Rediscover Regional Prejudice in China through the Lens of Weibo

Xi Wang*, Zhiya Zuo*, Yang Zhang†, Kang Zhao‡, Yung-Chun Chang§ and Chin-Shun Chou§
*Interdisciplinary Graduate Program in Informatics
†Department of Political Science
‡Department of Management Science
The University of Iowa
Iowa City, IA, USA
Email: xi-wang-1, zhiya-zuo, yang-zhang, kang-zhao@uiowa.edu
§Institute of Information Science,Academia Sinica
Taipei, Taiwan
Email:changyc, chouchris@iis.sinica.edu.tw

***Abstract*—Regional prejudice is prevalent in Chinese cities where native residents and migrants lack mutual trust. Weibo users actively discuss and argue about the issue of migration, which provides a good source of data to examine the communication network of regional prejudice. We are interested in the pairs of post and repost related to the topic of migrants. In a repost, on can add new content in addtional to the reposted content. In particular, we focus on the reposts in response to native residents' complaints about migrants. Based on the sentiment (negative or non-negative) and the direction (native resident→migrant or migrant→native resident), we classify the reposts into four categories. We find the evidence of homophily on regional prejudice in the communication network of Weibo: 72.7 percent of the time, native residents' complaints trigger more complaints from other native residents. What interests us most is the socioeconomic factors that can reverse the sentiment or direction of the initial posts. A multinominal regression model of the reposting patterns reveals that in a city with better housing security or a larger migrant population, migrant Weibo users are much more likely to argue with native residents who hold a negative view about migrants. A secure socioeconomic environment facilitates the communication between migrants and native residents and helps break the self-reinforcing loop of regional prejudice.**

***Index Terms*—Weibo, Regional Prejudice, social network, sentiment analysis**

I. Introduction

On the new year's eve of 2015, 49 people were injured and 36 died in a stampede in Shanghai when more than a million visitors rushed to the observation deck nearby the Chen Yi Square on the Bund. This accident caught great attention and triggered heated discussions on Weibo (a Chinese Twitter-equivalent) – the topic was mentioned 559,120 times within 10 days[1]. Although the criticisms were mostly cast toward local officials for their inadequate preventive actions, many Shanghai natives blamed visitors from other regions for the tragedy. They also complained that migrants brought a number of social problems to Shanghai. Shanghai, one of the most developed cities in China, is infamous for the problem of regional prejudice. Regional prejudice is rooted in distrust and conflicts between native residents and migrants, which is also a problem prevalent in Beijing and many other Chinese cities. In late 1980s, China started to seek market-driven economic development and allowed people to leave their birth towns for work opportunities in another city. Chinese people subsequently began to migrate at an unprecedented scale - the number of migrants has grown fast over time and they now make up a considerable part of the population in every large city. The National Health and Family Planning Commission provided a conservative estimate that the migrant population reached 253 million in 2014[2].

Although both visitors and migrant workers are important forces driving local economic growth, native urban residents tend to have negative impression over the undesirable side of the inflow of migrant population. They blame migrants for increasing crime rates, crowded living space, stealing local jobs, and taking scarce public resources. Some of these concerns are reasonable, while many others are rooted in stereotypes and prejudice toward migrants. Local resentment, along with discriminative government policies, leads to huge barriers for migrants and their children to settling down and getting assimilated in a new city. The Chinese national government calls for a fair treatment of native residents and migrants. However, without breaking regional prejudice, it is difficult to allocate a just portion of public resources to migrants and prevent the urban society from falling into segregation.

This study investigates regional prejudice among Chinese people using data from Sina Weibo. As the most popular micro-blog in China, Weibo's monthly active users reached 222 million in September 2015, with 100 million daily active users on average[3]. Like Twitter, Weibo allows user to publish short and instant posts to share personal stories and exchange opinions on various topics. With this valuable surveillance data of public opinion, we focused on Weibo posts that are relevant to the topic of migrants in Chinese cities. Our previous work [1] developed machine learning algorithms to identify posts

[1]http://data.Weibo.com/report
[2]http://politics.people.com.cn/n/2015/1112/c1001-27805401.html

that expressed regional prejudice. Based on these posts, we build a repost network among users who published content on regional prejudice.

In this study, we hope to rediscover the spread of regional prejudice opinions and explore the factors associated with opinion change from Weibo data. Specifically, we intend to answer the following three research questions:

- Regional prejudice can be detected by looking at three dimensions of Weibo posts: Who publishes a post, a native resident or a migrant? Does this post talk about migrants, native residents, or both? What is the sentiment of the post, negative or not? Our first research question is to find the frequent patterns of these three dimensions.
- One can include new content in a repost in additional to the reposted content. When a Weibo post is reposted, will the direction and sentiment change? In other words, will the reposter continue or reverse the original author's opinion?
- What socioeconomic factors are likely to reverse the sentiment or the direction of the initial posts with regard to migrants?

