Beyond the Wall: Mapping Twitter in China

Share Your Story

Beyond the Wall:
Mapping Twitter in China

Sonya Yan Song
Robert Faris
John Kelly

Beyond the Wall

Mapping Twitter in China

INTERNET MONITOR

INTERNET MONITOR is a research project to evaluate, describe, and summarize the means, mechanisms, and extent of Internet content controls and Internet activity around the world.

thenetmonitor.org

INTERNET MONITOR is a project of the Berkman Center for Internet & Society.
http://cyber.law.harvard.edu

November 2015

Beyond the Wall
Mapping Twitter in China

Sonya Yan Song
Robert Faris
John Kelly

INTERNET MONITOR

ABSTRACT

In this paper, we map and analyze the structure and content found on Twitter centered around users in mainland China. This study offers a rare look at the activity of Chinese Internet users on a platform that is largely unregulated by the state and only reachable through the use of tools that circumvent state-mandated Internet filters. For Internet users that reside in mainland China, Twitter offers access to news from around the world and a wealth of ideas and perspectives that might otherwise be unavailable there, as well as a platform for building online communities that is not under direct control of the government. This study of Chinese Twitter—to our knowledge the first such study—offers a unique window into the online activities and global connections of Chinese Internet users who actively circumvent content restrictions. Based on a mixed-methods approach, combining social network analysis and a qualitative review of the content and activity of Chinese Twitter, we are able to map and provide detailed accounts of the topically based clusters that form among these networks. We identify 36 clusters that focus primarily on three areas: politics, technology, and entertainment. From one perspective, the discourse in the politically engaged portions of Chinese Twitter suggests that Twitter serves an alternative public sphere. The political group is formed of journalists, lawyers, human rights activists, and scholars, who are free to discuss topics typically not permitted in China, such as the Tiananmen Square protests, Tibetan and Uyghur issues, political scandals, and pollution. Yet China's Internet repression is clearly succeeding. Chinese Twitter falls well short of supporting a broadly accessible networked public sphere. The proportion of the Chinese populace with direct access to the debates, communities, and shared resources on Twitter is relatively small, and the avenues by which such discourse might find its way into mainstream political discussion are severely constrained. The firewall between Twitter and the much larger social media platforms in China remains a formidable barrier.

AUTHORS

Sonya Yan Song is a Ph.D. candidate in Media and Information at Michigan State University. She has been named as a 2014 Berkman Fellow, a 2013 Knight-Mozilla OpenNews Fellow and a 2012 Google Policy Fellow.

Robert Faris is the Research Director at the Berkman Center for Internet & Society at Harvard University. His recent research has been focused on developing and applying methods for studying the networked public sphere.

John Kelly is the founder and CEO of Graphika, Inc, and a long-time Research Affiliate of the Berkman Center for Internet & Society at Harvard University.

ACKNOWLEDGEMENTS

CHINA HAS LONG BEEN THE EPICENTER of technological battles over Internet freedom. For the better part of the last two decades, the Chinese government has used a wide variety of strategies and tools to restrict Internet content, while technologists have built tools designed to thwart Internet filters and allow Chinese Internet users access to information blocked by government censors. In this study, we map and analyze the content and structure of user communities on Twitter centered on users in mainland China. Given that the popular microblogging platform has been blocked for approximately five years, studying Chinese Twitter offers a window into the actions of Chinese Internet users who actively circumvent content restrictions. While others in the past have sought to estimate the number of circumvention tool users in China as a rough measure of the impact of these tools, in this study we are able to describe in detail the nature of activity and discourse carried out by Chinese users on an open and public platform beyond the reach of government regulators.

This alternative venue is enjoyed by various groups of people with diverse shared interests that gravitate towards three main areas: politics, technology, and entertainment. The political group is formed of journalists, lawyers, human rights activists, and scholars, who express their opinions on topics considered taboo in China, such as the Tiananmen Square protests, Tibetan and Uyghur issues, political scandals, and pollution. A technology group is focused on less sensitive topics, and its members often follow software developers, tech entrepreneurs, and Internet freedom advocates in the United States. The entertainment group is divided between ACG (anime, comics, and games) lovers and users interested in broader topics, such as fashion, food, and photography.

The overarching question in this study is to what extent Twitter represents an alternative public sphere for Chinese users: whether Twitter constitutes an alternative arena for discussion and sharing of news and information outside of the bounds of traditional media controls. We also begin to offer answers to several key questions in the field: how effective are government actions in restricting online discourse, why do Chinese users seek to get around the Great Firewall, what is the impact of the development and distribution of circumvention tools, and to what extent are Chinese Twitter users engaging with international communities and discussions?

INTERNET RESTRICTIONS IN CHINA

The Chinese government is often described as employing the most sophisticated Internet content control regime in the world.[1] At the core is the so-called Great Firewall, a national filtering system that blocks access to a large number of websites that include content deemed objectionable by government censors. This includes a range of political, social, and religious topics such as Tibet independence movements, the Falun Gong (a banned religious group), and the Tiananmen Square protests of 1989. National filters use a variety of technical filtering approaches. They block IP ranges, tamper with the domain name system (DNS), and employ a system that breaks Internet connections when a sensitive word, phrase, or URL is encountered. Search engines in China are required to remove sensitive websites from search results. China employs a number of other strategies to rein in free expression online, including distributed denial of service (DDoS) attacks, cyberattacks, real-name registration policies for Internet users, licensing and legal requirements for Internet platforms and publishers, and surveillance and monitoring of Internet users, which at times leads to arrests related to online activities. The government is reported to engage in counter speech and information campaigns; users who are putatively paid to offer pro-government opinions in various forums online are commonly referred to as the "50 cent party."

Among the countries that aggressively block Internet content, China stands apart in its ability to selectively block social media content. This has been achieved by blocking external social media

INTERNET MONITOR

platforms while locally based alternatives have become established. The most popular international platforms, including Twitter, Facebook, and YouTube, are blocked in China. Chinese Internet users have access to a range of popular domestically hosted alternatives. For instance, Youku Tudou and other video streaming services fill the gap left by YouTube. Chinese social media serve the largest Internet market in the world. Sina Weibo and Tencent WeChat, in particular, each saw over 100 million active users on a daily basis in 2014.

Social media monitoring and filtering was initially applied to blogging platforms by requiring the platforms themselves to police content.[2] This system has been expanded to the full range of current social media platforms visited by a large and growing number of users. Social media companies combine automated review mechanisms and human review to examine the many millions of posts every day. By some estimates, Chinese social media companies employ tens of thousands of people to monitor and selectively block social media content.[3]

JUMPING THE WALL

There are many options for Internet users who wish to get around government filters, an activity colloquially known as "jumping the wall" in China. These include simple proxies, which are easily added to blocking lists when discovered by the government, and more sophisticated tools designed specifically to avoid detection and resist blocking. Virtual private networks (VPNs), which are a standard tool for businesses to ensure the security of online transactions, are also commonly used to get around Internet filters. The Chinese government typically blocks a broad range of proxies and circumvention tools and engages in a technological battle with circumvention tool developers. The government has never succeeded in blocking all circumvention tools, and doing so would be harmful to online commerce, a situation the Open Internet Tool Project (OpenITP) dubbed "collateral freedom."[4] The end result is that Internet users who are intent on circumventing Internet filters are able to do so. However, jumping the wall entails investing time to identify and install tools that work and requires a level of technological sophistication. Moreover, using circumvention tools often slows down connectivity speeds. Circumventing Internet controls also implies a willingness to defy the government standards for acceptable speech and to take on any perceived risks associated with using circumvention tools.

The implicit tradeoffs made by government regulators between information control and economic growth shift over time and changes in the regulatory climate and technological controls often mean that certain circumvention tools and strategies no longer work. For example, after the OpenITP report, the Google App Engine, crucial infrastructure for e-commerce, was blocked in 2014, and circumvention tools that operated through this platform stopped working. In 2015, various VPN providers reported being blocked as well.

It is unclear how many people in China have used or regularly use circumvention tools. A Berkman Center study estimated users to be less than 5% of the online population.[5] There is a debate over the potential market for such tools and whether low adoption rates are a function of low underlying demand for these tools or of the inconvenience, security, and usability issues with the tools. Of particular relevance to this report, a recent survey of circumvention tool users in China found that using blocked social media sites, including Twitter, was among the most popular reasons for using these tools.[6]

In addition to the technological proficiency required to acquire and use circumvention tools, users must also be willing to take on the risks of defying government content restrictions. For those that

INTERNET MONITOR

are able and willing to overcome these obstacles, we expect the users inside the Great Firewall (GFW) would find it more difficult to access Twitter, use it less often, and form fewer connections.[A] As we describe later, the data support these expectation. Users outside the GWF tend to have more followers and to follow more users on Twitter.

By contrast, in 2013, Sina Weibo had 556 million registered users, 50 million of whom were active on a daily basis. Other Chinese social media have also attracted a great number of users, including WeChat, QZone, and Renren. The popularity of Chinese social media suggests strong network effects: Chinese Internet users will place higher value on a platform if more of their family and friends are using this service or if the service provides access to a broader range of conversations. Based on network effects alone, Sina Weibo has a much larger pull than Twitter. Moreover, to use Twitter, Chinese users not only have to overcome the inconvenience imposed by the Great Firewall but also adapt to an environment with fewer potential familiar users and with different social norms.

The number of Twitter users in China is unknown. Even the platform operator itself will not have an accurate estimate of the number of users that physically reside in mainland China. Since Twitter users in China access the service through proxies or third party applications, it is not possible to accurately determine location for many Chinese users. There are several ways to infer location of users. In the user profile, there are location and time zone options, though no way to verify whether users disclose their true location. It may be possible to pinpoint a mobile user's location if the user allows Twitter or a third-party application to record and report the user's geographic coordinates.[B] We put the available location information to use later in the paper. We do not, however, attempt to estimate the number of active Twitter users in China. The available evidence suggests that active Twitter users comprise a small proportion of Internet users, but are also of sufficient number to maintain a vibrant online space for sharing ideas and information and creating content on online social networks.[C]

Activity on Twitter may serve different functions and denote different social processes.[7] It is also likely that there is substantial variation in the social meaning within each of these behaviors. A retweet or mention in one context might mean something entirely different in another context. Golder and Yardi assert that the "substantive nature of the social tie on Twitter is attention-based," which may cover a range of information seeking activities, whether focused on news, entertainment, or professional information.[8] Following another user may also be driven more by social motives: to stay in touch with existing friends, to find new friends, or to participate in a community of likeminded individuals. In their study of Twitter activity, Java et al. broke down activity into four categories: daily chatter, conversations, information sharing, and reporting news.[9] They divide users into three groups: information seekers, information sources, and friends, which span the range of information seeking and social activity. There is no clear consensus yet on how to interpret what it means to follow someone on Twitter,[10] which is unsurprising given the multiplicity of motives for forming ties on Twitter. Yet few doubt that these ties on Twitter are of social significance.

[A] In this paper, we use the term "Great Firewall" as term of convenience. In this context, it is meant to represent the wide range of Internet content restrictions that apply to users that reside within mainland China.

[B] For a detailed explanation of how we assessed user location, see Appendix 1.

[C] There have been wildly disparate estimates of the number of Twitter users based in China, with one market research firm reporting 35 million Twitter users in China. Alternative, more credible estimates place the number in the tens of thousands. See Jason Q. Ng, "There are NOT millions of Twitter users in China: Supporting @ooof's result and refuting GWI's conclusion," *Blocked on Weibo*, January 6, 2013, http://blockedonweibo.tumblr.com/post/39828699303/there-are-not-millions-of-twitter-users-in-china.

An interesting question is why Chinese netizens use Twitter at all. We expect to find users sharing information and ideas not permitted on more regulated platforms. We might also expect to find users that are interested in engaging with international peers and audiences. Given the inconvenience of using circumvention tools, we anticipate finding tech-savvy users. And given the risks of engaging in conversations on topics that are censored in China, we might also expect to find a high proportion of anonymous or pseudonymous users.

In the following sections, we describe the structure and content in this enclave of less regulated discourse and frame some of the many questions that emerge related to the reach and impact of this online forum.

METHODOLOGY

NODE SELECTION AND NETWORK MAPPING

The analytical basis for this report is a mixed methods research protocol: qualitative content analysis of Chinese Twitter supported by algorithmically drawn network maps and a diverse set of quantitative metrics calculated for each of the clusters in the network. The approach builds upon prior efforts to map online discussion spaces with boundaries drawn that correspond to physical geography.[11]

We generated a social network map of Chinese Twitter accounts based on the relationships between the users (see Figure 1). The network structure is visualized using a physics model layout algorithm (Fruchterman-Rheingold). The resulting network map and structures that emerge reflect the individual decisions of Twitter users to follow other users.

