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.

"Love Obama"

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

Positive  Negative
Conformity Influence Analysis

1. Peer conformity
   - I love Obama
   - Obama is fantastic
   - Obama is great!

2. Individual conformity
   - Positive
   - Negative

3. Group conformity
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]
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]
  - 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

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

All the datasets are publicly available for research.
A concrete example in Gowalla

Legend

- Grey circle: Alice
- Red circle: Alice's friend
- Blue circle: Other users

If Alice’s friends check in this location at time $t$, will Alice also check in nearby?
Notations

\[ G = (V, E, C, X) \]

\[ A = \{(a, v_i, t)\}_{a,i,t} \]

— 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

\[
i_{cf}(v) = \frac{|\{(a, v, t) \in A_v | \exists (a, v', t') : e_{vv'} \in E \land \epsilon \geq t - t' \geq 0\}|}{|A_v|}
\]

**A specific action performed by user v at time t**

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

**All actions by user v**
Peer Conformity

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

\[ pcf(v, v') = \left| \{ (a, v', t') \in A_v | \exists (a, v, t) : e_{vv'} \in E \land \epsilon \geq t - t' \geq 0 \} \right| \]

- A specific action performed by user $v'$ at time $t'$.
- User $v$ follows $v'$ to perform the action $a$ at time $t$.
- 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.

\( \tau \)-group action: an action performed by more than a percentage \( \tau \) of all users in the group \( C_k \)

A specific \( \tau \)-group action

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

All \( \tau \)-group actions performed by users in the group \( C_k \)
For an example

Conformity in the Co-Author Network

**Individual Conformity**
- KDD
- ICDM
- CIKM

**Peer Conformity**
- 2000
- 2005
- 2010
- KDD

**Group Conformity**
- Clustering
- Influence
- Recommendation
- Topic Model

KDD | ICDM | CIKM
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

\[ g(v_1, \text{icf}(v_1)) \]

\[ g(v_1, \text{gcf}(v_1, C_1)) \]

Group conformity factor function

Random variable \( y \): Action

Peer conformity factor function

Individual conformity factor function

Input Network

Group 1: \( C_1 \)

\( v_1 \)

\( v_2 \)

\( v_3 \)

Group 2: \( C_2 \)

\( v_4 \)

\( v_5 \)

Group 3: \( C_3 \)

\( v_6 \)

\( v_7 \)

Users

\[ g(v_1, y_1, \text{pcf}(v_1, v_3)) \]

\( y_1 = a \)
Model Instantiation

\[ O(\theta) = \log P_\theta(Y|G, A) \]

\[
= \sum_{i=1}^{N} \left[ \sum_{j=1}^{d} \alpha_{ij} f(y_i, x_{ij}) + \beta_i g(y_i, icf(v_i)) \right] + \sum_{e_{ij} \in E} \mathbb{I}[y'] \gamma_{ij} g(y_i, y_j', pcf(v_i, v_j)) \\
+ \sum_{i=1}^{N} \sum_{k=1}^{m} \mathbb{I}[c_{ik}] \mu_{ik} g(y_i, gcf(v_i, C_k)) - \log Z
\]

\[ g(y_i, gcf^\tau(v_i, C_k)) = \left( \frac{1}{2} \right)^{t-t'} gcf^\tau(v_i, C_k) \]

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

Individual conformity factor function

Peer conformity factor function

Group conformity factor function
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”).

---


Distributed Model Learning

\textbf{Input:} network $G$, action history $A$, and learning rate $\eta$;
\textbf{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 $\theta$ to all Slaves;
\for $l = 1$ \text{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 $\alpha_j$,
\[ \alpha_j^{\text{new}} = \alpha_j^{\text{old}} + \eta \frac{\partial \theta}{\partial \alpha_j} \]
\until convergence;

\textbf{Algorithm 1:} Distributed model learning.
Distributed Learning

Master
Global update

Slave
Compute local gradient via random sampling

Graph Partition by Metis
Master-Slave Computing

Inevitable loss of correlation factors!
Experiments
Data Set and Baselines

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

- 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

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

---

**t-test, p<<0.01**
Effect of Conformity

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

Achieve $\sim 9 \times$ speedup with 16 cores

Table 4: Running time of the proposed algorithm (hour).

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Flickr</th>
<th>Gowalla</th>
<th>Weibo</th>
<th>Co-Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confluence</td>
<td>1.602</td>
<td>0.245</td>
<td>1.083</td>
<td>0.512</td>
</tr>
<tr>
<td>Confluence (single)</td>
<td>19.637</td>
<td>2.395</td>
<td>11.229</td>
<td>6.464</td>
</tr>
<tr>
<td>CRF</td>
<td>3.864</td>
<td>0.387</td>
<td>2.547</td>
<td>1.823</td>
</tr>
</tbody>
</table>
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

Action 1: Comment Pic 1
Pic 1
[03/18]: Certainly some very beautiful scenes.

Group 1: Comment Pic 1
“B”
[03/17]: Like! @A, do you like it?
“C”
[02/16]: Puerile

Action 2: Comment Pic 2
Pic 2
[03/18]: Beautiful shot... lovely sky colors... Memorable Trip 😊

Group 2: Comment Pic 2
“E”
[03/16]: Wonderful!!! 😊
[03/17]: This is a beautiful shot. Wonderful trip! LOVE U!

“F”
[02/19]: Congrats on being on Flickr EXPLORE today! Cheers :) <3

“G”
[03/13]: Beautiful compo and light.

Group 3
“K”
[03/14]: What a lovely place here, have a great weekend.
“L”

“H”
I love Obama

Positive  Negative

Peer Conformity

Individual Conformity

1
I love Obama

Obama is great!

1. Peer conformity

Positive

Negative

2
I love Obama

Obama is great!

1. Peer conformity

2. Individual conformity

Positive

Negative

3
1. Peer conformity

2. Individual conformity

3. Group conformity

I love Obama

Obama is fantastic

Obama is great!

Positive

Negative