The paper is organized as follows. The next section reviews the related works on Weibo studies. Then, we introduce the reposting network regarding regional prejudice. We proceed to explore the socioeconomic factors that are likely to affect the reposting patterns. Finally, we conclude our work by summarizing and discussing our research findings.

II. Related Works

Stockmann and Luo [2] said that Weibo has the greatest potential to raise users' awareness, especially awareness of political events. Before February 2016, Sina Weibo had a "140-character limit" policy. Since a Chinese character is like an English word, a Weibo post actually can contain rich content. Weibo has drawn big attention from researchers who study public opinion. For example, Nip and Fu [3] found the import role of Weibo in guiding the public opinion in China. Qu et al. [4] investigated how the public responded to the Yushu earthquake in 2010 on Weibo. And Wang et al. [5] have studied the public's responses to air pollution via Weibo. By comparing the posting patterns across 21 salient social events, Guan et al. [6] found that differences in demographic characteristics are attributed to Weibo users' various behaviors.

Weibo posts are not independent because a post is likely to be triggered by another post from a friend, opinion leader, or organization. Previous research has shown that opinion diffuses via Weibo users ' networks. For example, a study shows that opinion leaders have considerable impact on their followers and negative emotions are more contagious than positive emotions [7]. The most studied type of social media networks is "follower-followee" network [8]–[10]. Comparing a group of verified users with a group unverified users, Chen and She [11] found that the verified users are effective in engaging other users. In a related research, Guo, Li, and Tu [10] discovered a group of verified users who have a large number of followers. Besides "follower-followee" networks, other networks that have been well studied include bipartite networks of Weibo users and posts [12] and a "post-repost" network of bloggers [13]. One limitation of these studies is that they only focus on basic network attributes such as in/out degrees and the number of follower/followees but ingore the rich content of Weibo posts.

Unlike Twitter, where only 35% of its posts are retweets, 65% of Weibo are reposted contents [14]. Therefore, to better understand the spread of information in this platform, it is more practical and reasonable building a post-based network. In other words, we hope to capture the path of information flow by observing the repost behavior. Many studies also analyzed the reposting behavior of Weibo users. Huang and Sun [9] discovered that a user ' s posts were more likely to be reposted if she/he has more followers. It was also found that posts containing pictures and published by verified users would attract people's attention [6]. Lei et al. [12] analyzed the motivations of users' reposting behavior – users with fewer friends were more likely to repost or comment on others' posts, which was beneficial to information diffusion. However, research mentioned above failed to consider shifting of opinions embedded in reposting behaviors. It is very likely a user is expressing an exactly opposite opinion when she/he reposts the Weibo post.

Specifically, public opinion varies along with users' profiles. The distribution of active Sina Weibo users is associated with regional economy and population structure. For example, in central and eastern China, the higher socioeconomic status leads to a larger amount of active Weibo users than the rest areas. Wang, Paul, and Dredze [5] analyzed air-pollution-related Weibo posts among 74 cities. They found that the number of pollution-related posts is significantly correlated with the particle pollution rate reported by the local government. Thus, various user behaviors in posting/reposting can reflect diverse economic and living environments of users' residential regions.

Our main contribution to the literature is that our study is the first attempt to explore the relationship between regional prejudice and socioeconomic status by studying Weibo data.

III. Repost Network of Regional Prejudice

To rediscover regional prejudice content and its diffusion pattern in Weibo, we build a social network to capture information flow among individuals.

A. Data

To the best of our knowledge, there is no publicly available corpus for measuring regional prejudice. Therefore, we compiled a corpus from Sina Weibo for analysis. To gather the Weibo posts and user information, we built a Weibo Webpage crawler. Based on 13 keywords related to the population migrant problem - "native" (本地), "permanent population" (常住人口), "census register" (户籍), "registered permanent residence" (户口), "residence permit" (居住证), "floating

[3]http://ir.Weibo.com/

population" (流动人口), "settle in a new place" (落户), "migrant workers" (农民工), "non-native" (外地), "coming from a different town" (外来), "people from other provinces" (外省人), "transient population" (暂住人口), "temporal residential permit" (暂住证), we retrieved related posts over four months from December 14, 2014 to April 15, 2015. These 13 keywords can be divided into two groups. One group of words, including "census register", "residence permit", and "temporal residential permit", are related to internal migration policies. The other group are labels indicating residential statuses. It is interesting to notice that some of the labels, such as "permanent population", "floating population", "transient population", were created by the Chinese government regarding its specific national condition. Other labels, such as "native", "non-native", and "migrant workers", are common words when people mention migrants. We kept the posts which match at least 1 of the 13 keywords. In total, we collected 4,641,398 Sina Weibo posts. To build a repost network, we only kept posts which have original posts in our dataset. After filtering, 34,187 original Weibo posts and 251,520 reposts from 81,600 users were extracted.