In the visualization, each node represents a single user account, and the size of each node reflects the number of followers from the network to that account. The location of each node relative to the others is based on the collective follow decisions of all of the nodes in the network. The network mapping algorithm is driven by two counteracting forces. One force repels, acting to separate all of the nodes from one another. A second force pulls together accounts that are linked by follow relationships and have followers in common as though by a spring or force of gravity. Thus, densely interconnected network neighborhoods "bunch up" in the map. In this way, one can think of the map as a picture of the pattern of influence and information flow in the network. This location of nodes on the map—based on follow relationships—reflects long-term stable relationships. A mapping algorithm based on mentions and retweets would produce a map with more emphasis on shorter term interests and influenced more by the content of tweets within the period in which data are collected.
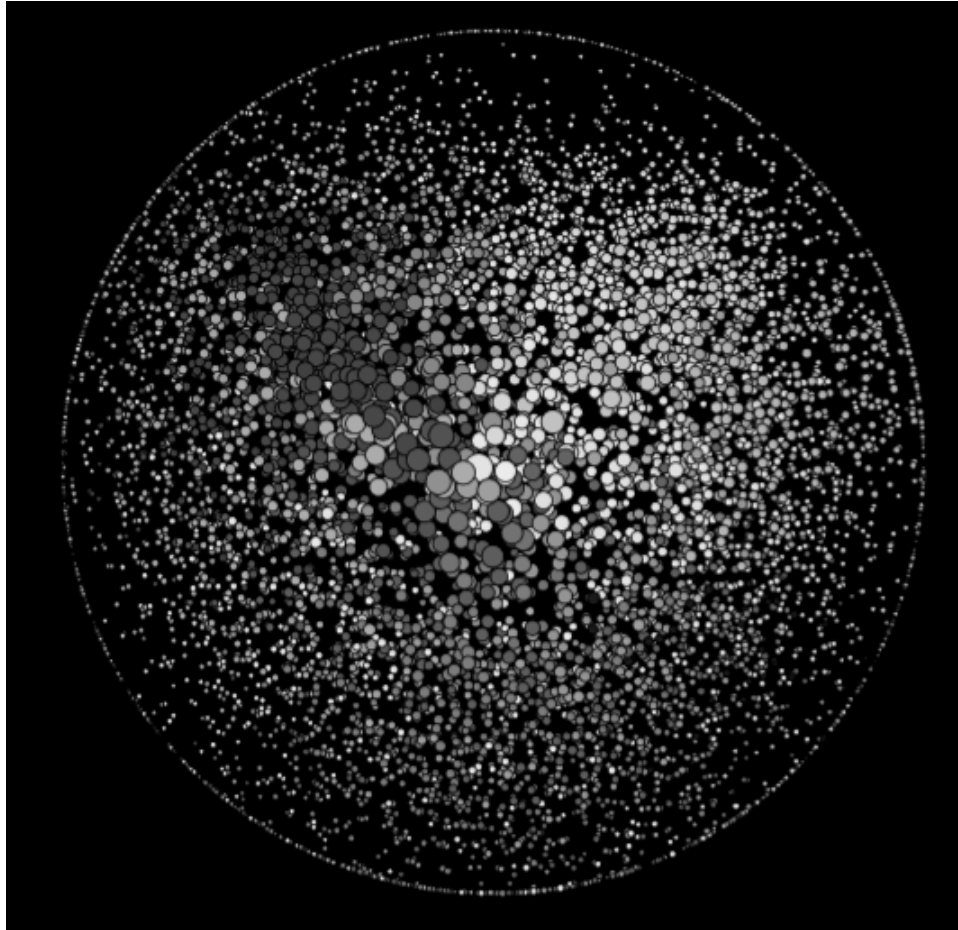
FIGURE 1. NETWORK MAP OF CHINESE TWITTER ACCOUNTS

Determining the accounts that populate the map is based on a multistage process for collecting relevant accounts and removing nodes that are less connected with the core network. The mapping starts with a seed set of approximately 150 accounts compiled by researchers. This set was expanded by adding all of the followers of the seeds, and then reduced by removing all accounts with no activity over the prior 90 days (February through April 2013). After producing a preliminary map based on these accounts, each of the clusters was evaluated for relevance based on participation of and interaction with users from mainland China. For example, clusters entirely comprised of users from Taiwan with no ties to mainland users were given low relevancy scores, while clusters comprised largely of users in mainland China were assigned high relevancy scores. These cluster weights were used to iterate the map, bringing in more nodes highly connected to members of positively weighted clusters and excluding those more connected to negatively weighted clusters. This has the effect of "centering" the map on the intended scope (mainland China), while removing accounts from Taiwan, expatriates, and other non-mainland Chinese accounts.

The resulting network map is comprised primarily of users in mainland China but is not limited to users there. International users are included if they are significantly connected with the core network of users. This includes some accounts that have large followings among mainland Chinese users, even if they follow few or no users in China, and international users that both follow and are followed by users in China. For example, users that are connected with Taiwanese user communities

are only included in the map if they are also highly connected with this network of users centered on mainland Chinese users.

The final map includes a total of 8,275 accounts, constituting a set that is rich in detail and resolution while being of a tractable size for quantitative and qualitative analysis. As described in Appendix 1, we are able to infer location for 84% of the users, and of those, 73% are within mainland China.

This approach to gathering accounts is in effect a snowball sample. However, the high density of Twitter networks and the use of automated discovery tools eliminates most if not all of the potential starting point biases. It is possible that there are smaller isolated enclaves of users that have few or no ties to these larger communities and were therefore not discovered by the automated spidering process. However, it is unlikely that we have missed relevant clusters of users from our target population: Twitter users in China that are openly engaged with others on social and political issues. While the study of smaller more isolated communities would be of interest in its own right, these communities are arguably not key participants in the networked public sphere and out of the scope of this study.

DEFINING CLUSTERS

The resulting map is overlaid with colors representing each account's assignment to a group based on a clustering algorithm (Figure 2). The assignment to different color-coded clusters is based on the commonality in outward attention. Accounts that use the same hashtags, use similar language, cite the same URLs, and follow the same accounts are drawn together into clusters. Qualitative human judgment is subsequently used to interpret the results, but not to generate the maps themselves.
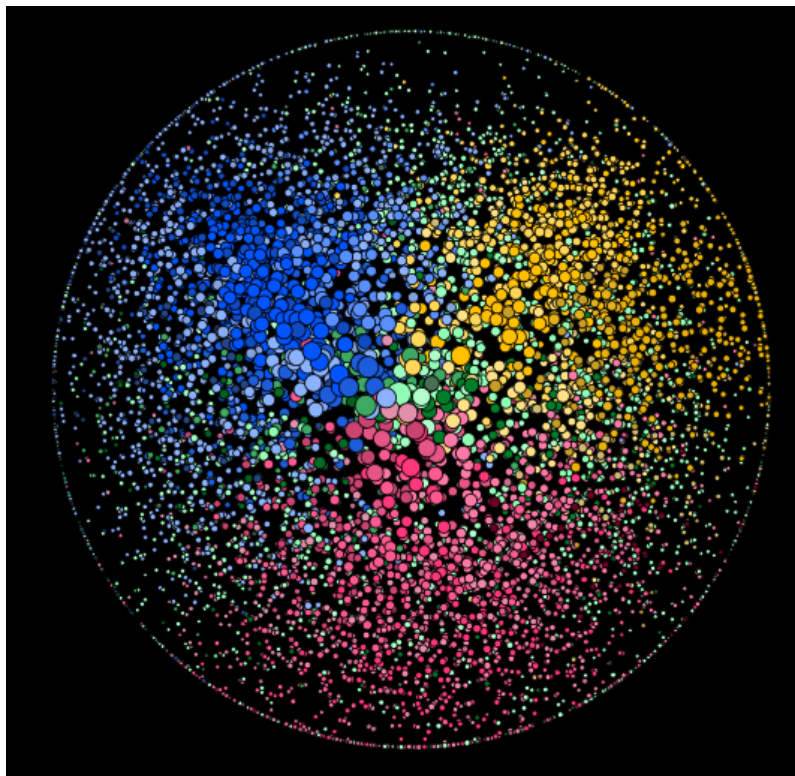

FIGURE 2. NETWORK MAP WITH COLORS DENOTING THE ATTENTIVE CLUSTERS

A variety of metrics are calculated to help researchers understand and describe the contours and structures of the map. These metrics provide a quantitative measure of which activities occur proportionately more often in each cluster compared to the others, including:

- Twitter accounts followed
- Twitter accounts mentioned, retweeted, and replied to within a timeframe
- URLs cited in tweets
- Words used, including bigrams and trigrams (pairs and triples)
- Countries, cities, and other locations mentioned in user profiles

Researchers apply labels to each of the clusters based on a review of these data. These labels are not meant to categorize each account within a given cluster but rather offer concise shorthand descriptions for the various clusters. As we will describe in the next section, the clusters vary by the interests and perspectives of the users in each clusters. There is, however, overlap between some of the clusters in the topics that they discuss. Each user offers a unique perspective that does not necessarily map perfectly with the set of common themes found in each cluster. For many of the accounts within a cluster, the cluster label offers a good summary of the primary focus and general orientation of the account; for other accounts, the cluster label may not strongly capture the interests and views of the user.

The final map of 8,275 users—either residing in China or strongly connected with Chinese residents—consists of 36 clusters oriented around three main topics: politics and human rights; technology; and culture and entertainment (see Figure 3). The political group consists of nine clusters and 2,334 users, the technology group comprises 2,029 users divided into nine clusters, and the entertainment group, made of eight clusters, contains 1,394 users. A fourth set of ten clusters with a total of 2,518 users spans a broader number of interests and includes bloggers, users based in Hong Kong and Taiwan, and groups that are oriented around celebrities.
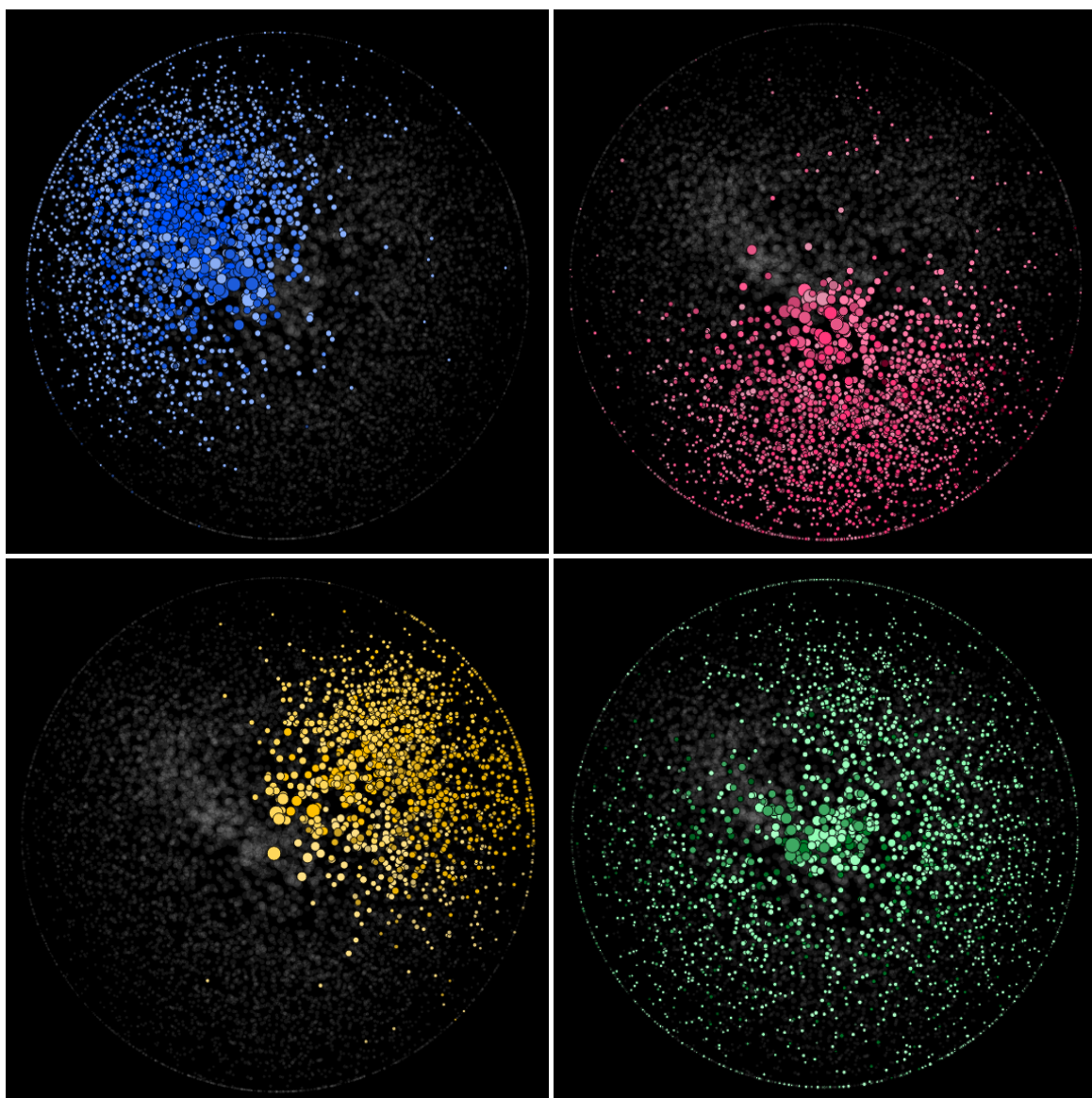
FIGURE 3. FOUR GROUPS OF CLUSTERS: POLITICAL GROUP (TOP LEFT, BLUE), TECHNOLOGY GROUP (TOP RIGHT, PINK), ENTERTAINMENT GROUP (BOTTOM LEFT, YELLOW), AND MIXED GROUP (BOTTOM RIGHT, GREEN)

