Role-aware Conformity Influence Modeling and Analysis in Social Networks

Jing Zhang[†], **Jie Tang**[†], **Honglei Zhuang**[‡], **Cane Wing-Ki Leung**[♯] and **Juanzi Li**[†]
[†]Department of Computer Science and Technology, Tsinghua University
[‡]Department of Computer Science, University of Illinois at Urbana-Champaign
[♯]Huawei Noah's Ark Lab
zhangjing12@mails.tsinghua.edu.cn, {jietang, lijuanzi}@tsinghua.edu.cn, hzhuang3@illinois.edu, cane.leung@huawei.com

Abstract

Conformity influence is the inclination of a person to be influenced by others. In this paper, we study how the conformity tendency of a person changes with her *role*, as defined by her structural properties in a social network. We first formalize conformity influence using a utility function based on the conformity theory from social psychology, and then propose a probabilistic graphical model, referred to as Role-Conformity Model (RCM), for modeling the role-aware conformity influence between users by incorporating the utility function. We apply the proposed RCM to several academic research networks, and discover that people with higher degree and lower clustering coefficient are more likely to conform to others. We also evaluate RCM through the task of word usage prediction in academic publications, and show significant improvements over baseline models.

1 Introduction

In social networks, conformity influence is the inclination of a person to be influenced by others by yielding to perceived group pressure and copying the behavior and beliefs of others. The earliest study on conformity influence dates back to the 1930's by social psychologists (Jenness 1932; Sherif 1935). Since then, precedent work extensively studied how conformity affects individuals' actions. The well-known experiments in (Asch 1955) showed that over 75% of people tend to conform to others in varying degrees. Existing work (Bernheim 1994; Cialdini and Goldstein 2003; Kelman 1958; Aronson, Wilson, and Akert 2007) has repeatedly verified the significant effect of conformity influence in our social life.

With the rapid proliferation of online social networks such as Facebook and Twitter, quantitatively estimating the conformity tendency of each individual becomes more and more critical for applications such as viral marketing, social influence maximization, etc. Yet, research on conformity influence in online social networks is just beginning. For instance, Li et al. (Li, Bhowmick, and Sun 2011; 2013) studied the interplay between the influence and conformity of an individual, while Tang et al. (Tang, Wu, and Sun 2013) proposed a probabilistic factor graph model that takes the effect of conformity into account when predicting user behavior. Both focus on modeling the effect of conformity influence at the individual level.

In this paper, we explore whether it is possible to summarize several "roles", i.e., prototypes, to concisely describe the correlation between the conformity tendency of an individual and her structural features in a network. Such role-based modeling and analysis of conformity influence is beneficial for applications such as recommendations, where data sparsity or the cold-start problem cannot be overlooked. It is also the key difference from existing social contagion studies (Kempe, Kleinberg, and Tardos 2003; Gruhl et al. 2004), which ignore the effect of roles in information diffusion.

There are several questions to address when modeling the conformity tendency of individuals based on their roles. First, how to formalize the conformity theory in social psychology? Second, how can roles be incorporated into such formalization to model user actions? Furthermore, how can role-based conformity be used to real applications? We summarize our answers to these questions and our main results in what follows.

Results We formalize conformity influence in terms of a utility function based on the conformity theory (Bernheim 1994), and justify the proposed utility function by proving the existence of Nash equilibria when users conduct actions according to it. We further incorporate the utility function into an application-oriented probabilistic model, known as the Role-Conformity Model (RCM), to describe user behaviors. To the best of our knowledge, this is the first attempt to formally connect the conformity theory from social psychology to a computational model.

We apply the proposed model on several academic networks to observe correlations between people's latent roles and their conformity tendency. Interestingly, people with higher degree and lower clustering coefficient are more likely to conform to others. The phenomenon may be explained as that when a person has more collaborators with the structure among them more diverse (i.e., the collaborations between the neighbors in the local network are infrequent), she may become more open-minded and tend to accept new ideas from others. However, when the social circle is restricted to a few collaborators, the person will limit her mind and tend not to accept other ideas.

We evaluate the proposed RCM through the task of word usage prediction, and results indicate that our model performs much better (+3.6-4.1% in terms of average MAP and +7.1-47.4% in terms of average AUC) than the basic TF-IDF and PLSA.

2 Formalizing Conformity

In this section, we formalize conformity influence based on the conformity theory from social psychology in terms of a utility function, and prove the existence of Nash equilibria if all users in a network behave according to it.

