政治科學論叢 / 第 78 期 / 民國 107 年 12 月 / 頁 1-22

Gary King on Reverse Engineering Chinese Government Information Controls*

Gary King
(Transcribed and Edited by Stephen B. Reynolds)

On May 29, 2018, Gary King, the Albert J. Weatherhead III University Professor at Harvard University, gave a speech at the International Conference about Innovations in Political Methodology and China Study, which was held at National Taiwan University. His first keynote speech focused on the methodological techniques that he and his co-authors Jennifer Pan and Margaret E. Roberts used to research the ways in which the Chinese government employs "50 cent party" members to censor online activity.

King showed that contrary to popular belief, the Chinese government generally does not censor criticism, rather it censors online activity perceived to be a potential cause for real collective action. He also found that the main role of the "50 cent party", who are a group of government officials and not ordinary citizens, is to steer discussions in a positive direction. They don't do this by arguing, but by cheerleading. By deciphering an archive of leaked emails, conducting a survey, and extrapolating the results to the rest China, King and his co-authors were able to verify the results, estimate the total number of fabricated posts, and observe how a majority of "50 cent party" posts are made in coordinated bursts during events which have the most potential for collective action. The accuracy of these results were later verified by the Global Times, which is considered to be a mouthpiece of the Chinese government. The discussion that followed King's keynote speech addressed some questions regarding the methodology used in this study.

臺灣大學學術期刊資料庫

* DOI:10.6166/TJPS.201812_(78).0001

This speech was transcribed and edited by Stephen B. Reynolds, a Ph.D. Candidate at the Department of Political Science at National Taiwan University. This speech is based on a publication that can be found here: gking.harvard.edu/50c For more information, please also refer to: GaryKing.org Gary King on Reverse Engineering Chinese Government Information Controls:

This was originally going to be a study about automated text analysis, conducted by Jennifer Pan, Margaret E. Roberts, and me. While we were looking at social media posts in China, we discovered that sometimes when clicking on posts, we would get the message: "This social media post is being investigated." We then discovered that we had the ability to download social media posts before the Chinese government could censor them. Therefore, we had access to the entire corpus of censored social media posts that the Chinese people were not allowed to read. So we thought, let's forget that paper on automated text analysis, and we designed a system to go back to every website in order to figure out whether or not the posts were still up, and we labelled every post as "censored" or "not censored", and we did this study.

The big thing that we found out is that the information controls in China, or the propaganda effort, is so large that even though its purpose is to slow the flow of information, it actually conveys a tremendous amount of information about the intentions, goals and purposes of the leaders of China. It's like a big elephant tiptoeing around and leaving big toe-prints. So, we did one paper about how censored and uncensored social media posts differ, and then we did an experiment where we randomly posted things that we wrote on the web in China to see whether or not they would be censored. We wrote some posts in support of government, some posts opposing the government, some in support of protests, and some against protests. We also have a paper on the Chinese 50c (50 cent) party and whether or not they fabricate social media posts to put on the web. There has never been any evidence of this before, but I'll tell you what we figured out. We also have a whole bunch of other papers on automated text analysis, clustering, keywords, and all kinds of methods, but usually when I give this talk nobody cares about those.

Everybody knows that the goal of censorship is to stop criticism, protests and collective action about the state, its leaders and their policies. What did we learn when we started studying this "obvious point that everybody knows"? What we learned was that it's utterly and completely wrong. What we knew to

臺灣大學圖書館期刊資料庫

be true was actually false. So, we asked ourselves, what could be the goal of censorship? We had a big pile of social media posts that were censored and ones that were uncensored, and we used some lenses to look at them, like whether the post was criticizing the state or whether it was "sensitive". The literature often says: "sensitive" posts are censored and "non-sensitive" posts are not censored. So, we looked at what the definition of "sensitive" was, and we had people code whether the post was "sensitive" or "non-sensitive". The percent censored among the "sensitive" posts was almost the same as the percent censored among the "non-sensitive" posts. So, the concept of "sensitive" is basically fabricated by us. It doesn't predict whether the Chinese government censors it or not. We then asked the question in a lot of different ways. By the way, the idea that theory is dead, because we can estimate everything empirically with so much data, is ridiculous. Once you have so much data, you need a theoretical lens to look at things. For a long time, we were looking at this through the wrong theoretical lens and we didn't see anything, until one day we decided to separate the goals and ask the question: "What could be the goal of censorship?" We wrote down that it could be to stop criticism of the state, or it could be to stop collective action. We then separated the questions. When we did that, everything tremendously clarified. We found out that the first was wrong and that the second was right. I will explain this in more detail.

In China, you can write on social media that the leaders of a town are all stealing money, how much they are stealing, the overseas bank accounts where they have their money stashed, and that they all have mistresses and these are their names. That won't be censored. However, if you say: "… and let's go protest", and there is a real protest on the ground related to that, then that would be censored. In fact, if you say that the leaders of a town are doing such a great job, let's have a rally in their favor, then that would be censored. They don't care what you think of them, rather what they care about is if you have the ability to stop them. This is because they have a pretty good thing going and they'd like to keep it going. What is it that they fear could stop them? It's not the West, it's not military power; rather, it's their own people that they fear. They have to fear people rising up, because that's what gets people who control autocracies thrown out of power. They don't care what you think of them. You can criticize them, you can say things in their favor, it doesn't matter. They care about whether you have the ability to move crowds, and if you do, they'll stop you. That's what we found.