GROUP	CLUSTER	DESCRIPTION
POLITICS	Human Rights Activists & Lawyers	Attention oriented toward prominent activists and human rights issues.
	Ai Weiwei Followers	Focus on Ai Weiwei and other prominent activists.
	Human Rights Advocates	Focus on human rights advocates including activists, lawyers, writers, filmmakers, and scholars; many live outside China.
	China Bridge	Includes Chinese expats overseas and foreign correspondents, scholars, and businesspeople with interest in China.
	Citizen Journalists	Users active in civic causes who live both inside and outside China.
	Dissidents & Reformists	Focus on exiled activists, many linked to Tiananmen Square protests.
	Journalists & Writers	Focus on prominent journalists, writers, and bloggers.
	Free China & Free Tibet	Interests related to both the Jasmine Revolution and Tibetan issues.
	Political News	Focus on news and information aggregators.
TECHNOLOGY	Software Development	Attention directed at software engineers inside and outside China.
	Open Source Software	Open source and tech orientation; many users from Beijing.
	Tech Gurus	Attention directed at popular figures in the tech world.
	Tech Entrepreneurs	Users follow and interact with famous names in tech.
	Tech & Media	Focus on media. Many users from Beijing.
	Tech & Culture	Interests include gadgets and devices.
	Apps & Misc Tech	Focus on Apple products and its apps.
	Internet Freedom & Tech	Focus on tech entrepreneurs and circumvention tools.
	Tech News	Attention to tech news, particularly @ifanr, a popular tech blog.
ENTERTAINMENT	ACG Shanghai	Fans of anime, comics, and games from Shanghai.
	ACG Guangzhou	Anime, comics, and game fans from Guangzhou, Hong Kong, and Chongqing.
	ACG Sharing	Anime, comics, and game fans from across the country.
	Fun Reads	Focus on entertaining Chinese websites.
	Curiosities & Humor	Interest in humor.
	Misc Entertainment	Interest in humor, anime, comics, and games.
	Beijing Hipsters	Attention to trendy mobile services. Users mainly from Beijing.
	Shanghai Hipsters	Users from Shanghai with interest in mobile services and websites.
MIXED	Hong Kong Users	Users mostly from Hong Kong.
	Taiwanese Users	Mostly Taiwan-based users.
	Japanese Celebrities	Focus on celebrities from Japan.
	US Celebrities	Focus on celebrities from the United States.
	Online Shopping	Interest in e-commerce sites. Many users from outside China.
	Foreign Media	Focus on English-language foreign media.
	Shanghai Bloggers	Bloggers based in mainly in Shanghai.
	News, Tech & Entertainment	Diverse interests including political news, tech, and entertainment.
	Tech & Satire	Interests include a mix of technology and humor.
	Misc	Uncategorized Twitter accounts.

TABLE 1. A BRIEF DESCRIPTION OF THE 36 CLUSTERS

INTERNET MONITOR

NETWORK DESCRIPTION

POLITICAL CLUSTERS

Nine distinct clusters emerge in the network that are primarily focused on politics and news. The popular topics and political orientation of discussions in these clusters all gravitate toward issues related to dissidents, pro-democracy movements, human rights, and the concerns of activists. There is substantial overlap in interests among many of the political clusters.

The **Human Rights Activists & Lawyers** cluster mainly follows and interacts with activists, e.g., Ai Weiwei, Chen Yunfei, and Zeng Jinyan;[D,E] and human rights lawyers, scholars, and writers, e.g., Zhang Lifan, Teng Biao, and Mo Zhixu. During data collection in 2013, many of the users in this cluster lived in China. One of the causes taken up by users in this cluster was demanding an investigation into the student casualties in the Sichuan earthquake. The hashtag #512birthday was created to memorialize the young victims from the earthquake, which occurred on May 12, 2008. Another popular hashtag, #发课税案 (fake case), was used to rebut the tax evasion charges against Ai Weiwei, which Ai and his supporters believe were made in retaliation for his efforts to reveal the government's role in the high death toll during the earthquake. Although the Chinese government admitted shoddy construction might have contributed to the deaths of thousands of young students,[12] it meanwhile suppressed pleas from citizens and media to investigate possible corruption or negligence of local officials by attempting to pay bereaved parents to stay silent[13] and detaining domestic and foreign investigators.[14] Another cluster, **Ai Weiwei Followers**, follows and interacts with a similar set of accounts, e.g., Chen Yunfei, Wang Lihong, etc. They pay particular attention to Ai Weiwei, who receives the most mentions, retweets, and replies in this cluster. The members also frequently use #aiflower as a show of support for Ai. The hashtag refers to flowers that Ai has been placing in a bike basket in front of his house since being put under house arrest on November 30, 2013.

The **Human Rights Advocates** cluster includes human rights advocates from diverse backgrounds. Activists, lawyers, writers, filmmakers, and scholars, many of whom live outside China, appear in this cluster. They pay attention to Chinese media operated overseas, e.g., 博讯 (Boxun), 参与 (Canyu), 维权网 (Rights Protection), and 民生观察 (Civil Rights & Livelihood Watch). Among the popular hashtags are those that refer to Zhang Anni, a 10-year-old girl who was denied permission to attend two schools because her father was a dissident, and to Sun Zhigang, a young man who was sent to a detention center because he carried no documents and was later found dead there.

The **China Bridge** cluster serves as a bridge between China and the outside world thanks to two groups of accounts: Chinese people with overseas connections and foreign correspondents, scholars, and business people who are living or used to live in China and are able to understand the Chinese language. This includes, for example, Hu Jia, a Chinese activist who has been recognized and awarded by foreign organizations; Qian Gang, a well-known journalist residing in Hong Kong; Isaac Mao, a Chinese technologist and investor, who has served as a board member to the Tor Project, an

---

[D] In this paper, we use Pinyin, a standard phonetic system from mainland China, to Romanize Chinese names with a family name preceding given names. For users who refer to themselves with a Romanized version of their name, we follow their usage. In some cases, this means that their given name precedes the family name.

[E] The names cited in this study are either prominent public figures or cited with the consent of the users. If we were unable to obtain consent for less prominent users, i.e., those not cited in major media, we do not report their names and user accounts.

INTERNET MONITOR

advisor to Global Voices Online, and a Fellow at the Berkman Center; Jeremy Goldkorn, a South African living in Beijing who runs Danwei, a China-focused blog; Melissa Chan, a former China correspondent for Al Jazeera, who was recently denied entry to China; Rebecca MacKinnon, a former CNN Beijing Bureau Chief who currently works with the New America Foundation's Open Technology Institute; Perry Link, a prominent sinologist whose entry to China has been denied since 1996 due to his involvement with the Tiananmen Papers; and ChinaFile, an online magazine focused on US-China relations.

Users in the **Citizen Journalists** cluster live both inside and outside China and are active in civic causes. For example, Zhu Chengzhi is a businessman turned activist; Suofeng Zhou has been in exile since 1989 and is the co-founder of Humanitarian China; Yaxue Cao is the founder and editor of ChinaChange.org. Users in this cluster mainly follow Chinese media operated overseas, e.g., 博讯 (Boxun), 参与 (Canyu), 良心中国 (Conscience China), and 维权网 (Rights Protection). They often use hashtags that refer to Liu Xia, the wife of detained Nobel Peace Prize winner Liu Xiaobo.

The **Dissidents & Reformists** cluster follows and interacts with people in exile involved with the Tiananmen Square Protests in 1989, including Bob Fu and Wang Juntao, and other exiles that left China before or after 1989 due to their advocacy of political reforms, e.g., Hu Ping, He Qinglian, and Yu Jie. There are also ties to dissidents and reformists remaining in China, e.g., Woeser, a Tibetan writer, and Wang Lihong, a human rights advocate. Common hashtags used in this cluster relate to democracy and advocacy efforts directed at freeing detained dissidents. The links shared in this cluster often lead to foreign media, e.g., VOA, RFI, Radio Free Asia, and Chinese media operated overseas, e.g., 博讯 (Boxun), 参与 (Canyu), Epoch Times, Human Rights in China, and 64 Tianwang.

The attention of the **Journalists & Writers** cluster is directed at famous journalists, writers and bloggers, including, for example, Xiao Shu, who is a historian who was prohibited from teaching for seven years since the Tiananmen Square Protests in 1989. His books were banned in 1999, he was forced from his position as editor-in-chief in 2005, his column was withdrawn in 2011, and his Weibo Sina account was deleted in 2012. Compared to other clusters, users in this cluster tend to employ more sophisticated language and sarcasm in their tweets, including concepts and terms that are not widely discussed by average users, e.g., #福柯必须跪 (Foucault must kneel), #democraticdev, and #h7n9 (a reference to an avian influenza virus).

The **Free China & Free Tibet** cluster distinguishes itself by sharing links and using hashtags that commonly relate to the Jasmine Revolution and Tibetan exiles. The first group includes Jasmine Action, Sound of Hope, China Spring of Freedom, Secret China, and China Democracy Party, while the second includes Tibet on the Map, Save Tibet, High Peaks & Pure Earth, Tibetan Centre for Human Rights and Democracy, and Central Tibetan Administration.

The **Political News** cluster follows and interacts with news and information aggregators, e.g., 推特观察 (Twitter Watch), 奇闻录 (anecdote collection), 博讯 (Boxun), Invisible Tibet, and Free Weibo, a full Weibo archive that includes deleted posts. Hashtags reflect an interest in diverse news topics, e.g., 盛世 (great dynasties, a phrase that is used to remind Chinese of the glorious history and the bright future), #512birthday (in memory of student victims during the Sichuan earthquake on May 12, 2008), and 科学上网 (scientific use of the Internet, a comical reference to "jumping the wall").

INTERNET MONITOR

TECHNOLOGY CLUSTERS

The **Software Development** cluster pays particular attention to software engineers inside and outside China. Their frequent hashtags are often related to development, e.g., #jquery, #git, #productivity, #osx, and #backend. Compared to other tech clusters, the **Open Source Software** cluster pays close attention to information related to open source software, as reflected by their frequent hashtags, e.g., #ubuntu, #vim, #linux, #bash, and #firefox. The majority of users in these clusters reveal they are located in Beijing.

The **Tech Gurus** cluster includes popular figures in the tech world, e.g., Feng Dahui, a prominent programmer and CTO of an IT firm, and Xi Qiao, author of the comic series Mysterious Programmers. The links they share are related to development, e.g., GitHub, Ruby China, Stack Overflow, CSDN (China Software Development Network), and SlideShare. The **Tech Entrepreneurs** cluster devotes attention to technology and entrepreneurship. The top users they follow and interact with include famous names in the field, e.g., Tim O'Reilly, Om Malik, and Robert Scoble. The links they share are from Kickstarter, Wired, and WSJ Technology, and the top hashtags include #sxsw, #wmc (World Mobile Congress), and #bigdata. Many members of this cluster are located in the Bay Area and New York.

The focus of the **Tech & Media** cluster is oriented toward media-related accounts, e.g., @guao, a news feed focused on Google products, and Jason Ng, a tech entrepreneur and blogger. The majority of the users reveal they are located in Beijing. The members of the **Tech & Culture** cluster show interest in gadgets and devices, reflected by the frequent hashtags: #gmail, #lastfm, #alipay (the payment service by Alibaba), #siri, #ipadmini, and #ux (user experience). Many of them also follow Jason Ng.

The **Apps & Misc Tech** cluster pays much attention to Apple products and related apps. During the time of data collection, the members were especially interested in apps related to fitness and weather, reflected in the hashtags, e.g., #maccn, #appcn, #fitstats, #fitbit, and #instaweather. The **Internet Freedom & Tech** cluster pays attention to tech entrepreneurs (Dash Huang), investors (Kai-Fu Lee), and tech bloggers (@keso). They express their eagerness to jump the Great Firewall with text and images in their profiles and tweets.

This **Tech News** cluster pays much attention to a tech news aggregator @ifanr, a popular Chinese tech blog. Their top hashtags and links are more fun than technical and occasionally refer to adult content, which is not observed in the other clusters. The majority of the users reveal they are located in Beijing.

ENTERTAINMENT CLUSTERS

Three of the clusters in the entertainment network share a love of anime, comics, and games (ACG), with many links to Japanese websites. Two of these, **ACG Shanghai** and **ACG Guangzhou**, are comprised of users from distinct geographic regions; the former include users predominantly from Shanghai and the latter from Guangzhou, although users from Hong Kong and Chongqing are also found in this cluster. The **ACG Sharing** cluster includes users from across the country, though with similar interests to the other two ACG clusters.

Two clusters, **Fun Reads** and **Curiosities & Humor**, tend to pay attention to entertainment-oriented Chinese websites such as 糗事百科 (encyclopedia of embarrassment). The latter cluster also shares an interest in ACG; cartoon characters are a popular choice for profile pictures in this

cluster. As shown later in Figure 4, this cluster is fairly far from the other entertainment clusters. The **Misc Entertainment** cluster occupies the middle ground between the ACG lovers and Fun Reads, which implies the assorted interests of its members.

The **Beijing Hipsters** cluster includes users with an interest in trendy mobile services, e.g., Instagram, FourSquare, and Last FM. They also often link to cnBeta, a website focused on latest technologies and gadgets. Most of them claim to be located in Beijing. The **Shanghai Hipsters** cluster, like the Beijing hipsters, shares information on similar mobile services and websites, with a majority of the members claiming to be in Shanghai.

MIXED CLUSTERS

This set of clusters displays a broader range of interests compared to the clusters that formed around politics, technology, and entertainment. Unlike the political, technology, and entertainment clusters, the accounts are distributed across the network map.