B. Detecting regional prejudice from texts

As it is almost impossible to label all 280k posts manually, we applied text classification algorithms to do this automatically. To train the classifier, we leveraged human annotated data. Specifically, we randomly selected 5,000 posts and asked five coders to annotate the Weibo posts via Crowdsdom[4], which is an annotation platform similar to Amazon Mechanical Turk. The Kappa statistic of the labeling process is 0.63, indicating a reliable result. In our previous paper [1], we described how we built the classifier in detail. Based on our proposed algorithm, Distributed Keyword Vectors (DKV), for sentiment analysis, we could successfully recognize polarity and direction of sentiment for Weibo posts. Specifically, given a Weibo post, we used machine learning methods to assign three labels to categorize its content:

- Owner Type (OT): Who published this post, native residents (NR) or migrants (M)?
- Direction (DR): What is the emotion direction of this post, toward migrants (M) or native residents (NR)?
- Sentiment (SEN): What is the sentiment polarity of this post, negative (Neg) or non-negative (Non-neg)?

Note that for each label, there is an "unknown" category for ambiguous posts. For example, a Weibo post published by an official account of the government talking about a new policy for the migrants is labeled as "unknown" for both OT and SEN. Meanwhile, some posts might contain information in multiple categories. For instance, a post with opinion for both native residents and migrants would have its DR labeled as "NR&M". Table I offered 4 examples of labeled posts. The total number of posts with assigned labels in each category were showed in Table II.

[4]http://crowdsdom.com/

C. The repost network

To understand how information is spread in a social media, one of the most popular ways is to build a directed user network and observe who makes a comment or reposts some others' posts. Both comments and reposts can express users' opinion directly. Reposts, however, will appear in the feeds board and thus can also be seen by followees of the reposter. In other words, the influential power of a repost is stronger than a comment. The goal of our study is to discover the diffusion pattern of regional prejudice opinions on Weibo. The effectiveness of information exposure to others is a key factor we need to consider. As a result, we selected repost network for further analysis.

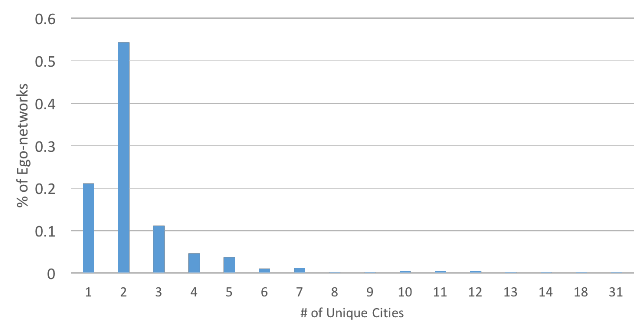


Fig. 1: Distribution of ego-networks with different number of unique cities

Different from the traditional social studies where networks are built with people being vertices, we build a "repost network". This is an effective approach to decrease the impact of users' versatility – A user might have published many different Weibo posts labeled in multiple categories, but none of these opinions are able to entirely represent the user. As a Beijing native, for example, a user made a negative judgment towards Shanghai residents about Shanghai stampeded on behalf of migrants, but she/he could also publish another post to criticize the migrants occupying Beijing resources. Taking a post as a vertex, it has four attributes: three labeled by our classifiers mentioned above (OT, DR, SEN) and one intrinsic property (regional information). To explore the regional prejudice problem, all the three labeled attributes are valuable for our analysis. Therefore, we only kept all vertices with known labels and removed those those with at least one "unknown" label. In total, we have 285,707 posts, including 34,187 original posts and 251,520 reposts.

1) Frequent network patterns: Obviously, all the networks we built are ego-networks. We take the original Weibo post as the focal vertex (ego), which is connected by reposts of this post (alters). The direction is from reposting vertices to reposted ones. Figure 2 is an example of one ego-network. Specifically, given an ego-network, 1-hop neighbors are direct reposts of the original Weibo post while 2-hop neighbors are reposts of 1-hop neighbors. Table III shows the attribute distribution of the original post, 1-hop, and 2-hop reposts, respectively. It can be seen that with the information spreading