Conformity utility function The conformity theory suggests that heterogeneous preferences do result in heterogeneous behaviors (Bernheim 1994). Everyone in a group expresses her own individuality. Yet, even individualists succumb somewhat to the desire for status (esteem or popularity) and shade their choices toward the social norm. This is because people seeking status care about how someone else feels about them through their actions. They are therefore willing to suppress their individuality and conform to the social norm, worrying that even small departures from the social norm may seriously impair their popularity.

We formalize the conformity theory in terms of a utility function. We use a binary value to represent whether a user v_i adopts an action ($y_i = 1$) or not ($y_i = 0$). Given the decision y_i, we model the utility v_i obtained from her decision from two aspects. One is the individual's intrinsic utility in the absence of all other neighbors, the other is the esteem acquired through conforming:

$$f(y_i) = (1 - \lambda_i)\, d(y_i, \hat{y}_i) + \lambda_i \sum_{j \in N(i)} d(y_i, y_j) \quad (1)$$

where $\hat{y}_i$ represents the intrinsic preferred selection of user v_i, λ_i represents the conformity tendency of v_i, and $N(i)$ denotes the neighbors of v_i at the time when v_i makes the decision. $d(.,.)$ is a metric that gives a utility of 1 when two decisions are the same, and 0 otherwise.

Nash equilibria We provide an induction method to prove that there exists Nash equilibria if all users in a network make the decisions for a given action according to the utility function defined by Eq. (1). For brevity, we assume that the parameter λ in Eq. (1) is fixed for different users. The proof is the same for different λ.

The proof is straightforward when there is only one user in a network. For a network with two users, when their intrinsic preferred selections are the same, a Nash equilibrium exists because they will make the same decision. When their preferred selections are different, λ determines the final selection. If $\lambda < 0.5$, a Nash equilibrium exists because they will select their own preferences respectively. If $\lambda > 0.5$, two Nash equilibria exist because they will both select $\hat{y}_1$ or $\hat{y}_2$.

Finally, we prove that if a Nash equilibrium exists in a network with k users (k-network), a Nash equilibrium will definitely exist in any $(k+1)$-network obtained by adding an additional user, v_{k+1}, to it. The general idea is to investigate whether the neighbors of v_{k+1} will change their decisions when v_{k+1} joined a k-network that has already arrived at a Nash equilibrium.

We assume that the preferred selection of v_{k+1} is 1, i.e., $\hat{y}_{k+1} = 1$. The proof is the same when $\hat{y}_{k+1} = 0$. Given an existing k-network, we denote the number of v_{k+1}'s neighbors with $y = 1$ as N^1, and the number of v_{k+1}'s neighbors with $y = 0$ as N^0. Thus, the utility of v_{k+1} is calculated as:

$$f(v_{k+1}) = \begin{cases} (1-\lambda) + \lambda N^1 & \text{if } y_{k+1} = 1 \\ \lambda N^0 & \text{if } y_{k+1} = 0 \end{cases}$$

The utility of a neighbor v_i of v_{k+1} is represented as:

$$f(y_i) = (1-\lambda)\, d(y_i, \hat{y}_i) + \lambda \sum_{j \in N(i)} d(y_i, y_j) + \lambda\, d(y_i, y_{k+1})$$

Suppose $(1-\lambda) + \lambda N^1 > \lambda N^0$, v_{k+1} will decide to adopt the action (i.e., $y_{k+1} = 1$). The proof is the same when $(1-\lambda) + \lambda N^1 < \lambda N^0$.

We observe that the neighbors with $y_i = 1$ will not change their decisions. Otherwise, the utility obtained from the k-network will decrease because the Nash equilibrium is damaged, and the utility obtained from v_{k+1} will also decrease because y_i is changed differently from y_{k+1}.

For the neighbors with $y_i = 0$, if they change their decisions, the marginal utility is $\lambda - c_i$, where $-c_i$ is the decreased utility triggered from the k-network because the Nash equilibrium is damaged. λ is the increased utility caused by v_{k+1} because y_i is changed to be the same as y_{k+1}. If $\lambda \leq c_i$, none of the neighbors will change their decisions. If $\lambda > c_i$, the neighbors will change their decisions. However, in such situation, v_{k+1} will not change back to 0, because the utility will be reduced from $(1-\lambda) + \lambda(N^1 + 1)$ to $\lambda(N^0 - 1)$.

To summarize, we can find a Nash equilibrium when an additional user v_{k+1} is added to any k-network with a Nash equilibrium already arrived.

3 Role-Conformity Model (RCM)

The aforementioned conformity utility function presents elegant theoretical properties, although it is too simple for real cases. In this section, we further extend it into an application-oriented probabilistic model, named Role-Conformity Model (RCM), to describe user behaviors. We introduce in the model discrete time slices from $t = 1$ to T, and two hidden variables for characterizing the "role" of a user as well as the "topic" of a certain action.