Two clusters, **Hong Kong Users** and **Taiwanese Users**, are comprised primarily of users from those specific locations, although with connections to the mainland China-based clusters and accounts. There are many other users from these and other locations with fewer connections to this network who are therefore not included in this study. Another two clusters, **Japanese Celebrities** and **US Celebrities**, are drawn together by their common interest in celebrity news.

An **Online Shopping** cluster appears to be comprised mainly of overseas Chinese who frequently tweet in Chinese. This cluster pays attention to e-commerce websites, e.g., Taobao and aihuishou.com (love to recycle/resell). The **Foreign Media** cluster follows and interacts primarily with English-language foreign media, e.g., AP, BBC, CNN, Huffington Post, New York Times, Washington Post, etc.

The **Shanghai Bloggers** cluster has many users from Shanghai who are interested in a variety of websites, e.g., Shanghaiist (a popular English website about China), Zhihu (a Chinese Q&A platform), and Douban (a social network gathering people with similar interests in music, film, literature, and other cultural products). Three other clusters include users with broader interests that span news, technology, and entertainment.

TOP USERS

The list of users with the most followers from this network consists primarily of political activists and journalists, but also includes popular figures from the technology and entertainment world.

NAME	SCREEN NAME	TWITTER FOLLOWERS IN THIS NETWORK	TOTAL TWITTER FOLLOWERS[F]	TOTAL WEIBO FOLLOWERS[G]	DESCRIPTION
Michael Anti	Mranti	4,499	71,087	60,611	Michael Anti is the pen name for Zhao Jing, a Chinese journalist and blogger known for his advocacy of press freedom in China.
艾未未 (Ai Weiwei)	Aiww	4,365	213,156	N/A	Ai is an artist and activist. He has been under surveillance and subject to travel restrictions since 2012.
韩寒 (Han Han)	Twocold	3,990	245,682	41,592,121	Han is a professional rally driver and writer, known for his blog posts questioning Chinese authorities.
冉云飞 (Ran Yunfei)	ranyunfei	3,943	76,369	N/A	Ran is a writer, activist, and endorser of Charter 08. He was arrested and remains under house surveillance.
闾丘露薇 (Lüqiu Luwei)	roseluqiu	3,930	95,806	3,695,178	Lüqiu is a female journalist working with Phoenix Television in Hong Kong. She was the first female war correspondent from China.
月光博客 (William Long)	williamlong	3,890	83,912	385,357	William Long is a popular tech blogger in China. He is active on both Twitter and Sina Weibo.
和菜头 (He Cai Tou)	Hecaitou	3,798	89,521	417,501	He Cai Tou is the pen name for He Jian, a popular writer and blogger with a focus on technology.
南方周末 (Southern Weekend)	NanZhou	3,769	159,759	5,140,869	Southern Weekend, based in Guangzhou, South China, is among the most outspoken and circulated newspapers in China.
李笑来 (Li Xiaolai)	Kristy_xiao	3,646	53,440	45,102	Li is a popular blogger and lecturer in China. He is now a partner of an educational enterprise. His current Twitter account is @xiaolai.
北风 (温云超, Wen Yunchao)	wenyunchao	3,572	83,541	N/A	Wen, also known by his pen name North Wind, is a blogger and activist from China. He was exiled to Hong Kong and was denied re-entry into China.
FMN 自曲新闻 (Free News)	freemoren	3,559	47,613	N/A	This account is a news curator with a focus on China. Its slogan is "true news coverage."
伍嘉贤 (Jason Ng)	jason5ng32	3,485	38,083	N/A	Ng is an entrepreneur living in Beijing and founder of several Internet products, including Qdan.me and kenengba.com.
王佩 (Wang Pei)	Wangpei	3,451	38,550	46,789	Wang is a playwright, a blogger, and the founder of two blog sites, heibanbao.com (black board) and baibanbao.net (white board).
西乔 (Xi Qiao)	arthur369	3,349	46,043	N/A	Xi Qiao is a female artist and the author of a comic series titled Mysterious Programmers.
佐拉 (Zhou Shuguang)	Zuola	3,338	47,143	7326	Zhou, better known by his pen name Zuola, is a citizen journalist covering controversial stories in China. He is now in exile in Taiwan.

[F] As of May 15, 2013.
[G] Those users that do not have an account on Sina Weibo are marked N/A.

INTERNET MONITOR

NAME	SCREEN NAME	TWITTER FOLLOWERS IN THIS NETWORK	TOTAL TWITTER FOLLOWERS[F]	TOTAL WEIBO FOLLOWERS[G]	DESCRIPTION
冯大辉 (Feng Dahui)	Fenng	3,329	42,930	1,762,154	Feng is a well-known programmer and the CTO of an IT firm, active on both Twitter and Sina Weibo.
蒲飞 (Pu Fei)	64tianwang	3,280	32,393	N/A	Pu's Twitter profile states that he/she is a volunteer for 64tianwang.com, a website focused on human rights in China.
冯正虎 (Feng Zhenghu)	fzhenghu	3,270	53,067	87	Feng is a dissident and supporter of the Tiananmen Square protests. He has been arrested and denied re-entry into China.
蒼井そら (Sola Aoi)	aoi_sola	3,260	392,390	15,761,989	Sola Aoi is a Japanese adult video star and actress popular both domestically and internationally. She has nearly 16 million followers on Sina Weibo with regular updates in Chinese.
滕彪 (Teng Biao)	Tengbiao	3,247	52,843	N/A	Teng is a human rights lawyer, activist, and endorser of Charter 08. He has been arrested several times in China.

TABLE 2. USERS WITH THE MOST FOLLOWERS FROM WITHIN THE NETWORK

This list of top users in this network differs markedly from the top users overall on Twitter and Sina Weibo. Overall top users on both platforms have attracted followers on a similar scale: 20-53 million for the top 20 Twitter users and 37-72 million for the top Sina Weibo users. Another similarity between top Twitter and Weibo users is that they are predominantly entertainers, along with a smaller number of politicians and businesspeople. By contrast, the top Twitter users in this network feature accounts from more diverse backgrounds, including entertainment, news media, and technology. Yet few of them have reached a milestone of one million followers. Overall, the lack of Chinese-speaking superstars on Twitter is a result of Twitter's low rate of market penetration in mainland China. By contrast, Sina Weibo has the largest user base in the region, including the most famous names hailing not only from mainland China, but also from Hong Kong, Taiwan, and Singapore.

Across Twitter and Sina Weibo, only one person appears among the top 20 on both lists: Han Han. Han is a professional rally driver and writer, and his blog is among the most widely followed in the world. Because Han openly questions authorities in China, he has been widely admired by younger generations of Chinese and extensively applauded by Western media, helping him to build strong connections with both Greater China and other parts of the world and reach different audiences across social media platforms. A few other top Chinese Twitter users are also active on Sina Weibo, such as @williamlong and @hecaitou, who focus on technology, which makes them less dependent on Twitter. However, a number of top Chinese Twitter users, such as Ai Weiwei and Ran Yunfei, do not have accounts on Weibo because they are political dissidents silenced by authorities in China.

LOCATION OF USERS IN THE NETWORK

As described earlier, the network map is focused on users in mainland China, but is not limited exclusively to users there. Although we are not able to accurately determine where each and every user in the network resides, we can make reasonable inferences about location based on the several sources of location information available. Of particular interest to us is whether users reside within

mainland China or elsewhere.[H] There are two sources of helpful information in the Twitter profile data: user-provided information on location and the user-selected time zone. Each of these fields is provided by users who may wish to mask their location, so we must treat these data with caution. In addition, the metadata for many of the accounts include longitude and latitude coordinates.

Where longitude and latitude data are available, we use that to determine location. This accounts for 17% of the users in the network. If coordinates are not available, we next turn to the user-provided location name. If we able to match the user-provided location name to a known geographic location, we use this to infer location. An additional 56% of the users fall into this category. Among the remaining 27% of users, 17% left the location field blank and 10% provided information that could not be associated with a specific location (e.g., "沦陷区" (enemy-occupied area), "潘多拉星球·拆迁办" (Demolition Office, Pandora Moon), "404 not found", and "Airstrip One, Oceania"). However, for approximately half of these users, we are able to infer from their choice in time zone that they reside within China. Using these various sources of information, we are able to estimate for 84% of the users whether they reside within China or not.

[H] In Appendix 1, we present an analysis of location at a more detailed level, showing the approximate distribution of users within China and other countries.

GROUP		% INSIDE GFW	MEAN # FOLLOWERS	% FOLLOWERS IN GFW	MEAN # FOLLOWEES	% FOLLOWEES IN GFW	FOLLOWER / FOLLOWEE RATIO
T	Open Source Software	96%	155	90%	119	84%	1.3
E	Curiosities & Humor	91%	242	87%	322	83%	0.7
T	Internet Freedom & Tech	90%	267	86%	357	81%	0.8
E	Shanghai Hipsters	90%	395	89%	455	87%	0.9
P	Human Rights Activists & Lawyers	89%	353	84%	492	80%	0.7
T	Tech News	89%	238	88%	231	85%	1.0
E	Misc Entertainment	88%	227	89%	223	87%	1.0
T	Software Development	88%	94	89%	109	78%	0.9
T	Tech Gurus	88%	206	88%	159	80%	1.3
E	Fun Reads	87%	250	88%	394	86%	0.7
T	Apps & Misc Tech	87%	244	88%	151	80%	1.6
T	Tech & Media	86%	126	86%	172	78%	0.7
E	ACG Sharing	86%	194	86%	217	84%	0.9
T	Tech & Culture	86%	227	87%	148	82%	1.5
M	News, Tech & Entertainment	85%	82	86%	102	79%	0.8
M	Misc	84%	140	86%	251	81%	0.6
E	Beijing Hipsters	84%	288	89%	217	87%	1.3
P	Political News	83%	201	83%	440	79%	0.5
P	Journalists & Writers	82%	192	84%	176	78%	1.1
E	ACG Guangzhou	80%	123	84%	118	79%	1.0
E	ACG Shanghai	80%	258	85%	223	82%	1.2
P	Human Rights Advocates	80%	461	82%	499	79%	0.9
P	Ai Weiwei Followers	78%	144	83%	189	79%	0.8
M	Online Shopping	76%	196	86%	175	83%	1.1
P	Citizen Journalists	76%	278	81%	267	77%	1.1
M	Shanghai Bloggers	74%	412	85%	169	78%	2.4
M	Tech & Satire	71%	329	83%	221	76%	1.5
P	Free China & Free Tibet	69%	194	83%	167	80%	1.2
P	Dissidents & Reformists	63%	168	80%	114	72%	1.6
P	China Bridge	51%	217	79%	137	71%	1.6
M	Foreign Media	26%	89	82%	30	67%	3.2
T	Tech Entrepreneurs	16%	101	77%	23	38%	4.1
M	US Celebrities	14%	46	80%	4	26%	7.7
M	Hong Kong Users	14%	177	70%	115	55%	1.6
M	Japanese Celebrities	13%	75	79%	4	29%	18.6
M	Taiwanese Users	9%	159	72%	87	53%	1.8

TABLE 3. LOCATION OF FOLLOWERS AND FOLLOWEES BY CLUSTER (E: ENTERTAINMENT, M: MIXED, P: POLITICS, T: TECHNOLOGY)

INTERNET MONITOR

As seen in Table 3, the great majority of users in a large majority of the clusters are in mainland China, although there is significant variation across the clusters. For 26 of the 36 clusters, three quarters or more of users appear to live in mainland China. The clusters within the technology and entertainment groups have a high proportion of users within China.

At the other side of the spectrum, there are six clusters that comprised primarily of users based outside of China: **Foreign Media, Tech Entrepreneurs, Hong Kong users, Taiwanese users**, and the US and Japanese celebrity-focused clusters. The users in these clusters tend to have asymmetric relationships with significantly more followers in this network than followees. The inclusion of these clusters in the network is driven by the outward facing interests of users based in mainland China.

Among the China-based clusters, the clusters from the political group tend have a greater mix of international and domestic users. The **China Bridge**, **Dissidents & Reformists**, and **Free China & Free Tibet** clusters have 51%, 63%, and 69% of their users in mainland China, respectively.

Comparing the clusters by the location of their followers in the network, there is less variation; between 70 to 90% of followers aggregated across each cluster come from within China. Looking across the groups, the entertainment and technology clusters tend to have a larger percentage of followers who reside in mainland China, compared to the political clusters. Similarly, the location of followees, which represents the external focus of the users in each of the clusters, does not vary greatly, except for the six clusters with a majority of users located outside of China. Again, the political clusters display a higher level of interest and engagement with users outside of China compared to the technology and entertainment clusters. Overall, considering the location of users, their followers, and their followees, it appears that users in this network based in mainland China are well integrated with users from outside the country.