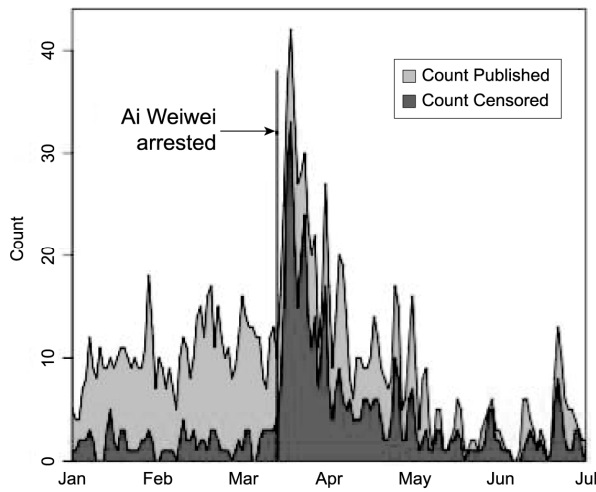
臺灣大學學術期刊資料庫

So what is the implication of this? The implication is that social media is actionable for them, and also for us as researchers. The Chinese leaders measure criticism. They don't ignore criticism, they allow it. They don't allow their local leaders to stop the criticism. They use the criticism to judge local officials. If you are running China, you have to worry about 700,000 towns and cities across China. How do you do that? How do you monitor the leaders of 700,000 towns? You could run public opinion polls in every area, but that's not exactly feasible, right? What they seem to do is monitor criticism on social media of the leaders of each of those towns, and if one of the leaders is engendering too much criticism, then they'll replace that leader. So, criticism is very useful for the Chinese leaders, because they see it. They then censor to stop events with collective action potential, witch is anything that either is a protest now or is likely to generate a protest.

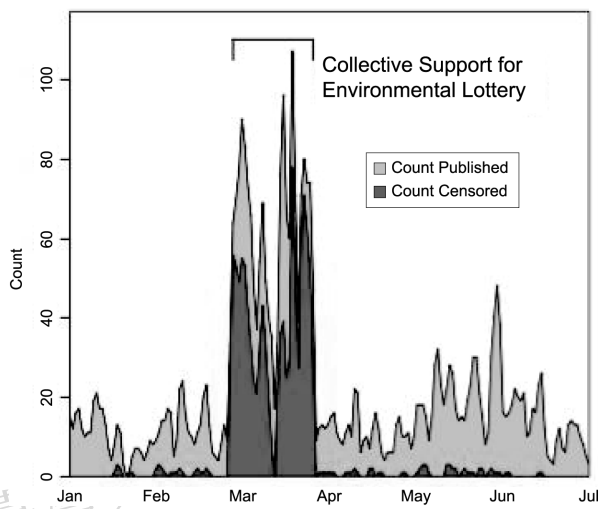
The interesting thing is that as researchers we can use criticism and censorship to predict all kinds of things that we wouldn't be able to otherwise. That's because this big elephant tiptoeing around leaves big footprints. We can predict which officials are in trouble and are likely to be replaced. The media can't tell ahead of time, and theorists can't tell ahead of time, but we can see the information that Chinese leaders have, so we can tell ahead of time. We can measure directly which policies are generating decent, and which are to the interest of the leaders. We can predict government action outside the internet, including predictions we had made that have been accurate: which dissidents were going to be arrested, which peace treaties were going to be signed, and emerging scandals. We don't actually know what the scandal will be or when it's going to appear, but we know something is up. We basically get a different view of the universe by looking at all of those millions and millions of social media posts all across China, at scale. We get a different view into China than could be had otherwise, all from doing political methodology. That was really our goal. We can also see things that we have theorized about for a long time, like disagreements between the central government and local leaders, which is not really something that is visible, but we can see it.

Let me tell you about censorship and the revealing of government goals. I'll give you a few examples of this.

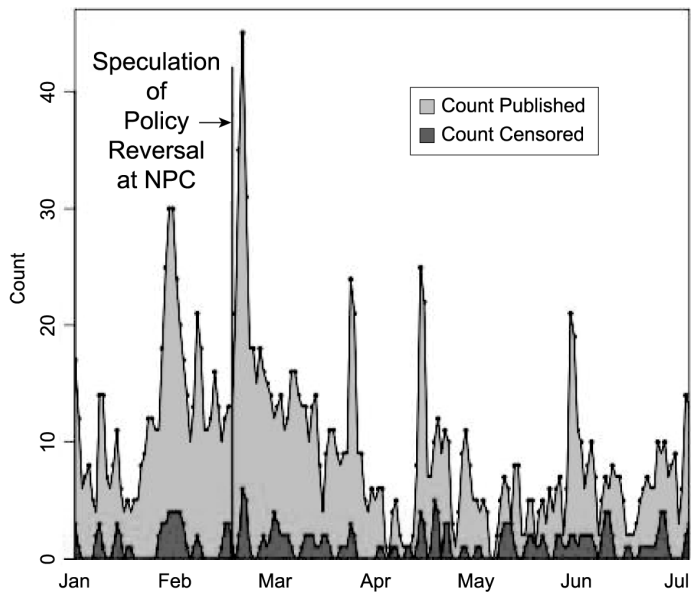
臺灣大學學術期刊資料庫



This is a small sample of data, where the vertical axis is the count of social media posts, and the horizontal axis is six months at a time. The light shade is the total number of social media posts. The spike is all about Ai Weiwei, a famous dissident artist. This is the number of social media posts about him when he was arrested. We know that he had the ability to create crowds and to move people on the ground, and because of that we knew that if there was a viral spike about him, it would be censored, and that's exactly what happened. The dark shade represents censored posts. What happened was a huge spike of activity, and the Chinese government followed them right up the hill, and then took it all down.