Definition 1 ***Individual attributes*** *At time slice t, each user v_i is associated with an attribute vector of length H, where the h-th attribute's value is denoted by $x_{i,t,h}$. Different network properties of v_i, such as clustering coefficient and degree, can be used as individual attributes, with the choice of which being application-dependent.*

Definition 2 ***Role distribution*** *We adopt the concept of "role" to summarize user attributes into several clusters. A user can play different roles at different time slices. Formally, we associate each user v_i at each time slice t with a vector $\rho_{i,t} \in \mathbf{R}^R$, where R is the number of roles in the model ($\sum_{r=1}^{R} \rho_{i,t}^r = 1$). Each element $\rho_{i,t}^r$ is the probability that user v_i belongs to role r at t.*

Definition 3 ***Topic distribution*** *In social networks, a user is usually interested in multiple topics. Formally, each user v_i is associated with a vector $\theta_i \in \mathbf{R}^K$, where K is the number of topics ($\sum_{z=1}^{K} \theta_i^z = 1$). Each θ_i^z is the probability (intrinsic preference) of user v_i choosing topic z.*

Model description Based on the above definitions, we explain the proposed Role-Conformity Model. The basic idea is that users' role distribution is determined by not only attributes but also actions. We use users' attributes to determine her role distribution, which is then used as a prior to guide the sampling process for users' actions. Specifically, the model consists of two parts. The first part models the generation of individual attributes. For an individual attribute, we first draw a role r from a multinomial distribution, and then draw the value of the attribute from a normal distribution with respect to r. The second part models the total utility of generating all the actions. Specifically, we extend the utility function in Eq. (1) to further incorporate the role and topic distributions of a user. Instead of binary actions, each user is now allowed to take a set of actions, denoted by $W = \{w\}$. When in role r, the utility function of v_i taking an action w is defined as:

$$\gamma_{i,r}^{w} = \left[(1-\lambda_r)\sum_{z=1}^{K}\theta_i^z\phi_z^w + \lambda_r\frac{1}{|N_i|}\sum_{j\in N_i}\sum_{z=1}^{K}\theta_j^z\phi_z^w\right] \quad (2)$$

where ϕ_z^w is a non-negative score of taking action w under topic z, satisfying $\sum_w \phi_z^w = 1$, N_i is the (directed) neighborhood of v_i in the social network, and λ_r is a weighting factor similar to λ_i in Eq. (1). Note that we utilize a unified λ_r for all the users in role r, to reduce the number of parameters. Since the neighborhood of different users in r can be different, we normalize the gain a user obtained from her neighbors by her neighborhood size. This modification does not affect the conclusion of Section 2 since it is equivalent to directly assigning the individual conformity tendency of each user v_i as $\lambda_i = \frac{\lambda_r}{|N_i|-(|N_i|-1)\lambda_r}$.

The extended utility function is more general than the one described in Eq. (1). It also has another probabilistic interpretation and can be regarded as the likelihood of generating action w, by tossing a coin s with distribution $Bern(\lambda_r)$. Then, if $s = 1$, w is determined by the individual's intrinsic topic distribution and is drawn from $P(w|i) = \sum_z \theta_i^z\phi_z^w$. Otherwise, w is influenced by the neighbors' topic distribution and is drawn from $P(w|N_i) = \sum_{j\in N_i}\sum_z \theta_j^z\phi_z^w/|N_i|$.

The complete generative process is summarized as:

- For the h-th attribute of user v_i at time slice t:
  - Draw a role r from multinomial distribution $\rho_{i,t}$;
  - Draw the value of the attribute $x_{i,t,h} \sim N(\mu_{r,h}, \sigma_{r,h})$.
- For an action w conducted by user v_i at time t:
  - Draw v_i's role r from multinomial distribution $\rho_{i,t}$;
  - Obtain the utility of action w denoted by $\gamma_{i,r}^w$ (or apply the probabilistic interpretation here).

For a given data set, we need to learn the parameters $\rho_{i,t}$, $\mu_{r,h}$, $\sigma_{r,h}$, as well as θ_i, ϕ_z and λ_r. We provide an application example of RCM in what follows.

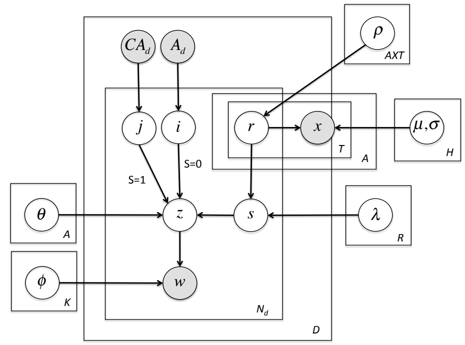


Figure 1: Role-conformity model.