Figure 4 shows that the users outside of mainland China (blue) have both more followers and followees from within the network than their counterparts inside China (red). This aligns with the expectation that the inconvenience of overcoming the GFW translates into fewer connections for those that reside in mainland China.
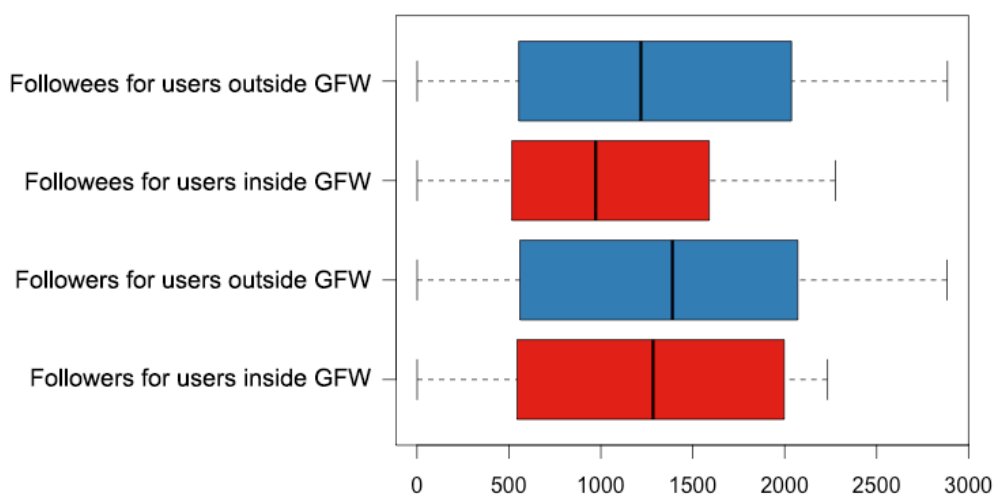

FIGURE 4. DISTRIBUTION OF THE NUMBER OF FOLLOWERS AND FOLLOWEES FROM WITHIN THE NETWORK FOR USERS INSIDE AND OUTSIDE OF MAINLAND CHINA

INTERNET MONITOR
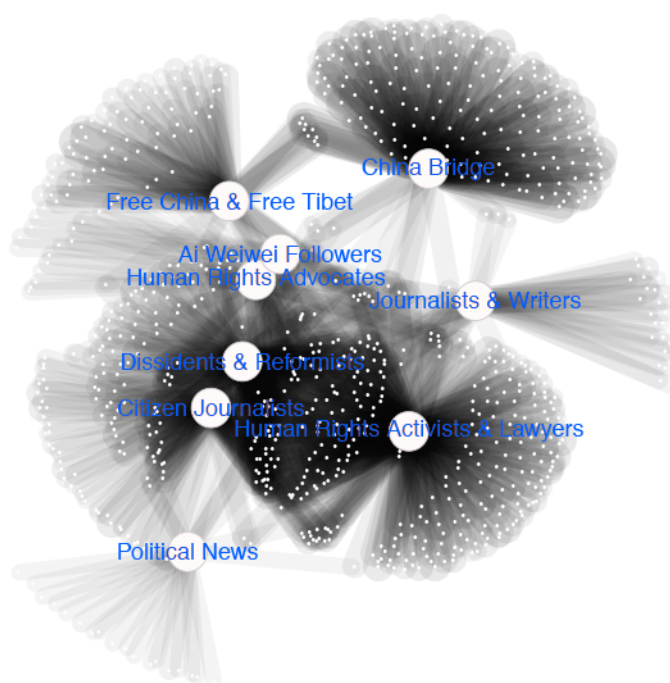
INTERACTIVE CONNECTIONS ACROSS CLUSTERS

Many of the 36 clusters in the network share areas of interest. To gauge the level of separation and common attention between the clusters, we map the clusters around areas of disproportionate attention, including the accounts that are followed, mentioned, replied to, and retweeted. These two-mode networks are created using iGraph for R for each of the political, technology, and entertainment groups.[15] Not only do these charts illustrate how the clusters are connected by common attention; they also reflect more dynamic information flows across clusters through retweets, replies, and mentions. We have visualized all the three types of interactions (retweets, replies, and mentions) but present here only the visualizations based on retweets as the results are similar for all three. In the visualizations, edges are between clusters and accounts that are disproportionately retweeted by a cluster; a normalized score was calculated for messages retweeted from each account to measure the level of disproportionate interest from each cluster. The width and color of the edges are proportional to the normalized attention score: higher levels of attention from a cluster to an account is displayed by a wider and deeper edge (see Figure 5).

The political chart shows where clusters share relatively more common interest in retweeting users. Several clusters are drawn together in this way: **Ai Weiwei Followers**, **Dissidents & Reformists**, **Human Rights Advocates**, and **Citizen Journalists**. The retweeting behavior of other clusters such as **Political News**, **Free China & Free Tibet**, and **China Bridge** moves them farther apart. Similar patterns were also observed among the technology clusters. More accounts were retweeted in common across **Tech Entrepreneurs**, **Tech Gurus**, **Software Development**, **Tech & Culture**, and **Open Source Software**, while accounts within the **Internet Freedom & Tech**, **Tech & Media**, and **Tech News** were retweeted more internally to these clusters. Quite differently, the entertainment clusters showed a clear divide between ACG lovers and other fun seekers. This gap may be induced by the nature of cultural products, which often arouse strong preferences and avoidances (Webster 2014).[16]

Examining hashtags helps to reveal the differences in interests and attention among the different clusters. The political chart shows the affinity between several of the clusters with a focus on dissidents and activists, and a gap between these clusters and the **China Bridge** and **Free China & Free Tibet** clusters. The dissident- and activist-focused clusters appealed for the freedom of their fellow members, including Chen Guancheng, Yang Kuang, Chen Yunfei, Xu Wei, and others. During the time of data collection, Liu Xia was at the top of the list, and many hashtags were related to her, including #刘霞 (Liu Xia), #freelx, #freeliuxia, #自由刘霞 (free Liu Xia), and #freeliu. On the other side of chart, the **China Bridge** cluster paid more attention to issues and regions rather than individuals, such as #sichuanquake, #xinjiang, #china, #pollution, #saudi, #northkorea, #southkorea, #thatcher, #senkaku, #nkorea, #dprk, #chine, #japan, #uyghur, #asia, #npc, #fdi, #sichaun, #nra, #birdflu, #nuclear, #eu, #southafrica, #us, #pyongyang, #nkorean, #sichuan, #dalailama, #india, #tibet, #syria, #quake, #environment, #tibetans, #afp, #weibo, #askeconomist, #xi, #iraq, #chinese, and #diplomacy. Members of the **Free China & Free Tibet** cluster talk more about Tibet and the Jasmine Revolution, which is reflected by their hashtag choices: #tibetonthemap, #tibetan, #tibet, #chinese, #tsamparevolution, #fujian, #tibetans, #gyama (Gyama Valley, where mine disasters occurred), #反日 (anti-Japan), #molihua (jasmine), #cnjasmine, #netdog (contemptuous name for Internet commentators), #lhasa, #hongkong, and #woeser (a Tibetan writer). The entertainment clusters also appear divided between ACG lovers and other users, and the hashtags reflect their distinct interests in ACG and other fun activities.

Many of the discussions on Twitter would have been prohibited by domestic social media. In particular, the most relevant political hashtags are commonly censored in China, including 1989, Tibet, freedom, democracy, national security (国保), and the names of both government officials and dissidents, such as Ai Weiwei, Liu Xia, and Xi Jinping. The hashtags most frequently retweeted by the technology clusters are less sensitive, including #backend, #移动社交 (mobile social network), #iosdev, #galaxysiv, #iwatch, and #nerds. The entertainment hashtags tended to be entertaining and ironic, such as #众推神早安 (good morning, tweeting gods and goddesses), #我的儿子是个奇葩 (my son is a weirdo), #会走路的av百科全书 (walking encyclopedia of adult videos), and #喜当爹 (happy to father a child).

"Retweet" Connections among Politics Clusters

INTERNET MONITOR

"Retweet" Connections among Technology Clusters
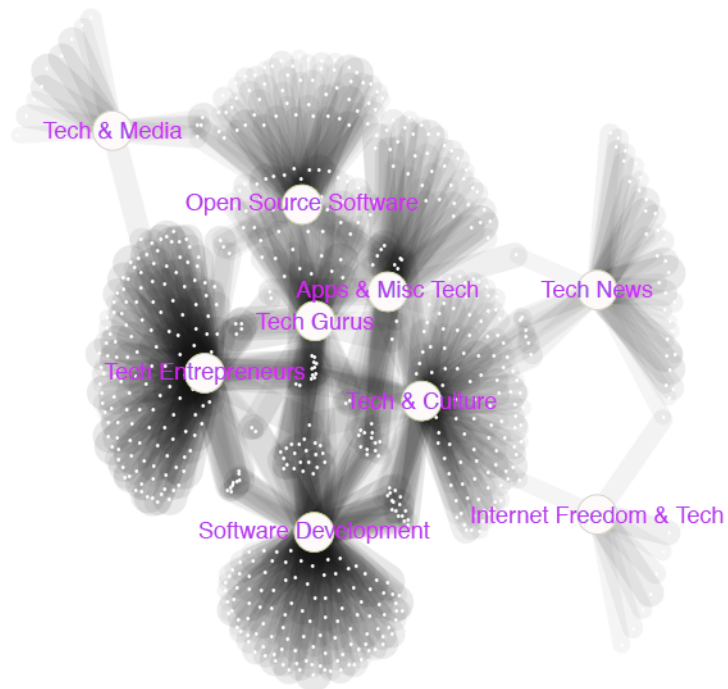
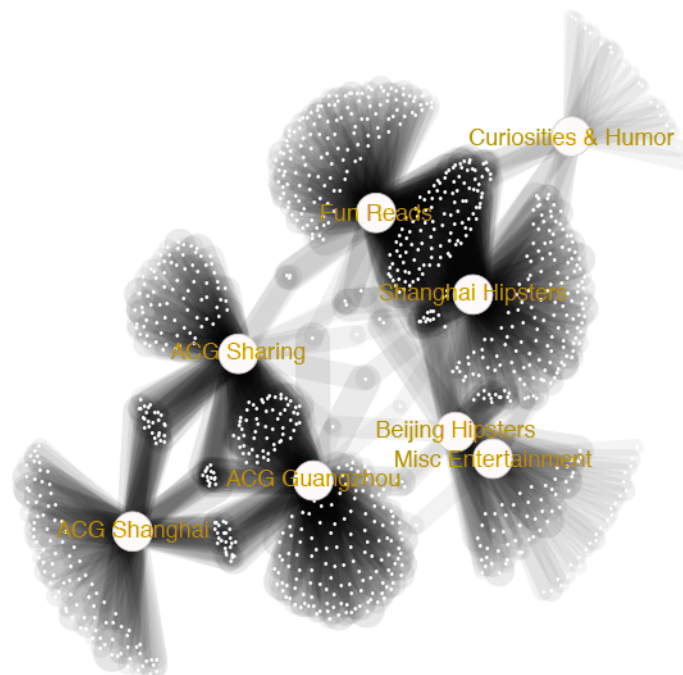"Retweet" Connections among Entertainment Clusters

FIGURE 5. TWO-MODE VISUALIZATIONS THAT REFLECT DYNAMIC INFORMATION FLOWS ACROSS CLUSTERS THROUGH RETWEETS

PROFILE PICTURES AND IDENTITY MANAGEMENT OF USERS

The fact that Twitter is outside of the reach of Chinese authorities does not mean that China does not seek to shape behavior on the platform. Some Chinese netizens have been arrested because of activity on Twitter. In 2012, Zhai Xiaobing published a political satire on Twitter before the 18th National Congress of the CPC and was charged with spreading terrorist rumors;[17] in 2014, Zhao Huaxu proposed on Twitter a technical idea for sharing knowledge about the Tiananmen Square Protest and was later arrested on charges of spreading illegal information.[18] Under such circumstances, we wondered whether Chinese users would prefer to hide their identities on Twitter to avoid risks. Because it is common for Internet users to adopt a nickname and be recognized by this handle, it is difficult to draw conclusions about a user's intention to remain anonymous solely by use of a pseudonym. Instead, we took profile pictures as a proxy for preferences regarding anonymity. Our assumption was that users' willingness to expose their identities would be correlated with the use of their photos in their profiles.

To answer this question, we retrieved user information of the 8,275 accounts in the sample through the Twitter API, downloaded the profile pictures of the existing users (about 1% of the accounts had been closed), and fed them into Face++, an online service specializing in detecting features of human faces, including gender, age, and race.[19] Face++ appeared to have identified real faces fairly well and accurately labeled cartoon characters and artistic portrayals as non-faces. Among the currently active users, Face++ recognized 28.4% as having one or more faces in their profile pictures, and found the remaining 71.6% to have no face at all. Among the profile pictures with faces, Face++ labeled 29.8% of them as female and the remaining 70.2% as male, and identified 66.3% as Asian, 3.0% as black, and 30.8% as white.

The automated approach to estimating use of real faces has a few limitations. First, Face++ missed some real faces when the photos were taken at an abnormal angle or heavily retouched. Another downside of the automatic processing is that we have no way to know when an account is using someone else's photo, e.g., a photo of a pop star. To assess the validity of the automated detection results, we manually checked a random sample of accounts for the presence of faces that appear to be of the account holder. Based on 367 cases drawn from the political cluster, we found that 25% of users included what appears to be their real face, which is close to the results of the automated assessment. For the manual check, we eliminated accounts that had no face, accounts with the face of a child, and accounts with the face of an easily recognized celebrity. For adult faces, we also conducted a Google Images search to look for matches that would indicate that a user had placed someone else's face in their profile.[I]

Based on the automated face detection results (see Figure 6), the users in the political group appear to show their real faces more than users in the technology and entertainment groups, instead of voicing opinions behind a mask. This result, which contradicts our expectations, is perhaps because openly voicing opinions with a disclosed identity is a calculated strategy to increase the impact and reach of advocacy efforts in full recognition of the risks involved. Celebrities and tech gurus inside and outside China also tended to publicize their faces, presumably to enhance their popularity. Interestingly, the entertainment group displayed real faces the least, even though they arguably bear a lower risk for their activities on Twitter. This phenomenon may come from the subculture among

---

[I] Based on the overall network of 8,275, the sample of 367 accounts corresponds to a confidence level of 95% and confidence interval of ±5%.