Weibo post	Translation	OT	DR	SEN
#上海踩踏事件调查报告#网上的心态啊，其实只要做到一点灾难就不会发生，那就是，让外地人回外地，哪儿来的回哪儿去，把上海留给上海人。。一切都完美了。。。	#The Shanghai Trampling Report# What people are thinking on the Internet! This tragedy could have been avoided by just doing one thing, that is, have the migrants go back to wherever they came from and give Shanghai back to the Shanghainese... Everything will be perfect!	**–**	***M***	***Neg***
早高峰高速路匝道堵了，都是外地车害的，大楼着火了，都是外地临时工害的，外滩踩踏了，估计也有不少外地人，外地人、外地车你们什么时候不是害上海之马！？	Because of migrants, there are traffic jams. Because of migrant workers, the buildings were on fire. There must be many migrants in the trampling on the Bund too. When can you stop bringing Shanghai down, migrants and your cars?	***NR***	***M***	***Neg***
外地人，本地人，来到温江一家人。	Whether you are migrants or native residents, we are families when you come to Wenjiang.	***NR***	**–**	***Non-neg***
同在一座城，请彼此包容也许你来自远方，也许你从一出生就在这里，也许你是深圳人，也许你是外来的新深圳	We are living in the same city. Please be tolerant to each other. Maybe you come here from far away. Maybe you were born here. Maybe you are a Shenzhen native. Maybe you are a newcomer to this city.	**–**	***NR&M***	***Non-neg***

TABLE I: Example of Weibo posts and their labels

	labels	number of posts
OT	NR	75,481
	M	16,884
	–	193,342
DR	NR& M	13,987
	NR	6,873
	M	18,460
	–	246,387
SEN	Neg	114,477
	Non-neg	171,230

TABLE II: Total numbers of posts in each category

	%	OT (NR)	DR (NR)	DR (M)	SEN (Neg)
Original	–	69.01%	69.23%	83.96%	81.54%
1-Hop	–	73.30%	56.19%	88.83%	82.80%
2-Hop	6%	89.80%	73.47%	89.80%	89.80%

TABLE III: Total numbers of posts in each category

Label Combination			1-Hop		2-Hop	
SEN	OT	DR	#	%	#	%
Neg	NR	M	837	67.6%	28	77.8%
Neg	NR	NR	448	36.2%	22	61.1%
Neg	M	M	178	14.4%	4	11.1%
Neg	M	NR	173	14.0%	5	13.9%
Non-neg	NR	M	124	10.0%	0	0
Non-neg	NR	NR	29	2.3%	1	2.8%
Non-neg	M	NR	16	1.3%	0	0%
Non-neg	M	M	11	0.9%	0	0%
Total			1,238		36	

TABLE IV: labels distribution of 1-hop and 2-hop reposts for negative regional prejudice original posts

out, original negative sentiment also propagates further away, (81.54%, 82.84% and 89.80%). Wherever the post is in the ego-network, more than 80% of posts are towards migrants. In addition, we summarized users' regional information. Figure 1 illustrated the relationship between ego-networks and the corresponding number of cities. Clearly, among 458 ego-networks, more than $3/4$ featured posts published by people from different cities.

Focusing on regional prejudice, we limited our analysis to one specific type of ego – original posts stated negative sentiment from native residents towards migrants. Given fixed ego attributes, Table IV shows the label distribution of 1-hop and 2-hop reposts. Obviously, 67.6% 1-hop reposts inherit the opinion of the original view, showing negative sentiments toward migrants published by native residents. This is the most frequent reposting pattern in our dataset. The top two frequent regional prejudice diffusion patterns were shown in Figure 3. Interestingly, some opinions of reposter were exactly opposite to the original Weibo – some migrants expressed negative sentiments against native residents in 1-hop reposts. This is somewhat intuitive – for example, native resident A posted a post to taunt migrant B, who would in turn repost this Weibo and fight back.

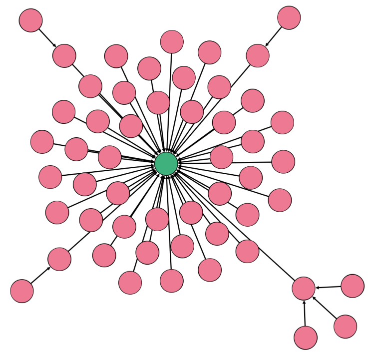


Fig. 2: Example of Ego-network
Green: Ego; Red: Alter

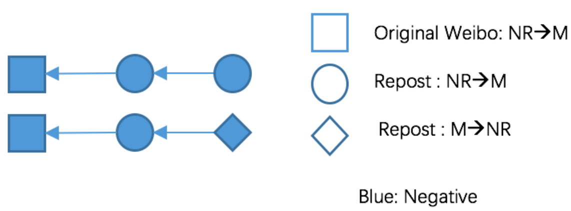


Fig. 3: Frequent Patterns

IV. Socioeconomic Environments and Regional Prejudice

A. Reposting patterns

In this section, we study two dimensions of reposts – direction and sentiment. Since prejudice is unreasonable mistrust between different social groups [15], we only focus on the exchange of posts between native Weibo users and the users who identify themselves as migrants, rather than the communication within a single group. Furthermore, we investigate how socioeconomic environments affect people's attitudes toward migrants and shift the patterns of online discussions over the issue of migrant population. Being filtered, 909 pairs of post and repost remain in the sample of our research interest.