Application example: We are given a bibliographic dynamic network $G^t = (V^t, E^t)$, where V^t is the set of authors up to time t and E^t is the set of coauthor relationships among them up to time t. Each author $v_i \in V^t$ is associated with an attribute vector $\mathbf{x}_i$, containing her individual attributes in the network. There is also a set of documents D, where each $d \in D$ can be represented by (A_d, W_d, C_d, t). $A_d \subset V^t$ stands for the author set of d, W_d is the list of words w in d, $C_d \subset D$ is the set of documents cited by d, and t is the time slice when d is published.

By regarding each word as an action, we can naturally plug this data set into RCM. The only concern is that we usually do not know which author wrote down which word w in a document d with multiple authors, thus we assume that each word is generated from an author randomly chosen from A_d. We also need to prudently define the neighborhood N_i of each author. Since the author is more likely to be influenced by the documents she is citing, we model N_i for author v_i in document d by $CA_d = \bigcup_{d'\in C_d} A_{d'}$. In this application, the individual features are defined based on the co-author network. The neighbors an author conforms with can also be their coauthors. However, we discover that conformity influence caused by coauthor relationships is not as significant as that by citation relationships. Thus we investigate conformity influence caused by citation relationships in this paper. We omit the analysis result for space limitation.

Without loss of generality, we continue our discussions based on the bibliographic data set to provide more technical details about how to apply our model in a real application. Figure 1 summarizes the RCM on a bibliographic data set.

Model learning We adopt the probabilistic interpretation of Eq. (2) and use maximum-likelihood estimation (MLE) for model learning. The likelihood for individual attribute generation can be written as:

$$\mathcal{L}_1 = \prod_{i=1}^{A}\prod_{t=1}^{T}\prod_{h=1}^{H}\sum_{r=1}^{R}\frac{\rho_{i,t}^r}{\sqrt{2\pi\sigma_{r,h}^2}}\exp\left[-\frac{(x_{i,t,h}-\mu_{r,h})^2}{2\sigma_{r,h}^2}\right]$$

The likelihood of action generation can be written as:

$$\mathcal{L}_2 = \prod_{d,w}\sum_{i\in A_d}\frac{\sum_{r=1}^{R}\rho_{i,t}^r\gamma_{r,i}^w}{|A_d|}$$

The unified likelihood function is $\mathcal{L} = \mathcal{L}_1\mathcal{L}_2$. It is intractable to directly solve $\mathcal{L}$. Thus we optimize $\mathcal{L}_1$ and $\mathcal{L}_2$

by EM algorithm respectively at each iteration. The product operation provides a theoretical guarantee that the product of lower bounds is also the lower bound of the product. We explain the EM steps for $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively as below.

To optimize $\mathcal{L}_1$, we first estimate the posterior distribution over r for each individual attribute $x_{i,t,h}$ in the E-step:

$$q^r_{i,t,h} = \frac{\frac{\rho^r_{i,t}}{\sqrt{2\pi\sigma^2_{r,h}}} \exp\left[-\frac{(x_{i,t,h}-\mu_{r,h})^2}{2\sigma^2_{r,h}}\right]}{\sum_{r=1}^R \frac{\rho^r_{i,t}}{\sqrt{2\pi\sigma^2_{r,h}}} \exp\left[-\frac{(x_{i,t,h}-\mu_{r,h})^2}{2\sigma^2_{r,h}}\right]}$$

Then in M-step, we update parameters $\mu_{r,h}$, $\sigma_{r,h}$ by:

$$\mu_{r,h} = \frac{\sum_{i=1}^A \sum_{t=1}^T q^r_{i,t,h} x_{i,t,h}}{\sum_{i=1}^A \sum_{t=1}^T q^r_{i,t,h}}$$

$$\sigma_{r,h} = \sqrt{\frac{\sum_{i=1}^A \sum_{t=1}^T q^r_{i,t,h}(x_{i,t,h} - \mu_{r,h})^2}{\sum_{i=1}^A \sum_{t=1}^T q^r_{i,t,h}}}$$

where $\mu_{r,h}$ and $\sigma_{r,h}$ are the mean and variance of the h-th attribute in role r.