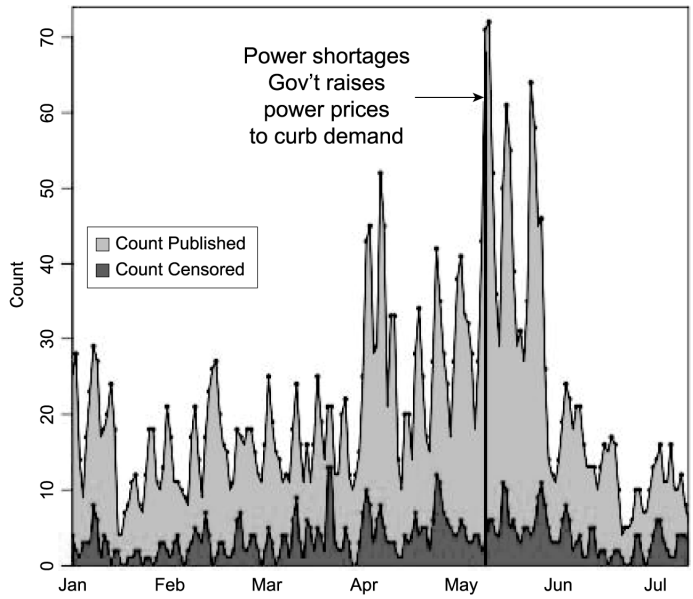


Here's a great example. Here is a city in China where they had an environmental lottery, which was a lottery where people put in a small amount of money and they randomly drew a winner. The lottery made a profit, so where did the profit go? This town announced that the profit would go to the environment. This was very popular and everybody really liked this policy. They liked it so much that they decided to have a big rally for the government officials that announced this policy, in favor of the government officials. What happened? It went viral: "Look what a great thing these leaders are doing!" There was a big spike of activity. What happened next? It was connected to actual protests on the ground, so they censored all of them. They don't care what you think about them, they only care about what you can do.



Here is the most controversial policy that I can imagine: telling people how many kids they can have. However, there was no protest on the ground related to this particular incident, and so there was not much censorship.

臺灣大學學術期刊資料庫



Here we have power prices, which generate protests all the time in the US and other Western countries, but there were no protests on the ground in China, so no real censorship.

Let me switch now to the Chinese 50c party. The idea of the 50c party is that individuals were rumored to be paid 50 cents, or just a few pennies, to fabricate social media posts, which means to write them and post them on the web in the name of ordinary people. The prevailing view of scholars, activists, journalists, and of people on social media who we found who accused other people of being 50c party members, is that the 50c party argues with and debates against those who criticize the government, its leaders and their policies. So, if you say something bad about the government, the 50c party supposedly jumps in and argues against you. This means that they argue against the people who argue against the government. The longest word in the English language is "antidisestablishmentarianism", which means "against the people who are against the government", so I got to use this word in a paper for the first time. What is the evidence that the 50c party argues against people who argue against the government? Well, there are a few anecdotes that we have no idea the veracity of, such as quotes from the literature that say: "There is no ground truth" or "There are no successful attempts to quantify 50c party activity." There are even several scholarly articles where the author makes up

臺灣大學學術期刊資料庫

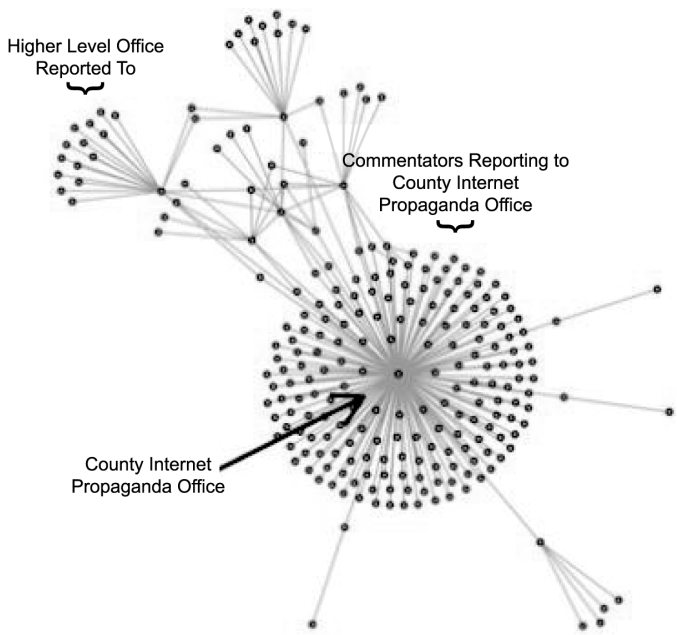
the dependent variable on the basis of the best plausible rules they could come up with, but still made up the dependent variable, and then spent the rest of the article trying to explain the variable that they just made up. These are very well-meaning articles, but they had no evidence.