INTERNET MONITOR

the culture/entertainment users, as this group contained a large number of ACG fans who used cartoon characters as profile pictures.
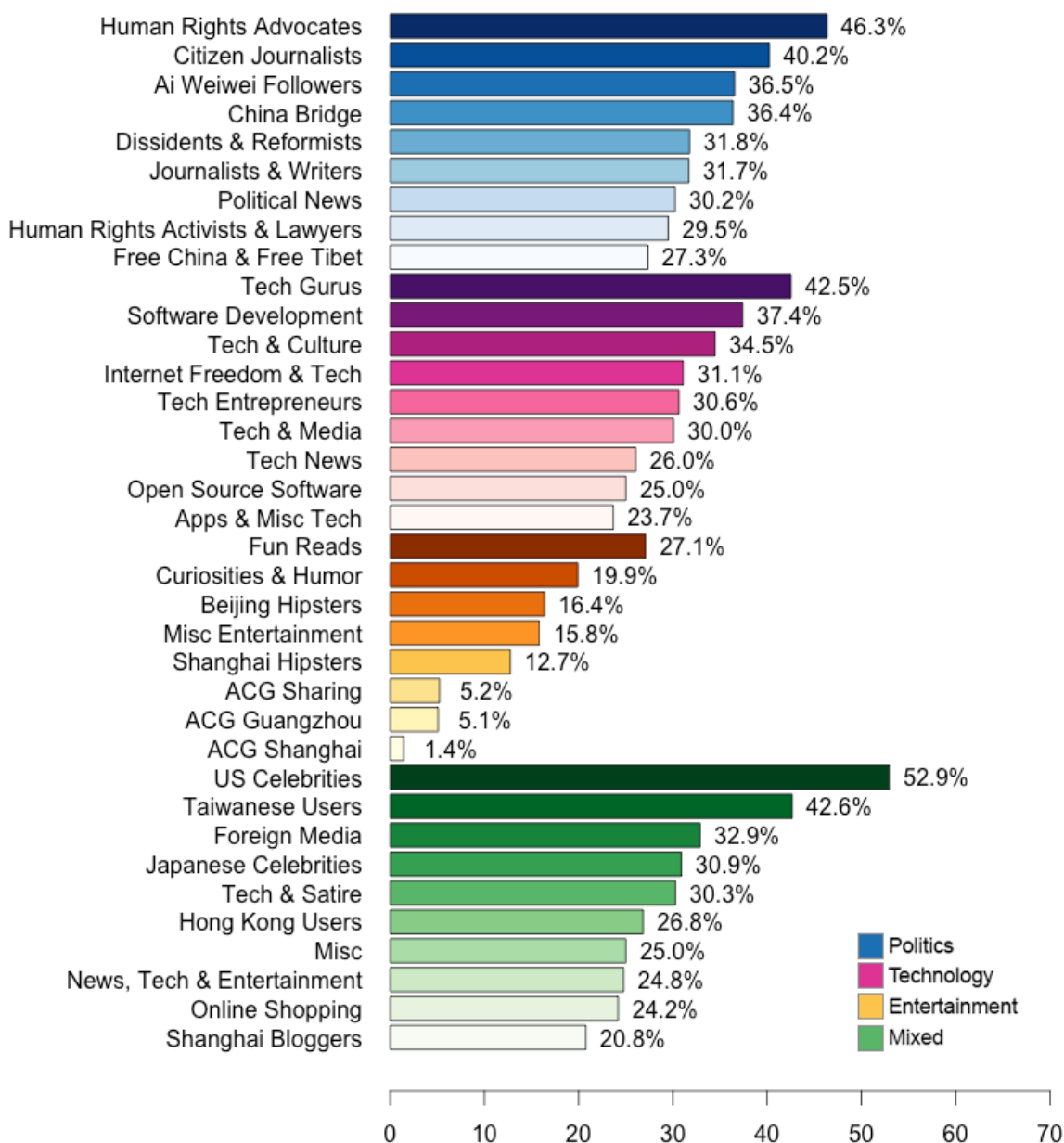

FIGURE 6. PERCENTAGE OF "REAL-FACE" PROFILE PICTURES ACROSS GROUPS AND CLUSTERS (FROM TOP TO BOTTOM: POLITICS, TECHNOLOGY, ENTERTAINMENT, MIXED)

SOCIAL NETWORK STRUCTURE

NETWORK STRUCTURE

In this section, we explore whether there are systematic differences in the structures of the various clusters. As described by Smith et al.: "Conversations on Twitter create networks with identifiable contours as people reply to and mention one another in their tweets. These conversational structures differ, depending on the subject and the people driving the conversation."[20] In the analysis that follows, we seek to describe in preliminary terms whether the structure of the clusters is correlated with the topical focus of these communities and offers insights into the functional relationships among users.

Twitter allows three possible follow relationships between each pair of users: no connection, one-directional connection, and mutual connection. Based on these follow relationships found in each of the clusters, we calculate three metrics to help interpret the differences in the structure of these clusters: density, mutuality tendency, and degree centralization.[J] Table 4 lists these descriptive statistics for each of the 36 clusters in the network, along with the number of users and connections within each cluster. The metrics are based solely on follow behavior and therefore denote longer-term static relationships and do not fully reflect users' interactions with others through retweets, replies, and mentions.

[J] In Appendix 2, we show a fourth measure, indegree centrality.

INTERNET MONITOR

GROUP	CLUSTER NAME	USERS	CONNECTIONS	% INSIDE GFW	DENSITY	MUTUALITY TENDENCY	DEGREE CENTRALIZATION
POLITICS	Human Rights Activists & Lawyers	215	8,357	89%	0.18	0.31	0.80
	China Bridge	86	1,021	51%	0.14	0.38	0.41
	Citizen Journalists	178	6,372	76%	0.20	0.36	0.74
	Dissidents & Reformists	252	2,824	63%	0.05	0.17	0.73
	Human Rights Advocates	355	49,186	80%	0.39	0.24	0.59
	Journalists & Writers	462	10,444	82%	0.05	0.18	0.84
	Free China & Free Tibet	273	3,800	69%	0.05	0.70	0.19
	Political News	278	8,851	83%	0.12	0.47	0.58
	Ai Weiwei Followers	217	1,538	78%	0.03	0.45	0.60
TECHNOLOGY	Open Source Software	48	865	96%	0.38	0.58	0.55
	Tech & Media	212	1,052	86%	0.02	0.19	0.83
	Software Development	222	1,806	88%	0.04	0.45	0.24
	Tech & Culture	292	9,194	86%	0.11	0.39	0.58
	Internet Freedom & Tech	177	3,638	90%	0.12	0.28	0.79
	Tech Gurus	254	9,928	88%	0.15	0.33	0.66
	Apps & Misc Tech	431	13,001	87%	0.07	0.39	0.47
	Tech News	214	4,689	89%	0.10	0.67	0.51
	Tech Entrepreneurs	167	1,387	16%	0.05	0.24	0.22
ENTERTAINMENT	ACG Guangzhou	118	1,181	80%	0.09	0.72	0.44
	ACG Sharing	133	5,384	86%	0.31	0.80	0.42
	ACG Shanghai	141	12,621	80%	0.64	0.69	0.29
	Beijing Hipsters	109	3,253	84%	0.28	0.81	0.49
	Misc Entertainment	272	14,391	88%	0.20	0.86	0.66
	Fun Reads	212	8,022	87%	0.18	0.76	0.48
	Curiosities & Humor	201	3,696	91%	0.09	0.36	0.72
	Shanghai Hipsters	202	16,862	74%	0.42	0.56	0.48
MIXED	Online Shopping	422	7,632	76%	0.04	0.64	0.16
	News, Tech & Entertainment	640	3,053	85%	0.01	0.45	0.16
	Foreign Media	69	48	26%	0.01	0.33	0.05
	Hong Kong Users	41	632	14%	0.39	0.46	0.49
	Japanese Celebrities	480	828	13%	0.01	0.26	0.04
	Shanghai Bloggers	106	3,264	74%	0.29	0.54	0.54
	Misc	433	5,419	84%	0.02	0.33	0.61
	Tech & Satire	193	3,980	71%	0.11	0.34	0.71
	Taiwanese Users	68	1,087	9%	0.24	0.47	0.39
	US Celebrities	34	24	14%	0.02	0.49	0.13

TABLE 4. DESCRIPTIVE STATISTICS OF THE 36 CLUSTERS

INTERNET MONITOR

Density is a measure of the overall level of connectivity among nodes in a network. This metric is calculated by the number of connections between users in a cluster as a proportion of the total possible number of connections.[K]
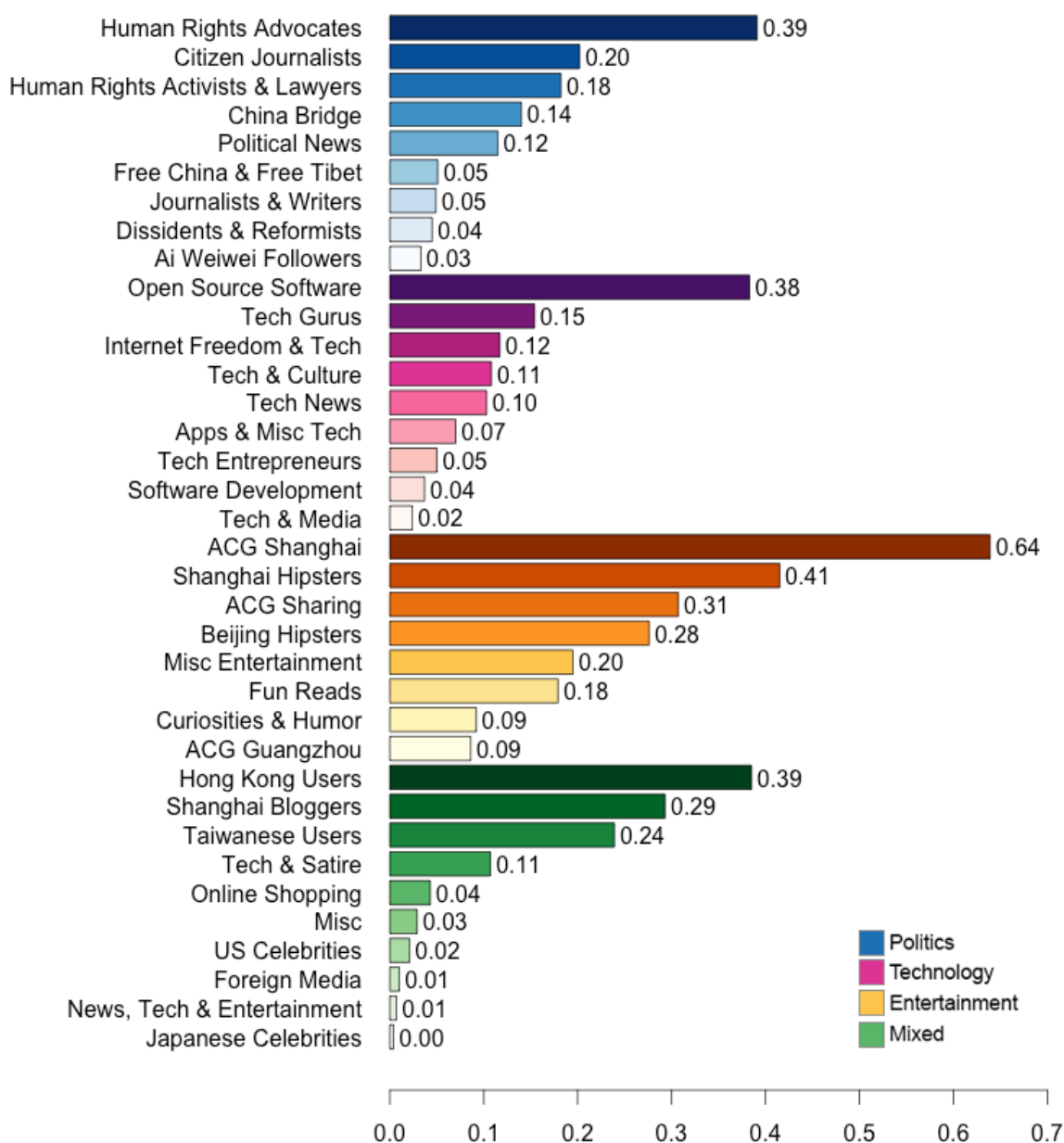

FIGURE 7. NETWORK DENSITY BY CLUSTER (FROM TOP TO BOTTOM: POLITICS, TECHNOLOGY, ENTERTAINMENT, MIXED)

[K] For each cluster with N nodes, there are N(N-1) possible connections (follows). Density is equal to the number of connections divided by N(N-1).