In order to optimize the likelihood of action generation $\mathcal{L}_2$, we first apply the E-step as:

$$a^r_{d,w,i} = \frac{\rho^r_{i,t}\gamma^w_{r,i}}{\sum_{r=1}^R \rho^r_{i,t}\gamma^w_{r,i}}$$

$$b_{d,w,i,r} = \frac{\frac{1}{|CA_d|}\sum_{j\in CA_d}\sum_{z=1}^K \theta^z_j \phi^w_z}{\sum_{z=1}^K \theta^z_i \phi^w_z + \frac{1}{|CA_d|}\sum_{j\in CA_d}\sum_{z=1}^K \theta^z_j \phi^w_z}$$

$$c^z_{d,w,i,r} = \frac{\theta^z_i \phi^w_z}{\sum_{z=1}^K \theta^z_i \phi^w_z}$$

where for each w in d conducted by user v_i, $a^r_{d,w,i}$ is the posterior distribution over r, $b_{d,w,i,r}$ is the posterior distribution of conforming, and $c^z_{d,w,i,r}$ is the posterior distribution over topic z. And then in M-step, we update θ^z_i, ϕ^w_z, and λ_r as:

$$\theta^z_i \propto \sum_{d=1}^D \sum_{w\in N_d} \sum_{r=1}^R a^r_{d,w,i} b_{d,w,i,r} c^z_{d,w,i,r} + \sum_{d=1}^D \sum_{w\in N_d} \sum_{j\in A_d} \sum_{r=1}^R \left[a^r_{d,w,i}(1-b_{d,w,i,r}) \sum_{j\in CA_d} c^z_{d,w,j,r} \right]$$

$$\phi^w_z \propto \sum_{d=1}^D \sum_{i\in A_d} \sum_{r=1}^R a^r_{d,w,i} b_{d,w,i,r} c^z_{d,w,i,r} + \sum_{d=1}^D \sum_{i\in A_d} \sum_{r=1}^R \left[a^r_{d,w,i}(1-b_{d,w,i,r}) \sum_{j\in CA_d} c^z_{d,w,j,r} \right]$$

$$\lambda_r = \sum_{d=1}^D \sum_{w\in N_d} \sum_{i\in A_d} \sum_{r=1}^R a^r_{d,w,i} b_{d,w,i,r}$$

where $\sum_z \theta^z_i = 1$ and $\sum_w \phi^w_z = 1$. The parameter $\rho^r_{i,t}$ is derived from both $\mathcal{L}_1$ and $\mathcal{L}_2$:

$$\rho^r_{i,t} = \frac{\sum_{h=1}^H q^r_{i,t,h} + \sum_{d,w} a^r_{d,w,i}}{\sum_{r=1}^R(\sum_{h=1}^H q^r_{i,t,h} + \sum_{d,w} a^r_{d,w,i})}$$

Please refer to the supplementary materials for derivation details.

4 Experiments

In this section, we apply our proposed RCM on a public available academic research data set[1] to investigate the conformity tendency of authors when they write papers.

4.1 Experimental Setup

Data sets We collect the data sets as follows. We first select eight domains from computer science, including database and data mining (DB&DM), human computer interaction (HCI), system and high performance computing (HP), software engineering (SE), computational theory (CT), artificial intelligence and machine learning (AI&ML), computer networks (CN), as well as computer vision and multimedia (CV&MM). For each domain, we then collect all the papers from the well-known journals and conferences in the domain and the citation relationships among them. There are in total 231,728 papers, 269,508 authors and 347,735 citation relationships, where each author has on average 3.44 papers and each paper has on average 1.68 citation relationships.

We design a task of word usage prediction to evaluate our proposed model. The objective of the task is to predict whether a user will write a given word in her paper title in a given time period. Using word usage patterns to study social behaviors has been adopted in existing literature such as (Danescu-Niculescu-Mizil et al. 2013). Specifically, we split each data set into training and test set. The training set contains the papers published in or before 2009, and the test set contains the papers published after 2009. We construct a coauthor network at each time slice and use the degree and clustering coefficient as the individual attributes at each time. Each paper can be viewed as a document with a list of words as the actions performed by the authors. We run our model on the training set and then predict the candidate words (all the words appeared in both the training and test set with stop words removed) that will be used for each candidate user (the user appeared in both the training and test set). The probability of one user using a word, $P(w|i)$, is calculated as the expectation of $\gamma^w_{i,r}$ in Eq. (2) over role r at time t, where t is the ending time of the training set, i.e., 2009 in our setting, and N_i in Eq. (2) is the collection of authors whose papers are cited by user v_i within $[t-\delta, t]$. We empirically set δ as 3 years.

Since the word usage prediction is more like a ranking problem, precision at top ranked results is preferred in evaluating the results. Specifically, given a candidate user v_i, we rank all the candidate words based on $P(w|i)$. We view the co-occurrence of word and user pairs in the test set as the ground truth and use P@5 (Precision of top-5 predictions),

[1] http://arnetminer.org/citation/

P@10, Mean Average Precision (MAP), and area under the ROC curve (AUC) to evaluate the ranking results for each user and then aggregate the results for all the users together.