Here's what we learned when we started this: the prevailing view is utterly and completely wrong. The 50c party does not argue against people who argue against the government. In fact, they don't engage with controversial issues at all. So, what do they do? What they do is distract and redirect public attention from criticism or arguments about central issues to cheerleading and positive discussions around those issues. They post social media posts that say: "It's a beautiful day today" or "I woke up this morning thinking about how important our martyrs were to the history of China." They post hundreds and hundreds of posts like this. Why are they doing this? Why does the Chinese government spend so much money posting cheerleading messages? Well, think of the last fight you had with your spouse, or your parents, or your kids, or your siblings. Suppose you wanted to stop the fight. Option one is that you could use the best reasonable academic argument about why they are wrong; or, option two is to say: "Hey, let's go get ice cream." Doesn't that work better? That's what the Chinese government is doing. Arguing with somebody doesn't end arguments, but saying "let's go get ice cream" sometimes does.

I'm going to provide the first systematic evidence about this point. There has never really been evidence before. First, we found a leaked email archive on the web, which was reported in the news. Someone got all of the emails directed to the Internet Propaganda Office of Zhanggong county in China. It was on the web, reporters basically said that they had the documents, but they never actually read them. What they contain is workers claiming credit for their 50c party posts. So, in a sense, it's the first evidence of actual 50c party posts, though it's just one county. It's a large, unstructured messy dataset, so it's no surprise that the journalists didn't spend a lot of time analyzing it, because it would've taken forever to do it by hand, and to do it by computer would require specialization in automated text analysis, and that's not what they do. It is what we do, however, so we thought it was a good opportunity, and we systematized it. What did we find? We found 2341 emails covering 2 years, 2013 and 2014, among which 1245 contained the text of 50c party posts. There were 43,797 known 50c party posts. There had never been anything like this before, so now we had something to study.

臺灣大學圖書館期刊資料庫

We studied them, and we mapped out the patterns of who posted them, what they posted, where they posted them (which social media sites), when they posted them, and why they posted them, or what was the substance. We dissected the whole mechanism. We knew who these people were, we had pictures of them, telephone numbers, email addresses, we knew what offices they worked in, and so on. We didn't publish that information, but we knew all of it. We then extrapolated the patterns from this one county to the rest of China. We built a statistical model based upon the patterns we found. We identified 50c party workers outside of Zhanggong, in other counties all across China. We then had a prediction throughout all of China about who was posting 50c party posts and how many there were. How did we know that the prediction was right? We did a survey and asked them directly, and we got them to admit that they were doing it. Then, although we didn't plan on it, we also got the Chinese government to admit that this was correct.



Here is the Zhanggong Internet Propaganda Operation. This just gives you a feel for what was going on with the 43,797 emails. The prevailing view is that ordinary citizens are getting paid 50 cents to make each social media post, but in fact, 99.3% of the posts are made by 200 government offices, so it's not people being paid to post them. Imagine if you were to have very low paid

臺灣大學圖書館期刊資料庫

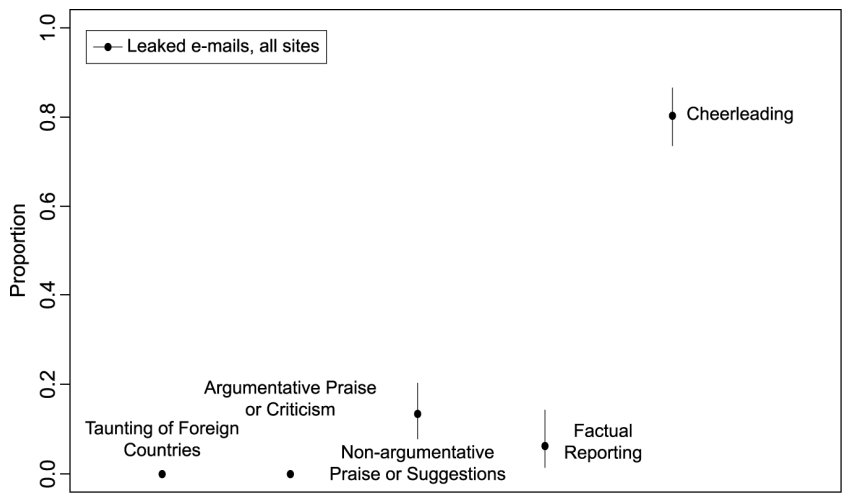
research assistants write social media posts, millions of them, and you wanted them to work accurately over the course of years. That would be really hard to do. Alternatively, if you are the Chinese government, you have millions of people working for you already, so you could just tell them: "When you get home tonight, I want you to post three things on the web. And, I want you to post five things, and I want you to post ten things…" If you think about it, of course that's what they do. These are highly paid employees and they know what they are doing. This is actually what is happening. So, this figure represents the Zhanggong County Propaganda Office, and each of these dots is one of the 200 offices that write into them saying: "OK, we've posted these posts," and are reporting that they did their job and created a certain number of social media posts on different social media sites. The Zhanggong County Internet Office then reports to higher levels. This is the structure of the archive. There is no evidence of 50 cent payments, so this rumor that people are being paid 50 cents per post seems to be totally false. Instead, it's probably just part of their ordinary jobs. That's the first thing that we discovered.

In order to go further, we needed to analyze what these 50c party posts were saying. To do that, we needed to put them in some categories. We estimated the distribution of these social media posts in different categories. The first two categories were: (1) taunting of foreign countries, and (2) argumentative praise or criticism. This is antidisestablishmentarianism, and what everybody thought was happening. Then there was: (3) non-argumentative praise or suggestions, (4) factual reporting, and (5) cheerleading. The first two categories make up the prevailing view of what 50c party members are doing, while we tended to find that the last three categories are what they are actually doing, with most of the evidence pointing to cheerleading. We used these five categories as our theoretical lens to look at the data, and then did separate analyses in stages starting from Zhanggong, and then all the way out to the rest of China, which is the goal.