Based on sentiment and direction, the reposts can be classified into four types (Figure 4): (1) a native resident complains about migrants (NR→M, Neg); (2) a native resident discusses about migrants with a neutral or positive tone (NR→M, Non-neg); (3) a migrant criticizes native residents (M→NR, Neg); (4) and a migrant holds a neutral or positive view about native residents (M→NR, Non-neg). The first type of reposts are consistent with the initial posts and continue a clear attitude of regional prejudice. Such reposts are dominant, making up 72.7 percent of the filtered sample. Negative sentiment toward migrants is more likely to be sustained than reversed, as the second type of reposts only constitute 11.1 percent. Unfortunately, the discussion of migrants on Weibo lacks the voice from migrants. Among the filtered reposts, only 16.1 percent come from migrants, among which 1.5 percent contain non-negative tone towards native residents while 14.6 percent show regional resentment.

Fig. 4: Regional Prejudice Patterns

The exchange of migrant-related posts on Weibo supports previous research that in fact the online community is highly polarized and segmented [16]. Compared to a face-to-face setting, the Internet provides many platforms where a person can easily find a group sharing his or her beliefs and ideologies, and block or avoid those with conflicting opinions. People have a tendency to seek the evidence that justifies their established opinions while neglecting the challenging evidence [17]. In the case of Weibo, homophily is the major mechanism underlying the link formation of the reposting network. On the one hand, native residents' complains about migrants often trigger more complains from other native residents. On the other hand, migrants and the native residents who get well with migrants tend to avert arguments with those holding strong regional prejudice. In consequence, regional prejudice becomes a self-reinforcing process.

Distrust between the natives and the migrants is likely to be a result of inadequate communication. Previous research has shown that inter-group contact and communication can help overcome racism [18] and cultural conflicts [19]. Although most reposts retain native Weibo users' criticisms with migrants, we are more interested in the reposts that either alter the sentiment or the direction of an initial post.

The reposting patterns reflect the dynamics of regional prejudice. We proceed to identify the factors that can explain the different patterns of reposting. The reposting pattern is thus the dependent variable in our explanatory model. Since there are a very few reposts in which migrants speak positively about native residents, these reposts are excluded in the analysis, leaving 895 cases. Here we regard the reposting pattern as a nominal variable with three categories.

We model the three reposting patterns with multinomial regression analysis. Naturally, NR→M (Neg) is set to be the baseline category, because it is consistent with the initial post. Concerned with regional heterogeneity, in making statistical inference, we rely on standard errors clustered on the pairs of the locations of the initial posts and reposts.

B. Socioeconomic environment

Socioeconomic contexts have substantial impacts on the relationship between different social groups. Interpersonal trust is likely to prevail in a wealthy society where people have no worry about food, housing, and other basic needs for survival

[20]. Some economic resources are limited and indivisible, and the competition for such resources resembles a zero-sum game. Therefore, when one's current political power or economic well-being is threatened, he or she is likely to show hostility toward new competitors and out-group members. Resource scarcity intensifies social conflicts, but resource abundance has the opposite effect. For instance, a study in the United States found that in wealthy and well-educated communities, as the Latino population increases the white residents become more liberal toward race-related policies [21]. One implication is that increased presence of migrants does not necessarily lead to a more intense relationship between native residents and migrants. In a region with a good socioeconomic environment, when a large number of newcomers arrive, local people may not feel insecure, but instead they may become more tolerant of diverse social groups via frequent interactions.

In explaining the reposting patterns, the key independent variables capture the differences in general socioeconomic status across Chinese cities. Socioeconomic status means the social standing of an individual or group. We retrieve six indicators of general socioeconomic status from the 2010 Census of China, the most recent census. The first indicator is %urban that measures the percentage of urban residents in a city. The level of education attainment is measured by %high school graduate and %college graduate. %unemployed measures the percentage of unemployed adults in the population that is over 16 years old. Two separate indicators are used to measure the percentage of house owners, because rural residents often build a house whereas urban residents often buy or rent a house.

We compare each indicator of socioeconomic status between the city where a repost was written and the city where the initial post was sent. Specifically speaking, we construct an independent variable by subtracting the latter's indicator from the former's indicator. For the cases missing the city identifier, we use the provincial indicators of socioeconomic status as a proxy for the city-level indicators.

The issue of regional prejudice is most salient in the cities with a large migrant population. Therefore, we control for the percentages of migrants in the exploratory model. The 2010 Census distinguishes between two types of migrants. Within-province migrants live in a new city but in their home province. Cross-province migrants left their home provinces to find a job. The two types of migrants have similar impacts on local economy and labor market. However, due to distinct dialects, living styles, and even looks, cross-province migrants are more likely to receive higher prejudice from native residents. In addition, we control the distance between the locations of the post and repost using three binary variables–same province, neighboring provinces, and non-contiguous provinces.