Baselines We compare our model with TF-IDF, the traditional probabilistic latent semantic analysis (PLSA) (Hofmann 1999) and the citation influence model (CIM) (Dietz, Bickel, and Scheffer 2007).

TF-IDF: In TF-IDF, the probability of a user using a word in the test set is calculated as the TF-IDF value of the user writing the word in the training set. We view a document as the aggregation of all the paper titles of a user to calculate TF-IDF value.

PLSA: In PLSA, the probability of a user using a word is calculated as $P(w|i) = \sum_{z=1}^{K} \theta_i^z \phi_z^w$. This method ignores conformity influence and assumes that users write words only based on their intrinsic preferences.

CTM: In CTM, the probability of one user using a word is calculated as $\gamma_{i,r}^w$ in Eq. (2), where λ_i is directly learned for each user in CTM, instead of for each role in RCM.

4.2 Experimental Results

Table 1 shows the performance of word usage prediction in the collected data sets from eight different domains.

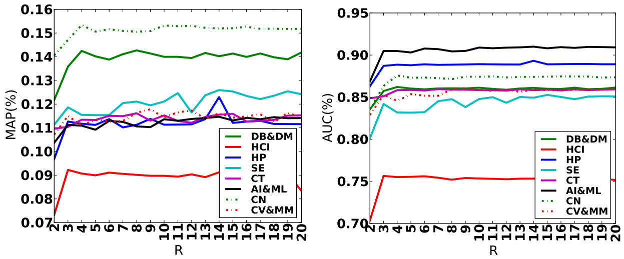
Table 1: Performance of word usage prediction (%).

Query	Method	P@5	P@10	MAP	AUC
DB&DM	TF-IDF	15.84	12.67	6.68	36.20
	PLSA	20.10	15.49	9.26	77.61
	CIM	22.26	17.98	11.59	85.50
	RCM	**30.40**	**24.94**	**14.16**	**86.90**
HCI	TF-IDF	13.57	11.42	5.40	27.59
	PLSA	14.25	11.65	5.71	67.37
	CIM	18.67	15.34	8.12	73.39
	RCM	**19.16**	**15.40**	**8.92**	**75.32**
HP	TF-IDF	15.71	12.95	7.08	38.70
	PLSA	17.33	14.39	8.47	79.96
	CIM	19.62	16.25	10.83	88.67
	RCM	**20.57**	**17.12**	**11.37**	**89.21**
SE	TF-IDF	16.81	13.21	7.82	38.07
	PLSA	4.20	2.60	2.60	81.15
	CIM	21.43	16.42	12.16	**85.55**
	RCM	**25.31**	**19.98**	**12.54**	85.27
CT	TF-IDF	19.18	15.10	11.56	46.80
	PLSA	17.52	13.37	9.88	81.09
	CIM	19.36	14.50	11.04	85.31
	RCM	**20.13**	**15.20**	**11.46**	**85.93**
AI&ML	TF-IDF	19.14	15.39	8.25	42.02
	PLSA	19.92	15.50	9.40	84.10
	CIM	21.24	16.41	10.85	90.70
	RCM	**23.60**	**18.02**	**11.41**	**90.92**
CN	TF-IDF	20.03	17.51	8.71	37.23
	PLSA	26.68	20.33	12.99	80.63
	CTM	29.36	21.62	14.75	86.92
	RCM	**31.20**	**23.35**	**15.22**	**88.41**
CV&MM	TF-IDF	17.19	14.19	8.18	41.65
	PLSA	19.88	14.78	09.64	78.85
	CTM	22.09	16.12	11.10	85.02
	RCM	**24.49**	**18.37**	**11.50**	**85.63**
Avg	TF-IDF	17.18	14.06	7.96	38.53
	PLSA	17.49	13.51	8.49	78.85
	CTM	21.75	16.83	11.31	85.13
	RCM	**24.36**	**19.05**	**12.07**	**85.95**

Better performance From Table 1, we can see that RCM clearly outperforms TF-IDF and PLSA on all the eight data sets (+3.6-4.1% in terms of average MAP and +7.1-47.4% in terms of average AUC). TF-IDF and PLSA predict word usage only based on the intrinsic preference of a given user, and ignore the situation where a user's topic distribution may change and become closer to her neighbors' topic distribution over time. TF-IDF performs worse because it directly counts the frequency of words, which may be sparse in paper titles. RCM also outperforms CIM on almost all the data sets. Although CTM also considers both the intrinsic preference of a user and her conformity tendency, it suffers from the problem of data sparsity. Specifically, CTM directly learns the conformity tendency of each user, which is very difficult to be estimated accurately when very few historical actions of the user and/or her neighbors are available for model learning. In contrast, our model clusters similar users (with similar individual attributes) into roles, and then learns the conformity tendency of each role. The data sparsity problem can be well avoided by RCM.