First of all, we had the 43,797 leaked 50c party posts in Zhanggong, then we narrowed them down to the Weibo posts, because Weibo posts are easier to find in other counties. Then, we had to see if the Weibo posts were approximately the same as all the posts across 1400 different social media sites. If they were, then it would give us some evidence that we could use Weibo to extrapolate. We then partitioned the accounts into two types: ordinary accounts, like a mom who writes about her kids and occasionally other things, and then there is another type of account where the mom who works in government does

臺灣大學學術期刊資料庫

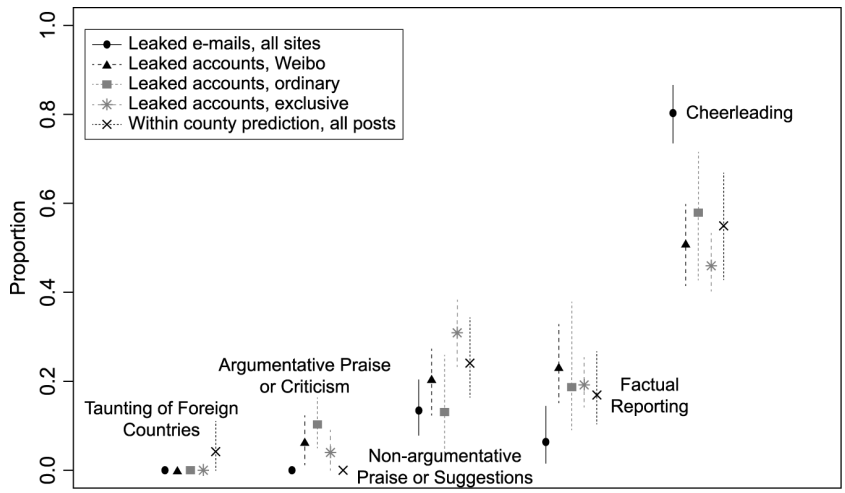
not want to disrupt her social media account with these posts that she's required to make, so she instead makes a completely different kind of account that we call an exclusive account, which is used exclusively for making 50c party posts. For the first type of account, the posts are hidden inside the ordinary account, and in the second they are much easier to spot. We hoped that people were posting the same types of things on the two different kinds of accounts. If we could see that they were, then we could extrapolate to the rest of the country and find the ones that were exclusive accounts much more easily. Exclusive accounts were ones that are engaged with and follow the Zhanggong Weibo, and they had very few followers. We went another stage up and expanded to exclusive 50c Weibo accounts that weren't in our leaked archive, and then we went to unleaked exclusive Weibo accounts in counties all across China, which was our final stage. We took the five categories and applied them five times. Let me give you the results:



First, this is an analysis of the 43,797 posts in Zhanggong. The vertical axis is the proportion of posts, and this is basically a histogram. There is almost no taunting of foreign countries. There is almost no argumentative praise or criticism. There is a little bit of non-argumentative praise or suggestions, and little bit of factual reporting, and there is mostly cheerleading. So, that's what 50c party people are doing. They are mostly saying: "It's a beautiful day today." We used the method that we call "ReadMe", of which Daniel Hopkins and I published a paper about in 2010, where we used a training set by having undergraduates take posts and put them into categories,

臺灣大學學術期刊資料庫

and then we took that training set, and with a much larger test set, we tried to figure out the percentage of each category. That's what we used here, and it estimated these quite accurately.

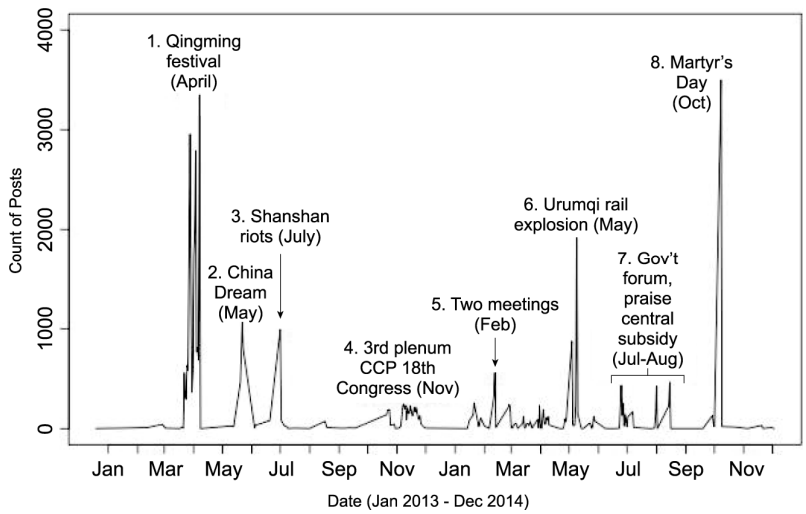


Next, we looked at just the Weibo accounts, and all of the posts from those Weibo accounts, not only the ones in the archive. We would identify the account, and once we had the account, we would scrape all of the posts from that account. This is the histogram of that. Then we divided them up into "ordinary" and "exclusive" accounts. We wanted both types of accounts to be the same, which they seemed to be, which was good because we could use the exclusive accounts to extrapolate to the rest of China. We then looked at all of the posts within the county, even the ones that were not in our leaked archive, and they were getting similar kinds of results. Then there was one more stage where we went out of the county to the rest of China. These are what our basic results were.

