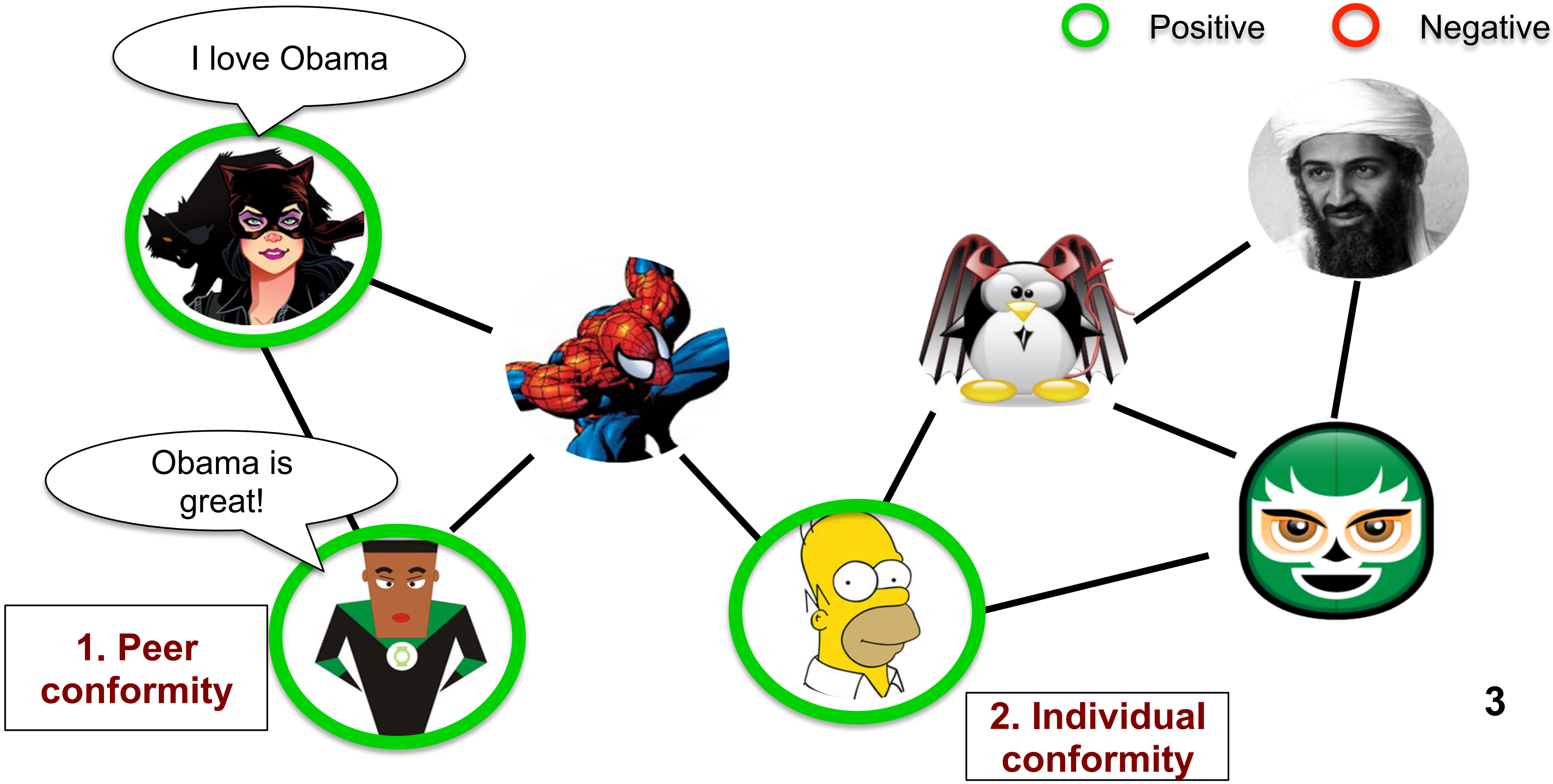
Qualitative Case Study

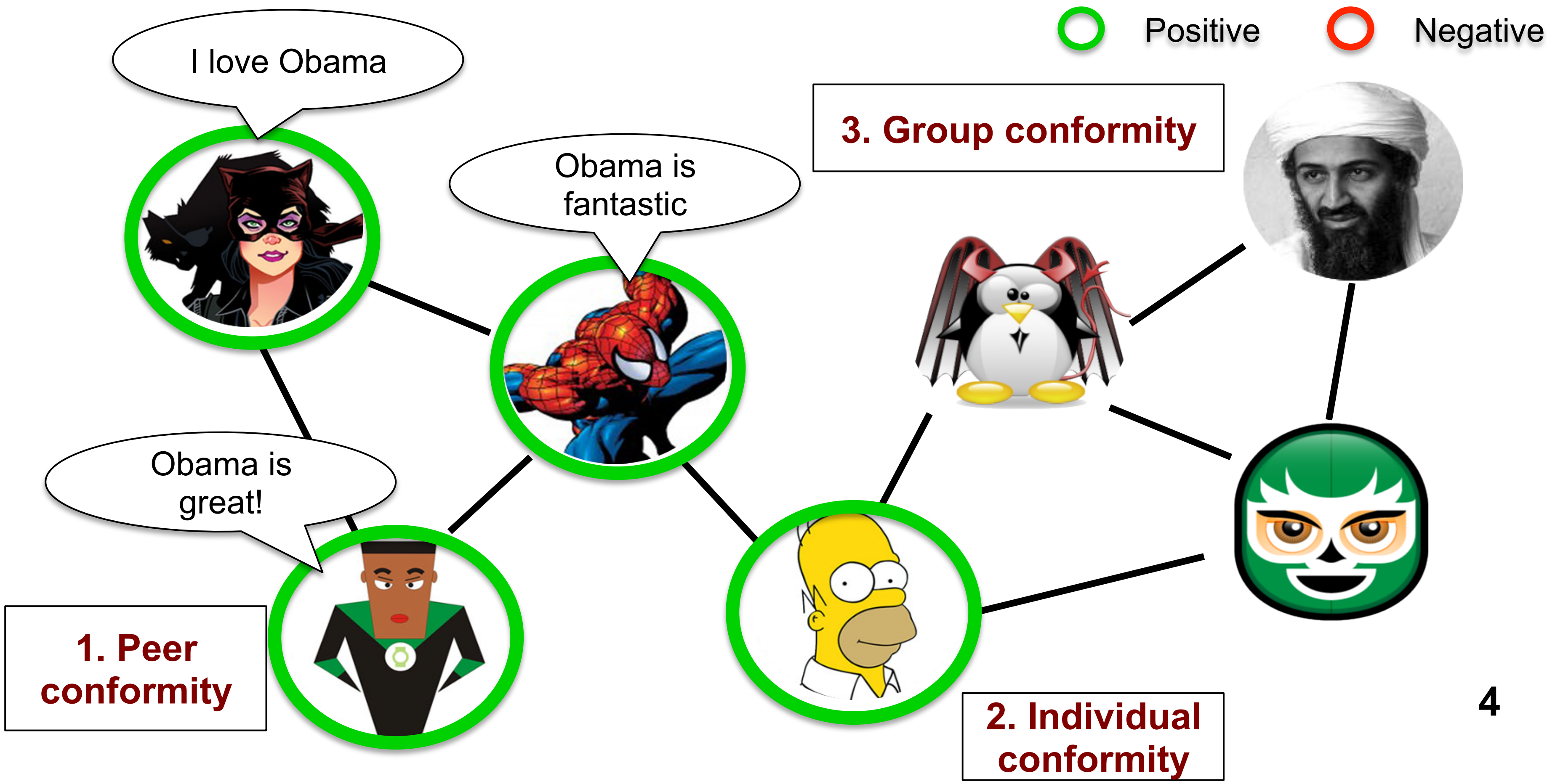




Positive
Negative
I love Obama
1


Positive
Negative
I love Obama
Obama is great!
1. Peer conformity
2


Positive
Negative
I love Obama
Obama is great!
1. Peer conformity
2. Individual conformity
3


Positive
Negative
I love Obama
Obama is fantastic
Obama is great!
1. Peer conformity
2. Individual conformity
3. Group conformity
4