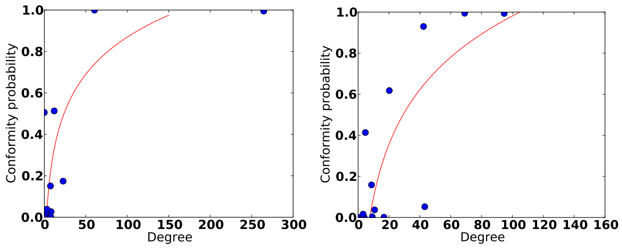
Parameter Analysis There are two tunable parameters, K and R, in RCM. K is the number of topics, which has been analyzed in many previous research (Blei, Ng, and Jordan 2003). We experiment with different values of K, and observe that perplexity first rises and then stabilizes as K gets large. We then fix $K = 30$ where the perplexity is stable and then analyze the number of roles R. Figure 2 plots the correlation between MAP/AUC and the number of roles on the eight data sets. We find that both MAP and AUC present increasing trend at the very beginning and soon become stable when R gets large. The results indicate that our RCM model is insensitive to the number of roles. Finally, we empirically set $R = 13$ in our experiments.

Correlation between role and conformity influence The learned parameter λ by RCM represents the conformity tendency for different roles. The model also learns the mean value of each individual attribute for a role, i.e., $\mu_{r,h}$. Thus we can represent each role as a vector of the mean values of different network attributes and analyze the correlation between role and conformity tendency. We select two domains, DB&DM and HP for further discussions. Figure 3 shows the correlation between a role's mean degree and its conformity probability. We discover that the correlation follows a logarithm function. When fitting the data points, we first remove the roles with a small number of related users, where the number of related users with respect to a role r is estimated by summing up the probability $\rho_{i,t}^r$ over all v_i and t. We try different forms of functions to fit the remaining data points and select logarithm function with the largest R^2. Figure 4 shows the correlation between a role's mean clustering coef-



(a) MAP (b) AUC

Figure 2: Role number analysis.



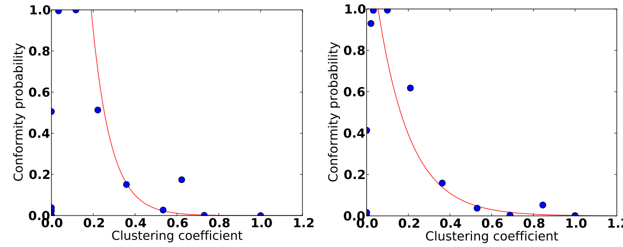
(a) DB&DM (b) HP

Figure 3: The correlation between mean degree of role and conformity probability.

ficient and its conformity probability. We discover that one kind of roles have clustering coefficient close to 0, and the other kind of roles follows an exponential function. Since papers are publicly published, anyone could read others' papers and have almost the same opportunity to be influenced by others. Thus the phenomenon in the two figures may be explained as: when a person collaborates with more authors and the coauthors are more structurally diverse (i.e., with a small clustering coefficient), she may become more open-minded and tend to accept new ideas from others. However, when the social circle of the user is restricted to a few coauthors forming a dense collaboration network, the person will be more conservative and tend not to accept other ideas.

5 Related work

Conformity is a type of social influence involving a change in opinion or behavior in order to fit in with a group. Considerable research (Asch 1955; Bernheim 1994; Cialdini and Goldstein 2003; Kelman 1958) has been conducted on the issue of conformity in social psychology. Recently, several studies on conformity have been conducted on large social networks. For example, Li et al. (Li, Bhowmick, and Sun 2011; 2013) studied the interplay between the influence and conformity. Tang et al. (Tang, Wu, and Sun 2013) proposed a factor graph model to quantify the effects of different conformity factors. However, both the studies do not consider the problem of data sparsity. An individual's conformity cannot be estimated accurately if her historical actions are few. To overcome this problem, we assign hidden roles to users and then learn the correlation between roles and conformity tendency. Besides, to the best of our knowledge, this is the first attempt to formally connect the conformity theory



(a) DB&DM (b) HP

Figure 4: The correlation between mean clustering coefficient of role and conformity probability.

from social psychology to a computational model.