Since everything beyond the first result is a prediction, we had to make sure that the predictions were reasonable. First, let me tell you how big this effort is in China. How many posts are they actually fabricating? This is not being done by computer, they are doing it by hand. We know this because we know the people who write them. We could look up each person on the web to see if they are real. This was not a fabricated archive. They don't do this by computer, although they may eventually get to that level, but at the moment they have a lot of workers to utilize. How big is the 50c party? In 2013, there

臺灣大學圖書館期刊資料庫

were about 80 billion total social media posts in China, which included Weibo, blog posts, chat rooms, and other kinds of things. The number of 50c party posts in Zhanggong was 154,216. The number of 50c party posts at the provincial level (Jiangxi) was 10.65 million. The number of 50c party posts that the Chinese government fabricated in 2013 was 448 million. Imagine being in charge of that operation: writing 448 million fabricated posts a year. That is a big job. They are putting an incredible amount of effort into this. About half of these, 53%, are on government sites, where they are a noticeable proportion of all the posts. About half, 47%, are on commercial sites, where it's one out of every 178 posts. If the rate is really that low, then it would have no impact, but it turns out that they are using the posts in a strategic way and that it is a highly coordinated operation.



Here is a plot of the 43,797 posts by 50c party members in Zhanggong. The vertical axis is the count, and the horizontal axis shows the dates over these two years. You can see that during most times, there were almost no posts at all. Then every once in a while they fired a big canon when they needed it, which is when they thought that there might be protests. When they are concerned about protests, or some kind of collective action, then at those times they fire a canon and basically dilute the whole web. If you are someone trying to create some kind of collective action, or some kind of protest, and you put something on social media saying "let's all go to the town square", somebody would ideally repost your message, but if at the same time there are

臺灣大學學術期刊資料庫

also hundreds and thousands of posts saying “it’s a beautiful day today”, then it’s going to be very hard to find your message. That’s what they are doing, and they only do this every once in a while and only when they need it. It’s highly coordinated. The events that they use this on appear to be ones with high collective action potential. It’s a big operation that conveys the government’s intent. Even though it’s an organ of the government designed to slow the flow of information, it is also an elephant tiptoeing around that conveys a lot of information about itself.

How did we validate the 50c party membership predictions? How did we know that these people who were predicted to make 50c party posts were actually making 50c party posts? We figured, why not just try asking them? Of course, these are not your optimal survey respondents; their job is to intentionally mislead people. They take orders from an uncompromising government, their employer is that government, and this potentially puts their jobs at stake. We don’t want to get any individuals in trouble. It turns out that there is extensive literature in the social sciences about asking highly sensitive questions. Imagine public health workers who have to do a survey to figure out about private sexual behavior, or income, which is often an even more sensitive topic. Or, in a country that is even more overbearing than China, asking questions about the leadership might not be something people would be forthcoming about. It turns out that there are special ways that social scientists have come up with to ask questions in these situations, so we used every one of these methods that we could. We also studied the local social media context, so as to not seem like three crazy Americans asking an insane question. We first took a random sample of the accounts that we predicted to be 50c. We then administered a double blind survey via direct messaging on Weibo, so they didn’t know who we were, and we didn’t know who they were. We then specially designed and pre-tested a survey question. I’m not going to tell you the exact text that we used for the survey question, because we don’t want people to be able to search to find these respondents, but I’ll give you a version of it that is translated into English: “I saw your comment, it’s really inspiring, I want to ask, do you have any public opinion guidance management, or online commenting experience?” We did this survey to be our validation. Of course, we also needed to make sure that the validation itself was right, because this was a crazy survey, and why should it even work at all? We needed to validate the validation by asking the same question to a sample of known 50c party members in that leaked archive. This was to see if we would get roughly the same answers as we did for the predicted 50c party members. So, we had a way

臺灣大學圖書館期刊資料庫

to validate our predictions. Then we asked the same question to people pretty much known not to be 50c party members to see whether we would get very different answers. These are the results.

50c Status	Origin	Percent Yes
Predicted 50c	Across China	**59%**
Known 50c	Leaked Zhanggong Archive	57%
Known "Not" 50c	Random sample	19%

We had people who were predicted to be 50c party members all across China, and we pretty much asked them: "Are you a 50c party member?" The percent who we got to admit were 50c party members was 59%. That's pretty good by itself. Roughly 60% of the people said: "Yeah, I'm fabricating social media posts." Then we validated the validation by asking people who we knew for sure were 50c party members, from the leaked archive, and asked the same question to them, and 57% said that they were. As a methodologist, 59% and 57% is the same number. Then we asked a random sample of people who were likely not 50c party members, and about 19% said that they were, which is much lower. The difference between 59% and 57% is not statistically significant, while the difference between 59% and 19% is statistically significant, so it is a pretty good confirmation of our validation, and also a good confirmation that our predictions were accurate.