C. Multinominal regression analysis of the reposting patterns

The multinominal regression model has the following functional form:

$$Pr(y_i = k) = \frac{exp(w_k^T x_i)}{exp(w_0^T x_i) + exp(w_1^T x_i) + exp(w_2^T x_i)} \quad (1)$$

where $k = 0, 1, 2$. w_0, w_1, and w_2 correspond to the coefficients for the reposting patterns NR→M (Neg), NR→M (Non-neg), and M→NR (Neg), respectively. To guarantee that w_1 and w_2 are identifiable, the constraint $w_0 = \mathbf{0}$ is added to the model.

Table V shows the results from the multinominal regression analysis. Among all the indicators of general socioeconomic status, only Δ%House Built has a statistically significant effect on the reposting pattern NR→M (Non-neg). The effect of Δ%House Built, however, was in an unexpected direction: as this independent variable increases, native residents are less likely to hold a neutral or positive view about migrants.

With regard to the reposting pattern M→NR (Neg), urbanization level, education level, and unemployment rate do not have statistically significant impact. But in a city with more secure housing, migrants are more likely to criticize native residents in response to their negative posts about migrants. As Table V shows, both Δ%House Built and Δ%House Bought raise the probability of M→NR (Neg) in comparison to the probability of NR→M (Neg), and the effects are statistically significant at the 95% confidence level.

We find that migrants are empowered by their numbers. With regard to M→NR (Neg), Δ%Within-Province Migrant and Δ%Cross-Province Migrant are statistically significant at the 90% confidence level. In cities with a larger migrant population, migrant users become more engaged in the debate on Weibo with local users who complains about migrants. The magnitudes of the coefficients for the two migrant-related variables are fairly close. It appears that regardless of their origins, the migrants living in the same city form a common identity as opposed to the native. We also find that resentful reposts toward native residents are most likely to be sent by the Weibo users in a neighboring province.

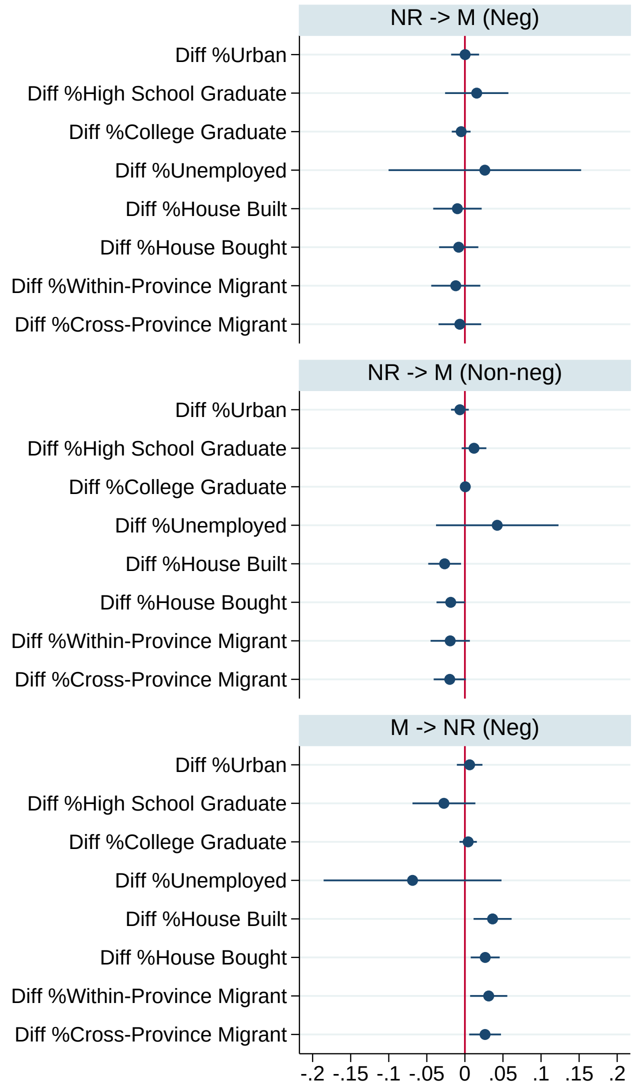
Table V presents the statistical significance of the independent variables which shows whether the observed effects are systematic or largely by chance. But a statistically significant effect may be trivial, only accounting for a tiny portion of the variation in the dependent variable. Thus, it is also necessary to compare the substantive significance which refers to whether an observed effect is large enough to be meaningful. There are multiple metrics for substantive significance, such as first difference and marginal effect. Figure 5 compares the marginal effects of each independent variable in the multinominal regression model. Margin effect is defined as $\frac{\partial Pr(y=k)}{\partial x_j}$. For a multinominal regression, a marginal effect is non-linear that it is dependent on the combination of the values of all the independent variables. Average marginal effect takes the average of the marginal effects at all the data points. So, we rely on average marginal effect to measure the overall substantive effect of a variable.