Density varies substantially across the different clusters (Figure 7). Generally speaking, if people get online to seek information and entertainment or to create and solidify social ties, we imagine those who seek out several of these purposes may use the platform more frequently and consequently form denser networks. Among the four groups (politics, technology, entertainment, and mixed), the entertainment group has the highest density scores, followed by the political and tech groups. The clusters with the highest density scores in each group are the **Human Rights Advocates** and **Citizen Journalists** clusters in the political group, **Open Source Software** and **Tech Gurus** in the technology group, **ACG Shanghai** and **Shanghai Hipsters** in the entertainment group, and **Hong Kong Users** and **Shanghai Bloggers** in the mixed group. One plausible explanation is that these groups are populated by more engaged users that seek out more information and are more motivated to network with likeminded people. It could be that some of the clusters have stronger commonality of interests than others, or some topics could attract more engaged users. Another possible explanation is that a higher proportion of members of these clusters know each other personally.

There are many clusters at the opposite end of the scale with low density levels (e.g., **Dissidents & Reformists**, **Journalists & Writers**, **Software Development**, **Tech Entrepreneurs**, **Curiosities & Humor**, **Fun Reads**, and **News, Tech & Entertainment**). The variation in density scores is higher within the four groups (politics, technology, entertainment and mixed) than the variation across the four groups, suggesting that the set of factors that influence density in these clusters applies to all of the topics.

The indegree centralization scores measure the extent to which connections are evenly divided across the nodes in each cluster or are more highly directed at a small number of nodes. A high centralization score means that a small number of prominent nodes are the focus of attention. Low density scores combined with high centralization scores indicate that some of these clusters were formed around star writers, activists, and technologists, as seen in the **Journalists & Writers**, **Tech & Media**, **Human Rights Activists & Lawyers**, **Dissidents & Reformists**, and **Curiosities and Humor** clusters. The members of these clusters may be satisfied to receive information from the hub users and less motivated to build ties with others in the cluster. In the case of the **Tech & Media** cluster, the low density measures may be explained by the fact that the top users in this cluster, @hecaitou and @williamlong, are also active on Sina Weibo and have attracted about 400,000 followers on that platform. This cluster on Twitter could be a shadow of a more active conversation on Weibo. The two top users mainly discuss technology, which is acceptable in China, thereby reducing the need for an alternative, less regulated platform.

The **Dissidents & Reformists** cluster also had an extremely low density score, suggesting the users in this cluster either lack interest in forming connections or are deterred from doing so. If nothing else, this suggests that the Great Firewall is effective at isolating dissidents from each other and from those outside their circles. The low level of connectivity could reflect low public attention Chinese dissidents receive because human rights violations and dissidence are rarely reported by Chinese media. Although the *New York Times* (NYT) and the Hong Kong-based *South China Morning Post* (SCMP) cover these issues much more than the *China Daily*,[21] non-Chinese media may fail to boost the fame of human rights activists among mainland Chinese for two reasons, implied by Krumbein's findings. First, the SCMP and NYT mainly cover Chinese issues around major events, such as the US President's visit to China and the Beijing Olympics. Second, they pay repeated attention to the Tiananmen Square anniversary crackdowns and provide less coverage for other events. In addition,

the language barrier hinders Chinese speakers from learning about dissidents from their own country.

RECIPROCITY AND EQUALITY

Since Twitter allows asymmetric connections across users, mutuality (reciprocity) may reflect different underlying social dynamics. In particular, high reciprocal connections may indicate familiarity, comparable prominence, or the exchange of information around shared interests. For example, friends make mutual connections on Twitter, whereas entertainers, politicians, and tech gurus rarely reciprocate the connections with their fans, supporters, or admirers. Homophily and equality appear to be important aspects in forming reciprocal relationships on Twitter. Weng et al. find that a large portion of reciprocal ties on Twitter might be explained by common interests.[22] Wu et al. find significant homophily within different categories of Twitter users with "celebrities following celebrities, media following media, and bloggers following bloggers."[23] Kwak et al. show that the number of followers in reciprocal follow relationships ("r-friends") are correlated, implying a measure of equality in mutual relationships on Twitter.[24] Golder and Yardi say that "mutuality might be a proxy for equal status."[25]

As a measure of reciprocity, we calculated mutuality tendency, which measures the conditional probability of a reciprocal follow relationship for each follow relationship.[26] Among the four groups, entertainment had a higher tendency towards mutuality compared to the political and tech groups (Figure 8). The high tendency might indicate that more of these people are friends, or could reflect the exchange of media and information, such as animation, comics, games, and fun activities, or that the users tended to share more similar levels of prominence in their clusters. Within the tech group, **Tech Gurus** and **Tech Entrepreneurs** were less likely to follow back, whereas the **Tech News** and **Software Developers** clusters were more likely to make mutual connections, perhaps because they sought each other's technical and emotional support and relied on each other for the latest tech updates. In the same light, among the political clusters, **Free China & Free Tibet** advocates and **Political News** followers had the highest reciprocity score, perhaps also driven by information and support seeking. The least reciprocal clusters include **Dissidents & Reformists**, **Tech & Media**, and **Journalists & Writers**, which feature high-profile dissidents, tech critics, and famous writers, respectively. These top users had more followers than followees, suggesting these clusters were highly biased toward the "stars."
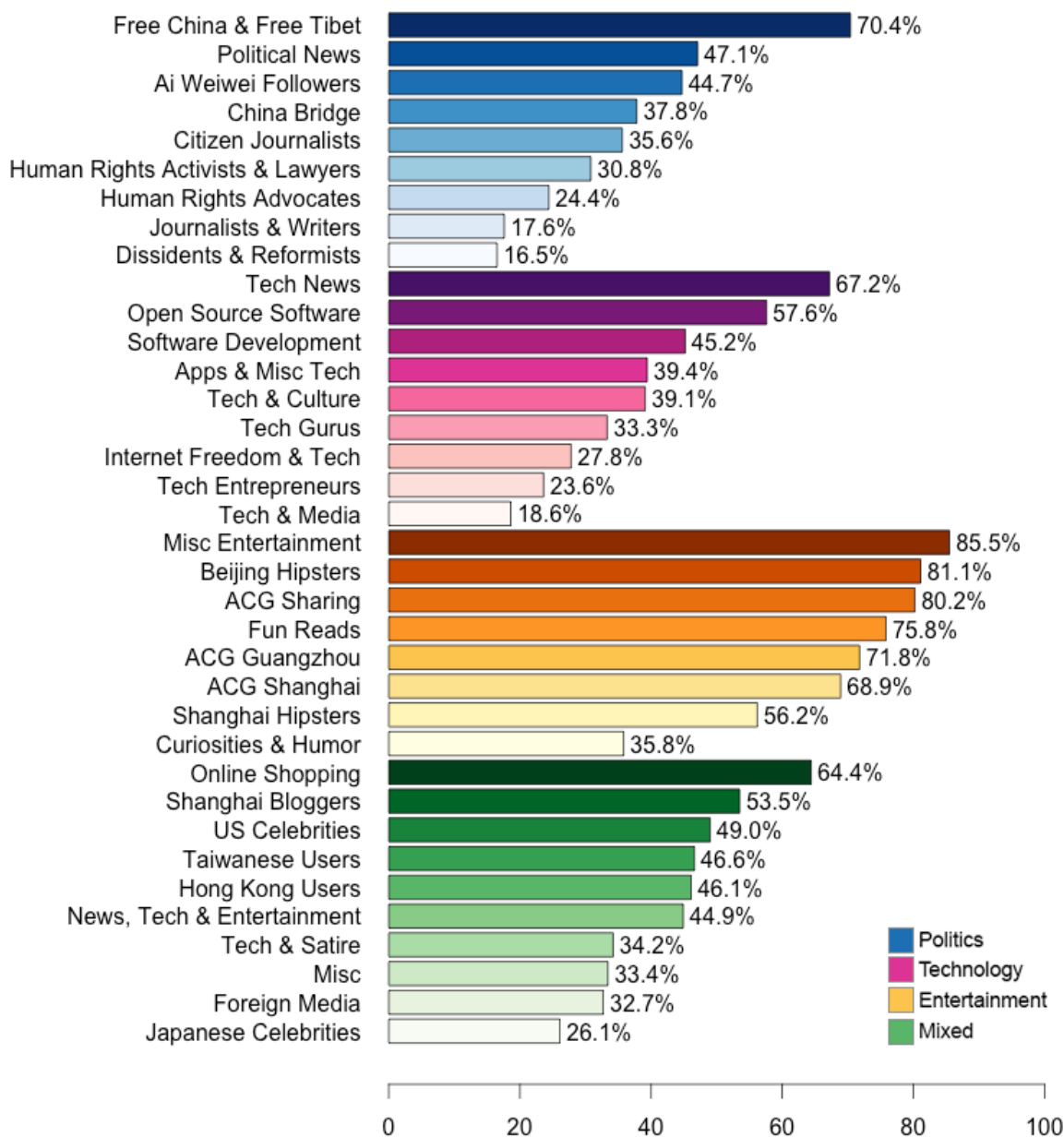
INTERNET MONITOR

FIGURE 8. CONDITIONAL PROBABILITY OF RECIPROCITY (MUTUALITY TENDENCY) BY CLUSTER (FROM TOP TO BOTTOM: POLITICS, TECHNOLOGY, ENTERTAINMENT, MIXED)

We find considerable variation in network structure across the various clusters, which likely reflects different informational and behavioral forms. These different structural forms are found whether discussing politics, technology, or entertainment. In Chinese Twitter, topical coverage does not determine network structure. There are, however, broad differences between the clusters, as shown in Tables 5 and 6. The clusters in the entertainment group are more highly connected and more likely to form reciprocal ties. The political and technology groups are comparatively less dense and more centralized, and their users less likely to form reciprocal ties.

	% INSIDE GFW	DENSITY	INDEGREE CENTRALIZATION	MUTUALITY TENDENCY
Politics group	78%	0.12	0.60	0.36
Technology group	88%	0.10	0.55	0.39
Entertainment group	85%	0.24	0.48	0.74
Mixed group	49%	0.03	0.28	0.46

TABLE 5. MEDIAN CLUSTER STATISTICS BY GROUP

	% INSIDE GFW	DENSITY	INDEGREE CENTRALIZATION	MUTUALITY TENDENCY
Politics group	63%	0.09	0.77	0.35
Technology group	79%	0.04	0.61	0.33
Entertainment group	77%	0.12	0.51	0.70
Mixed group	55%	0.02	0.43	0.36

TABLE 6. NETWORK STATISTICS BY GROUP[L]

CONCLUSIONS AND FUTURE DIRECTIONS

This study confirms that Twitter is used by residents of mainland China to follow popular accounts that are not found on Sina Weibo or other Chinese microblogging platforms. The focus of attention includes political, technological, and cultural topics. For instance, users follow entrepreneurs from San Francisco and New York and human rights activists and organizations inside and outside China. Content from and about many of the prominent figures on Twitter is blocked in China and not available on other platforms. Despite the scale and scope of Chinese social media activity on domestically hosted platforms, there is a demand for content and conversations not found within the Great Firewall.

Several of our expectations prior to carrying out this research are confirmed by the analysis. We find a major focus of Chinese Twitter to be on politically contentious topics that would be blocked on domestically hosted platforms. The presence of these groups is consistent with the commonly stated proposition that there is a small group of highly motivated political activists who are willing to take the necessary steps to get around Internet censorship.

The political crowd, who would face the highest risks for their online speech, appear to be the least likely to seek anonymity. The fact that many of these advocates openly tweet under their real names and include their face in their profiles suggests a willingness to risk their personal freedom, or alternatively a belief that international recognition may help to shield them from persecution.

It is not surprising to find groups of technologists on Chinese Twitter. This is consistent with the idea that it is relatively simple to circumvent the firewall for those that are technologically adept. The presence of several clusters that focus much of their attention on culture and entertainment would have been harder to predict.

In some ways, the discourse in the politically engaged portions of Chinese Twitter suggests that this is indeed an alternative public sphere in which networked individuals can cooperate in a peer-

---

[L] Table 6 treats each of the groups as a single analytical unit, pooling all of the users from each group into a single network.

produced system that collectively filters and highlights areas of public concern with less reliance on large media organizations and under less government control, allowing a shift towards more bottom-up agenda setting and framing of issues.[27] However, the seemingly sparse connections between Chinese dissidents and their compatriots may be the result of limited domestic media coverage and a language barrier between users and Western media outlets.

Chinese Twitter falls well short of supporting an inclusive and broadly accessible networked public sphere. The proportion of the Chinese populace with direct access to the debates, communities, and shared resources on Twitter is very small, and the avenues by which such discourse might find its way into mainstream political discussion are severely constrained. The firewall between Twitter and the much larger social media platforms in China appears to form a formidable barrier. Future research may shed further light on the effectiveness of indirect methods of information diffusion across these separate networks. One possibility is word of mouth. Another is the injection of perspectives, topics, and frames from Twitter into Weibo, although any such transfer would have to survive the domestic filters.

Tracking the evolution of Chinese Twitter over time may help to answer some of the outstanding questions. One possible trajectory would be the addition of new users interested enough in more open discussion of political issues to surmount the obstacles of the Great Firewall. It is also easy to imagine a growing number of technologically proficient Chinese netizens joining international social media platforms. Another complementary path would be the addition of more users to Twitter and similar platforms that are neither among the most highly engaged politically nor among those interested in technology, which would cover a much wider cohort of Internet users. It may be instructive to learn more about the motives for users in the culture and entertainment groups to join Twitter.