That was supposed to be the end of our paper, but then something else happened. We were privately sharing the draft of our paper with colleagues for comments and revision, and somehow a reporter from a Western news outlet got a copy of it. They sent me an email saying that they were going to write a news article about the paper and asked if I could answer some questions. I responded: "Hold on a second, we haven't published this paper, I haven't even put it on my website. In fact, we aren't even finished with the paper, could you please wait?" And they said: "No…" So, Molly, Jen and I thought that we probably shouldn't have an argument with someone about freedom of the press in the context of a paper about fabricating social media posts. We answered the reporter's questions and put the paper in its almost-finished state on our website. The news article was published, and within a few hours it was in the *New York Time*s and the *Wall Street Journal*, *Bloomberg*, and other papers all over the world. According to Google News, there are 5000 newspaper articles on our little

臺灣大學學術期刊資料庫

preprint. So many, I fact, that the Chinese government responded. They wrote an editorial in the *Global Times*, which is a CCP controlled nationalist paper. It's the extreme pro-China end of things. First of all, they published it only in Chinese. Even though almost everything in the *Global Times* is published in both English and Chinese, they only published this article in Chinese. This is actually consistent with our hypothesis that the main thing that they are interested in is their own people. They didn't care about explaining to the people who had read the 5000 newspaper articles that they read the "wrong thing" or something like that, they only wanted to explain to their people what they were doing. So, we thought "OK, that's a confirmation". Then they actually admitted to the existence of the 50c party for the first time essentially, or something close to it. They also confirmed the veracity of our leaked archive. They summarized our results accurately. They took no issue with any of our conclusions. They called us "naïve" or "nerds" or something like that, but they did not at any point disagree with anything we said. They also explicitly acknowledged that the purpose of "public opinion guidance" is to stop the spread of grassroots social issues with collective action potential. They literally stated our hypothesis as confirmation. This is roughly like posing a survey question not to a respondent, not to a former member of the government, not to a current member of the government, but to the government itself. We roughly asked them: "Do you agree with our results?" To which the government issued a proclamation saying: "Yeah." I am exaggerating a bit, of course. So, why would they do this? The editorial said that Chinese society is generally in agreement regarding the necessity of public opinion guidance. We thought, that is an interesting empirical hypothesis which we could test, so we tested it.

We first went to the nationalist *Global Times* website where there are comments, and then we went to Weibo. We expected things to be consistent with what they said on the *Global Times* website, and inconsistent with what they said on Weibo. So, 82% of people on the *Global Times* website agreed that public opinion guidance is essential, which is no surprise. But, the real question was what percentage were supportive of it on a more representative Weibo? The answer is 30%. So, this hypothesis is incorrect. The results indicate that the figures are accurate, and the figures are accurate because they don't censor criticism and they don't make up posts that are about the policy area, meaning that the 30% figure is probably right. This indicates that the regime has a problem, because the people do not support this policy. So, they probably wrote this piece because they wanted people to understand what they were doing, and they thought that conveying this was important to them and

臺灣大學圖書館期刊資料庫

important to their people, and if they could convince them that it was important to do this activity, then maybe their people could be a little more forgiving. That's what I think they were trying to do.

What are the theoretical implications of all of this? For China, they do not engage on controversial issues, they do not censor criticism, they do not fabricate posts and post them on the web in the name of other people in order to argue against criticism. What they do is try to stop collective action. They respond to grievances with different public policies in order to prevent collective action. They also give incentives to their local leaders to respond in some way to prevent collective action. They censor any discussion of real-world collective action by distracting with cheerleading and positive valence "it's a beautiful day today" -kinds of posts. More generally for authoritarian regimes, scholars have thought that autocrats want to prevent people from knowing about grievances because they might rise up, which would increase the cost of coordination. But, we find that China grievances are generally allowed since everybody has a grievance living in an autocratic country, and knowing additional ones isn't going to create any big problems. What is not allowed is collective action, which they consider to be a big problem, and they do whatever they can to stop it. It's sort of like a "hair-on-fire" event. We have also learned that arguments never end arguments, but distraction is usually pretty effective. The one thing I want to leave you with to keep in mind for future research is that, when studied at scale, Chinese government information controls leave big footprints, and we can study those big footprints and learn a lot about the leaders, their intentions, their policies, and what they are going to do in the future, better than we could by using other methods. Thank you very much, and you can find some more information on the three authors' websites:

GaryKing.org; JenPan.com; MargaretRoberts.net

Question:

Both the stopping of collective action and the stopping of criticism don't necessarily have to be mutually exclusive. For example, once criticism is identified, that account could be blocked, meaning that there would be no more criticism from that account. In other words, the identity of that account matters. From the perspective of network analysis, how could data from accounts like these be represented? For example, if someone famous is raising criticism, this person could have a high potential to create collective action because so many

臺灣大學學術期刊資料庫

people follow them, and so for someone like Ai Weiwei, his criticism would not be allowed because he is the source of collective action.

Gary King:

This is a really good question which we actually looked into. It turned out that for famous people, or people with a lot of followers, or anyone who can create a big spike of activity on the web, some of them would be censored and some of them wouldn't be. We kept looking, and it turned out that if a famous person like Ai Weiwei had created collective action on the ground in the past, then when he says something that goes viral, it would be censored pretty much no matter what. However, if somebody creates something that goes viral for either good or bad reasons, or it looks like it's going to create collective action, but there was no demonstration of having created collective action in the past, then it wouldn't be censored. I'll give you one example. There was a school principal that had been caught with six kindergarten girls in his hotel room. The mom of one of those girls made a placard and went outside the school, and the placard said: "Leave the girls, take me." It went viral on the web, and you could imagine that this would create collective action, but it didn't. She was the only one there, even though on the web people were taking pictures of themselves holding signs saying "leave the girls, take me" or different variation of it, and it went enormously viral, but there was no collective action. It has to be related to actual collective action on the ground for it to be censored. Maybe it's because the people who are making the censorship decisions are low-level employees, it's not a very fun job, and they are given instructions which are probably: "If you find somebody who has created collective action in the past, and something goes viral about them, take them all down." I think that's what happens.