In Figure 5, the blue lines represent the confidence intervals of the average marginal effects at the 90% confidence levels, the vertical red line indicates the zero average marginal effect. If the confidence interval is broken by the red line, the corresponding average marginal effect is not statistically significant. Marginal effect is meaningful only when it is

	NR→M (Non-neg)	M→NR (Neg)
Δ%Urban	-0.06	0.05
	(0.07)	(0.09)
Δ%High School Graduate	0.09	-0.22
	(0.10)	(0.21)
Δ%College Graduate	0.01	0.04
	(0.04)	(0.06)
Δ%Unemployed	0.37	-0.53
	(0.51)	(0.60)
Δ%House Built	-0.24*	0.28**
	(0.14)	(0.13)
Δ%House Bought	-0.16	0.20**
	(0.12)	(0.10)
Δ%Within-Province Migrant	-0.17	0.24*
	(0.17)	(0.13)
Δ%Cross-Province Migrant	-0.18	0.20*
	(0.14)	(0.11)
Same Province (Baseline Category)		
Neighboring Province	-0.06	1.58*
	(1.63)	(0.88)
Other Province	-0.40	0.09
	(0.34)	(0.27)
Constant	-1.82***	-1.79***
	(0.08)	(0.09)
N	895	
Log-Likelihood	-641.10	

Notes: The baseline category is NR→M (Neg) where the repost did not change the tone or the direction. The standard errors are in parentheses and are clustered by the locations of initial posts and reposts. * indicates p-value $<$ 0.1, ** indicates p-value $<$ 0.05, and *** indicates p-value $<$ 0.01.

TABLE V: Multinominal Regression Analysis of Locations and the Dynamics of Regional Prejudice



Notes: The average marginal effects are estimated based on the multinominal regression model in Table V. The blue points indicate the average marginal effects of socioeconomic differences between the cities of initial Weibo posts and the cities of reposts on each pattern of reposting. The blue lines show the 90% confidence intervals of the average marginal effects. The vertical red line corresponds to zero average marginal effect.

Fig. 5: Average Marginal Effects of Socioeconomic Differences on Reposting Patterns

systematic across random samples. Here we focus on the four average marginal effects (Δ%House Built, Δ%House Bought, Δ%Within-Province Migrant, and Δ%Cross-Province Migrant) with regard to the reposting pattern NR→M (Neg). The average marginal effects are approximately the same, with 0.036 for Δ%House Built, 0.027 for Δ%House Bought, 0.031 for Δ%Within-Province Migrant, and 0.027 for Δ%Cross-Province Migrant. On average, as the gap in one of these four socioeconomic indicators increases by one percent, the probability of the reposting pattern NR→M (Neg) rises by about 3 percent, which is substantial.

V. Discussion and Conclusion

Regional prejudice represents unreasonable hatred and distrust of people from a different place. It is a widespread problem across Chinese cities. Analyzing data from Weibo provides a valuable alternative to traditional social science research methods, such as surveys and experiments, in studying prejudice and other public opinion questions. People are more likely to convey their true opinions on Weibo than in a face-to-face social setting. Prejudice, discrimination, and intolerance, these concepts are hard to directly measure with precision, because they are subject to social desirability effects [22]. In contrast, Weibo users can post controversial or provocative posts under an anonymous identity, so they are less restrained to say what they truly believe.

We are interested in understanding how Weibo users respond to the posts related to regional prejudice. In particular, we focus on the reposts of the Weibo posts that contain a clear tone of regional prejudice – native residents' complains about migrants. These reposts are classified based on the direction (native resident→migrant or migrant→native resident) and the sentiment (negative or non-negative). We find that homophily is the major mechanism underlying the reposting network: 72.7 percent of the time, native residents' complains about migrants lead to more negative posts about migrants from other native residents. Indeed, 11.1 percent of the reposts reversed the initial sentiment by native residents who expressed positive attitudes toward migrants, and 14.6 percent of the reposts changed the initial direction by migrants who criticized native residents. In many cases, regional prejudice is a result of inadequate communication between native and migrant residents. Regional prejudice is thus likely to be weakened when both native residents and migrants get engaged in the debate over the issue of migration.

Furthermore, we examined the socioeconomic factors that could alter the sentiment or the direction of the initial posts. Secure socioeconomic environments breed mutual trust among different social groups. We studied the relationship between regional prejudice and a host of indicators of socioeconomic conditions from the 2010 Chinese Census. Through a multinominal regression model, we find that in the regions with housing security and a large migrant population, migrant Weibo users are more much more likely to get engaged in the argument with native residents who hold a negative view about migrants.

VI. Limitation and Future Work

The outcome of this study has implications for the study of social inequality, which is a area that needs more attention by researchers. The limitation of this work is that we only collected static data without tracking the posts dynamically. Sina Weibo implemented "comprehensive and proactive censorship" to help the government control the content diffused in the Internet [23]. Consequently, we might miss many firsthand and aggressive opinions on public affairs. In the future, we can improve the capacity of our crawler to retrieve posts in a more dynamic way. In addition, it is interesting to work on influential users detection, whose opinions on regional prejudice are more likely to arouse discussions.