Social influence has been studied throughly. Kempe et al. (Kempe, Kleinberg, and Tardos 2003) first proposed two fundamental diffusion models to estimate the expected influence of given seed users. Bakshy et al. (Bakshy et al. 2012) and Bond et al. (Bond et al. 2012) conducted randomized controlled trials to identify the effect of social influence. Dietz et al. (Dietz, Bickel, and Scheffer 2007) and Liu et al. (Liu et al. 2012) used topic models to learn the influential strength between papers or users. Tang et al. (Tang et al. 2009) and Tan et al. (Tan et al. 2011) used discriminative models to learn the weights of different influence factors. Gruhl et al. (Gruhl et al. 2004), Saito et al. (Kimura et al. 2011) and Goyal et al. (Goyal, Bonchi, and Lakshmanan 2010) learned the influence probabilities of the time-decayed diffusion models. However, they all focus on modeling how users influence others, and ignore the inclination of the users to be influenced.

6 Conclusion

We present the first attempt to connect the conformity theory from social psychology to a computational model. We first formalize conformity theory in terms of a utility function, and validate the utility function by proving the existence of Nash equilibria. We then extend and incorporate the utility function into a probabilistic topic model that takes the role and topic distributions of users into account. Our model allows for mining the correlation between users' hidden roles and conformity tendency. Our experiments on academic research networks show an interesting result that people with higher degree and lower clustering coefficient are more likely to conform to others. In addition, our method also outperforms several baselines on the task of word usage prediction in academic papers.

ACKNOWLEDGMENTS

The work is supported by the National High-tech R&D Program (No. 2014AA015103), National Basic Research Program of China (No. 2014CB340500, No. 2012CB316006), NSFC (No. 61222212, No. 61035004), NSFC-ANR (No. 61261130588), the Tsinghua University Initiative Scientific Research Program (20121088096), a research fund supported by Huawei Technologies Co. Ltd and Beijing key lab of networked multimedia.

References

Aronson, E.; Wilson, T.; and Akert, R. 2007. *Social Psychology*. Prentice Hall.

Asch, S. 1955. Opinions and social pressure.

Bakshy, E.; Eckles, D.; Yan, R.; and Rosenn, I. 2012. Social influence in social advertising: evidence from field experiments. In *EC'12*, 146–161.

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

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *JMLR* 3:993–1022.

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

Cialdini, R., and Goldstein, N. 2003. Social influence: Complicance and conformity. *Annual Review of Psychology* 55.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *WWW'13*, 307–318.

Dietz, L.; Bickel, S.; and Scheffer, T. 2007. Unsupervised prediction of citation influences. In *ICML'07*, 233–240.

Goyal, A.; Bonchi, F.; and Lakshmanan, L. V. 2010. Learning influence probabilities in social networks. In *WSDM'10*, 241–250.

Gruhl, D.; Guha, R.; Liben-Nowell, D.; and Tomkins, A. 2004. Information diffusion through blogspace. In *WWW'04*, 491–501.

Hofmann, T. 1999. Probabilistic latent semantic indexing. In *SIGIR'99*, 50–57.

Jenness, A. 1932. The role of discussion in changing opinion regarding matter of fact. *Abnormal and Social Psychology* 279–296.

Kelman, H. C. 1958. Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution* 2(1):51–60.

Kempe, D.; Kleinberg, J.; and Tardos, E. 2003. Maximizing the spread of influence through a social network. In *KDD'03*, 137–146.

Kimura, M.; Saito, K.; Ohara, K.; and Motoda, H. 2011. Learning information diffusion model in a social network for predicting influence of nodes. *Intell. Data Anal.* 15:633–652.

Li, H.; Bhowmick, S. S.; and Sun, A. 2011. Casino: Towards conformity-aware social influence analysis in online social networks. In *CIKM'11*, 1007–1012.

Li, H.; Bhowmick, S. S.; and Sun, A. 2013. Cinema: conformity-aware greedy algorithm for influence maximization in online social networks. In *EDBT'13*, 323–334.

Liu, L.; Tang, J.; Han, J.; and Yang, S. 2012. Learning influence from heterogeneous social networks. *DataMKD* 25(3):511–544.

Sherif, M. 1935. A study of some social factors in perception. *Archives of Psychology* 187.

Tan, C.; Lee, L.; Tang, J.; Jiang, L.; Zhou, M.; and Li, P. 2011. User-level sentiment analysis incorporating social networks. In *KDD'11*, 1049–1058.

Tang, J.; Sun, J.; Wang, C.; and Yang, Z. 2009. Social influence analysis in large-scale networks. In *KDD'09*, 807–816.

Tang, J.; Wu, S.; and Sun, J. 2013. Confluence: Conformity influence in large social networks. In *KDD'13*, 347–355.