Question:

I'm particularly Interested in the group who you said were known not to be 50c party members. Did you try to match the characteristics of that sample to those who were known 50c party members, or was it just a random sample?

Gary King:

That's a good question. We just tried to find people who were not 50c party members to the best of our knowledge, or at least had a low probability of being 50c party members. That's basically what we did. We did some matching in the way that you are describing, in that we would survey them from the same places as the ones who were 50c party members.

臺灣大學學術期刊資料庫

Question:

I have a question about the cheerleading findings. As you said, there are some exclusive accounts which have the purpose of posting cheerleading posts. I assume that very few people follow those accounts, so what's the point of posting cheerleading posts if there are no people following?

Gary King:

The same thing that was posted on the exclusive accounts was also posted on ordinary accounts, it was the same exact kind of information, posted in the same way, by the same people. The way Chinese social media works is not like Twitter. In Twitter if you have 10 followers then basically nobody sees your posts, but on Weibo if you put a post on some famous account, then it goes in the stream of that account, and we can then find it because it also appears on the individual's account. So, it did appear on someplace important.

Question:

I agree with you on the interpretation that allowing criticism of local officials is a mechanism of monitoring and sampling, but I think that it's possible that the dichotomy between collective action and criticism may be too simple, because as a user of Weibo I feel that if you criticize a high-level official, even if you were to mention "Paddy the Pig" or "Winnie the Pooh" you would be censored because it relates to a certain leader. So, what do you think of the more complex situation of censorship in terms of politically sensitive issues?

Gary King:

I of course have heard these media reports, and I think that there is actually a very big difference, in this area in particular but perhaps all big data areas, between the methods of journalists and the methods of social scientists. The journalists will come and they'll find one person that had a social media post censored, and they will spin a great story from that one example. I'm not even sure whether the theory that "Winnie the Pooh" is censored is actually correct systematically across the web, although it is testable, and I want to test it. Since the time we wrote our paper and we did the original research, every few weeks someone says: "So, it's changed now, right?" We can't test it every day, but every once in a while we go back and test it again, and it's consistent and the same things still seem to be happening. When we did the experiment where we fabricated posts ourselves, we wrote posts criticizing the leader and they were not censored any more than any other post. It may be that it changes

臺灣大學圖書館期刊資料庫

tomorrow, we don't really know, and maybe some of these points are right, but I'm just very weary of the method of journalism in this particular case because we didn't see the patterns until we were able to measure millions and millions of posts, and nobody has done that for "Winnie the Pooh".

Question:

I would just like to know if you want to evaluate the effectiveness of this information control strategy. Did you try to find a connection between the cheerleading messages and the effectiveness of the outcome? Are these strategies really useful for controlling so-called collective action?

Gary King:

I think this is a really great question, and we haven't run an experiment to try to stop collective action ourselves, so I don't know for sure. I think in some sense because the human attention span is limited and there is only so many things that we can pay attention to at once, if the web is filled with things that are distractive, I think to some extent it has to work, and the fact that they are doing this in big canon bursts every once in a while makes it more likely to work. But, you are right, we need to do a study to figure out whether or not it reduces collective action. To generalize your point, just because the leaders have a rule to stop collective action doesn't mean that collective action is stopped. It may be that they are conveying to the citizens the things that really annoy them and really get them motivated. So, if you are in China, and you are willing to deal with some risk, and you want to change a public policy, collective action is probably the best possible thing you could do. The rumors are that there is a lot of collective action in China. This is another thing that there isn't really any good evidence of, but there are supposedly lots of protests. The thing that we are studying is not just something that the government is doing to the people, because the people are responding. It's a cat and mouse game. It's an arms race on both sides, and we should figure out how to study the other side. I think this is a good point.

Question:

My question is, currently in Mainland China, many in the younger generation are actually very proud to say that they are a part of the 50c party and support the government. So, I think that it is not a very big deal to ask them whether they are a 50c party member or not, because it is not a sensitive topic at all and they would be proud to say that they are. Also, in China there is now another party, something like the "5 US dollar party", which don't agree

臺灣大學圖書館期刊資料庫

to admit to be paid by the US government, so if you are trying to do research on this kind of party, what methodology would you use to explore how many people in China are part of this "5 US dollar party"?

Gary King:

It's an excellent point. The answer is that we don't use the content of the messages to decide whether they are 50c or not. It's not possible to use the content to figure out who is doing it, because there are plenty of people in China that hold nationalist positions that support the government as much as the government itself, and are proud to put that message out on the web. There are also plenty of other people that want to take anti-American views and plenty of people that take the opposite views. We can't use the views to figure out if they belong to a certain party or not, but we can use the kinds of things we did: looking at who they were, where they were posting, what types of accounts they were posting on, etc. That's how we did it. We then used the content to figure out what the people who we found to be 50c party members were posting. One of the great things about all of this is that in social media they are not censoring criticism, so that means that criticism exists, and it seems to be real. The arguments against criticism also seem to be real, so this means that social media analysis is valuable for us scholars because it's real to some extent. Every manner of opinion exists in China, and they do express themselves.

臺灣大學學術期刊資料庫

臺灣大學學術期刊資料庫