Acknowledgment

We would like to thank The Grand Data Challenge of 2015 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP 2015), which provided us an opportunity to inspire a research question about "Inequality". We would also like to acknowledge the feedback from other researchers at SBP 2015. In addition, we also want to express gratitude to Mr. Shih-Wen Huang and his company Jizhi.co for their generous help to our study.

References

[1] Y.-C. Chang, C.-S. Chou, Y. Zhang, X. Wang, and W. Hsu, "Sentiment analysis of chinese microblog message using neural network-based vector representation for measuring regional prejudice," *accepted by PACIS 2016 (details forthcoming)*, 2016.

[2] D. Stockmann and T. Luo, "Which Social Media Facilitate Online Public Opinion in China?" Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2663018, Jul. 2015. [Online]. Available: http://papers.ssrn.com/abstract=2663018

[3] J. Y. M. Nip and K.-w. Fu, "Networked framing between source posts and their reposts: an analysis of public opinion on China's microblogs," *Information, Communication & Society*, vol. 19, no. 8, pp. 1127–1149, Aug. 2016. [Online]. Available: http://dx.doi.org/10.1080/1369118X.2015.1104372

[4] Y. Qu, C. Huang, P. Zhang, and J. Zhang, "Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake." ACM Press, 2011, p. 25. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1958824.1958830

[5] S. Wang, M. J. Paul, and M. Dredze, "Social media as a sensor of air quality and public response in China," *Journal of Medical Internet Research*, vol. 17, no. 3, p. e22, 2015.

[6] W. Guan, H. Gao, M. Yang, Y. Li, H. Ma, W. Qian, Z. Cao, and X. Yang, "Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events," *Physica A: Statistical Mechanics and its Applications*, vol. 395, pp. 340–351, Feb. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437113009369

[7] R. Fan, J. Zhao, Y. Chen, and K. Xu, "Anger Is More Influential than Joy: Sentiment Correlation in Weibo," *PLoS ONE*, vol. 9, no. 10, p. e110184, Oct. 2014. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0110184

[8] H. Zhang, Q. Zhao, H. Liu, J. He, X. Du, H. Chen *et al.*, "Predicting retweet behavior in weibo social network," in *Web Information Systems Engineering-WISE 2012*. Springer, 2012, pp. 737–743.

[9] X. S. Ronggui Huang, "Weibo network, information diffusion and implications for collective action in China," *Information*, vol. 17, no. 1, 2014.

[10] Z. Guo, Z. Li, and H. Tu, "Sina Microblog: An Information-Driven Online Social Network," in *2011 International Conference on Cyberworlds (CW)*, Oct. 2011, pp. 160–167.

[11] J. Chen and J. She, "An analysis of verifications in microblogging social networks–sina weibo," in *Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on*. IEEE, 2012, pp. 147–154.

[12] K. Lei, K. Zhang, and K. Xu, "Understanding Sina Weibo online social network: A community approach," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2013, pp. 3114–3119.

[13] F. Li and N. Lin, "Social network analysis of information diffusion on Sina Weibo micro-blog system," in *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Sep. 2015, pp. 233–236.

[14] L. L. Yu, S. Asur, and B. A. Huberman, "Artificial Inflation: The Real Story of Trends and Trend-Setters in Sina Weibo," in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, Sep. 2012, pp. 514–519.

[15] G. W. Allport, *The nature of prejudice*. Basic books, 1979.

[16] E. Morozov, *The net delusion: The dark side of Internet freedom*. PublicAffairs, 2012.

[17] Z. Kunda, "The case for motivated reasoning." *Psychological bulletin*, vol. 108, no. 3, p. 480, 1990.

[18] S. Welch, *Race and place: Race relations in an American city*. Cambridge University Press, 2001.

[19] S. Ting-Toomey and L. C. Chung, *Understanding intercultural communication*. Oxford University Press, New York, NY, 2012.

[20] R. Inglehart, "Trust, well-being and democracy," in *Democracy and Trust*, M. E. Warren, Ed. Cambridge: Cambridge University Press, 1999.

[21] R. P. Branton and B. S. Jones, "Reexamining racial attitudes: The conditional relationship between diversity and socioeconomic environment," *American Journal of Political Science*, vol. 49, no. 2, pp. 359–372, 2005.

[22] J. H. Kuklinski, M. D. Cobb, and M. Gilens, "Racial attitudes and the "new south"," *The Journal of Politics*, vol. 59, no. 02, pp. 323–349, 1997.

[23] J. Sullivan, "China's Weibo: Is faster different?" *New Media & Society*, vol. 16, no. 1, pp. 24–37, Feb. 2014. [Online]. Available: http://nms.sagepub.com/content/16/1/24