There are several areas where future research will help us to better understand the impact and reach of such Internet enclaves set apart from their larger media spaces by technological and regulatory barriers. More theoretical and empirical work is needed to understand the relationship between the structural form of these networks and the functional aspects of communication and community building via digital platforms. One limitation of this study is the difficulty in interpreting how the characteristics of these digital networks may or may not reflect a growing and vibrant public sphere. Studying the evolution of these networks over time will provide some answers. Comparing the form and function of Chinese Twitter to other similar networks in different regulatory contexts may be fruitful as well. Another useful extension of this work would be to conduct interviews with participants to help to bridge the gap between the digital record and user perceptions on their motivations, intent, and purpose of Twitter activity.

APPENDIX 1: ASSESSING USER LOCATION

We cannot establish with certainty the location of Twitter users. There are, however, several sources of information that help us to assessing the likely location of the users in the network. The metadata for many of the accounts include longitude and latitude coordinates. We also have access to user-supplied information in three fields of Twitter profiles: language, time zone, and location. Since each of these fields is user defined, we do not expect them to be fully reliable and treat these data with caution.

INTERFACE LANGUAGE

Users can select the interface language on Twitter. In addition to choosing between English, Spanish, or Chinese, for instance, Chinese can be specified as simplified or traditional, the two major Chinese systems exclusively adopted by various countries in the world. Users of simplified Chinese are more likely to come from mainland China, whereas those of traditional Chinese are more likely to come from Taiwan or Hong Kong. In our sample, however, the mostly frequently chosen language is English, rather than Chinese of either type. English is the default option, and many users might not find it necessary to change this, even if they still tweet mostly in Chinese. Even though language is not a precise proxy for mapping users' locations, we found that language preference was somewhat correlated with users' interests.[M] Figure 9 below shows that the users in the political group are more likely to configure the interface as Chinese, while English is more common in the technology group.
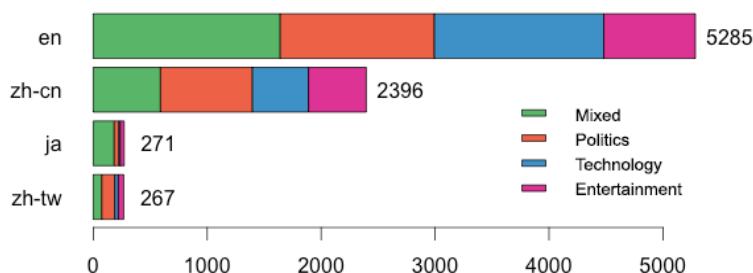

FIGURE 9. DISTRIBUTION OF LANGUAGE PREFERENCES (EN – ENGLISH; ZH-CN - SIMPLIFIED CHINESE FOR MAINLAND CHINA; ZH-TW - TRADITIONAL CHINESE FOR TAIWAN; JA - JAPANESE)

[M] This relationship is statistically significant: *Chi-squared*=157.73, *df*=6, *p-value*<2.2e-16.
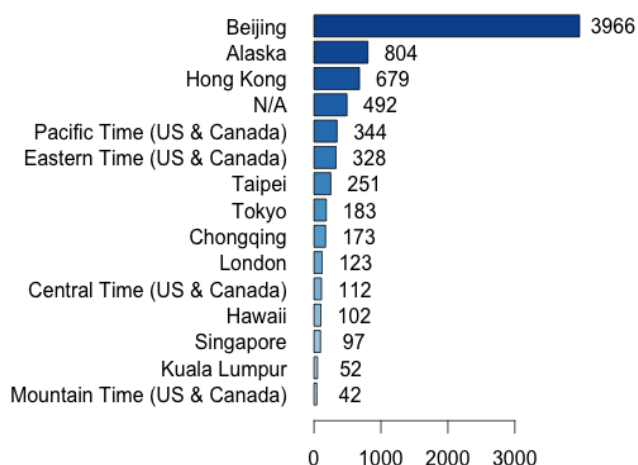
INTERNET MONITOR

TIME ZONE

FIGURE 10. DISTRIBUTION OF TIME ZONE SETTINGS

The majority of the users selected Beijing as their time zone. However, there were at least two complications that prevented us from using time zone as a precise measure for the geographic distribution. First, time zone is too widespread, with one time zone marking approximately one 24th of the earth from the North to the South Pole. For example, a number of places are situated in the same time zone as Beijing (UTC+8), such as Taiwan, Hong Kong, Singapore, Western Australia, and the Philippines. Second, users may deliberately choose a time zone that does not coincide with their actual location.

Figure 10 shows that, following Beijing, Alaska was the second most popular time zone. This anomaly might be explained by discussions on Twitter revealing that some Chinese users purposely selected an American time zone (Alaska or Pacific) to enable an archive feature (see Figures 11 and 12). This feature, which enables users to download an archive of one's tweets, was released in December 2012, and was reported to be available only for users in the United States.[N]

[N] Mollie Vandor, "Your Twitter Archive," Twitter Blog, December 19, 2012, https://blog.twitter.com/2012/your-twitter-archive.
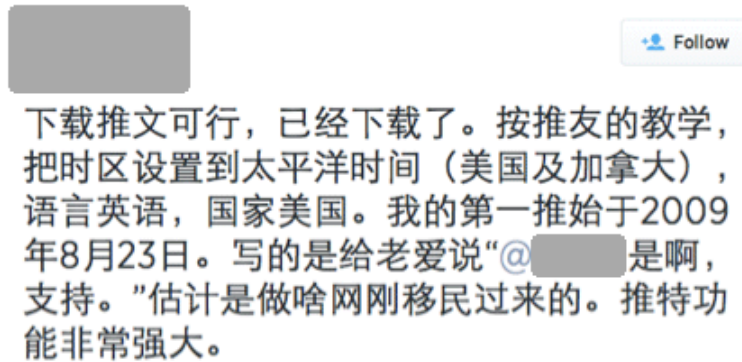
FIGURE 11. A USER REPORTED THAT HE HAD SUCCESSFULLY DOWNLOADED HIS TWEET ARCHIVE AFTER SETTING HIS TIME ZONE AS PACIFIC TIME.

FIGURE 12. ONE USER INDICATED THAT SHE FAILED TO DOWNLOAD HER TWEET ARCHIVE, PERHAPS BECAUSE HER TIME ZONE WAS BEIJING; THE SECOND USER INDICATED THAT HE HAD SUCCESS BECAUSE HIS TIME ZONE WAS SET AS ALASKA.

Another reason why some Chinese users chose a non-Chinese time zone is because they apparently want to distance themselves from the country. For example, a user explained that she always uses Hong Kong as the time zone because a candlelight vigil is held there every year as a memorial to the Tiananmen Square Protest (Figure 13). Chinese users who choose a non-Chinese time zone often find Hong Kong and Taipei more convenient than others because these are consistent with Beijing, meaning the user's experience on Twitter is not altered.

FIGURE 13. THE PHOTO EMBEDDED IN THIS TWEET SHOWS A CANDLELIGHT VIGIL IN HONG KONG AS A MEMORIAL TO THE TIANANMEN SQUARE PROTEST. THE USER CLAIMED THAT SHE ALWAYS CHOOSES HONG KONG AS HER TWITTER TIME ZONE BECAUSE THIS EVENT IS HELD IN HONG KONG ANNUALLY.

USER REPORTED LOCATION

Although language and time zone fail to provide reliable information about user locations, the "location" field on the user profile provided better insights. The location data, which were retrieved through the Twitter API, generally fell into one of five categories: known places, recognizable nicknames for certain places, coordinates, unrecognizable places, and blank. First, many users set their locations with known places written in various languages, such as Boston, 北京 (Beijing/Peking), and (Seoul). Occasionally, users listed multiple known places in this field, such as Nanjing/New York. In this case, we assigned the first value to the given user. We then converted these known places to coordinates using the GeoNames API.[O] When an upper level region was given as a location, a lower-level place (mostly capitals or populous places) was returned by the GeoNames, e.g., US converted to New York, China to Beijing, and California to Los Angeles. For this reason, Beijing and New York were overrepresented because they were used to represent users living in China or the US who did not specify a location at a lower level.

The second category consisted of places with a known nickname. For instance, China was nicknamed as 西朝鲜 (West Korea), 天朝/天朝 (Heaven Dynasty, a historic and sinocentric way for a Chinese dynasty to refer to itself, and a term now used as satire by Chinese netizens), "Ceramic Country," "Country of Harmony," and "拆哪" (which resembles the pronunciation of Chi-Na and means "demolish where" in Chinese). In addition, Beijing, Shanghai, and Guangzhou, the largest

---

[O] GeoNames, "About GeoNames," http://www.geonames.org/about.html.

cities in China, are often referred as 帝都 (imperial capital), 魔都 (magic capital), and 妖都 (yōkai capital). Other "capitals" included 伪都 (pseudo-capital, Wuhan), 旧都 (old capital, Nanjing), and 废都 (abandoned capital, Xi'an). These nicknames were manually translated to standard geonames in English and automatically converted to coordinates using the GeoNames API.

LONGITUDE AND LATITUDE COORDINATES

The fourth source of location information for some accounts is longitude and latitude coordinates which was included as part of the metadata. Some of these coordinates included a prefix such as ÜT or iPhone, which we suspect are fingerprints left by Über Twitter or other third-party applications. After the prefixes were removed, these coordinates could be used directly to map users. The dataset for this study included longitude and latitude coordinates for 1,378 users, which accounts for 16.7% of the sample.

Another 10.4% of the users reported unrecognizable geonames. Among them, some were metaphorical locations, e.g., 墙内/墙外 (inside/outside the [Great Fire] Wall) and "404 not found"; some were fictitious places, e.g., "Gotham City," "Airstrip One, Oceania," and "Oz"; some were statements, e.g., "can't tell you because I may be chased after." The remaining 16.7% of the users left the location field blank.

MAPPING GROUPS AND CLUSTERS

After we normalized the locations and converted them into longitude and latitude, we created heatmaps with the help of Google Fusion Tables to illustrate users' geographic distribution (Figure 14).[P] Heatmaps better illustrate the density of users, and when weighted on follower count, influence can be observed geographically. Overall, most users in the sample were from Beijing/China, Shanghai, and Tokyo. However, if weighted by follower count, we found more influential users were from Tokyo and the US coasts, which indicates the "super nodes" in China had many fewer followers compared to their counterparts elsewhere because Twitter does not have a mass audience in the country.

Twitter users in the political, cultural, and technology groups were mostly inside China, whereas those in the mixed group were mostly from outside the country. On the cluster level, some showed tight geographic distribution, e.g., the **Hong Kong Users** and **Taiwanese Users** clusters, like various clusters in the Arabic blogosphere (Etling et al. 2010). There are various reasons to explain why geographic proximity helps cluster people. First, people living in the same area have more chances to meet in person and build relationships. Second, geographic clustering represents other similarities among people, such as cultural background and socioeconomic status, and therefore facilitates connections among people. The **Tech Entrepreneurs** cluster had its leading accounts based in San Francisco and New York, and the **ACG Guangzhou** cluster had influential users from Guangzhou, where the ACG industry prospers.

[P] To interact with the Google Fusion Table, visit the following link and find instructions in "about the table": http://brk.mn/chinatwitterfusiontable.

INTERNET MONITOR

HEATMAPS

Political Group
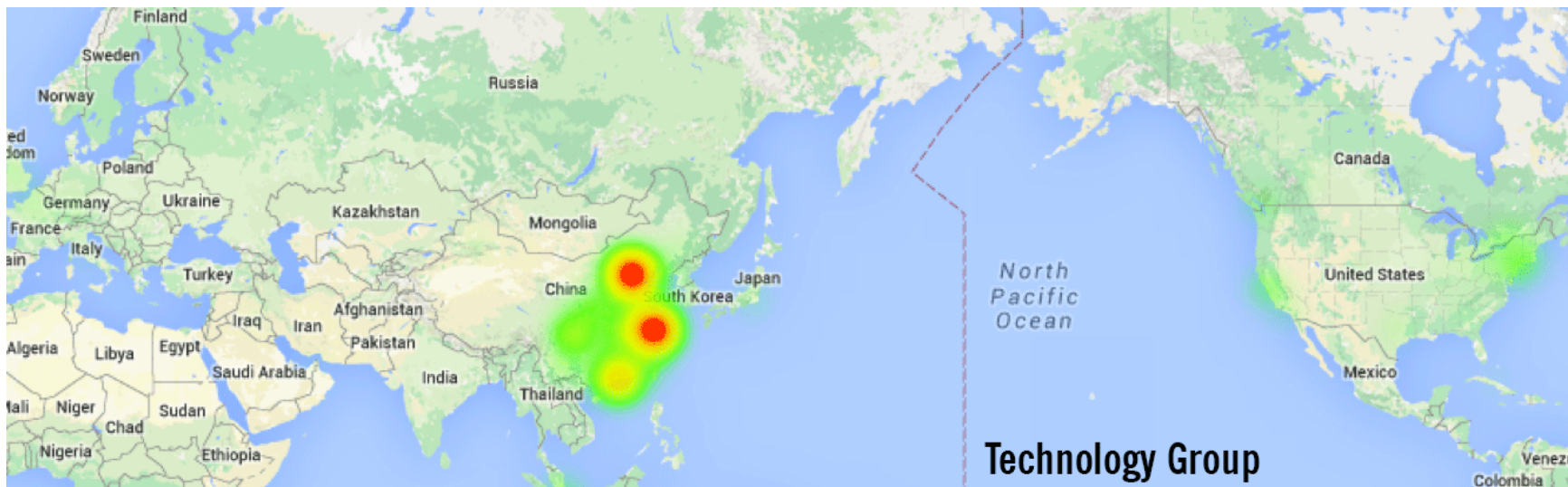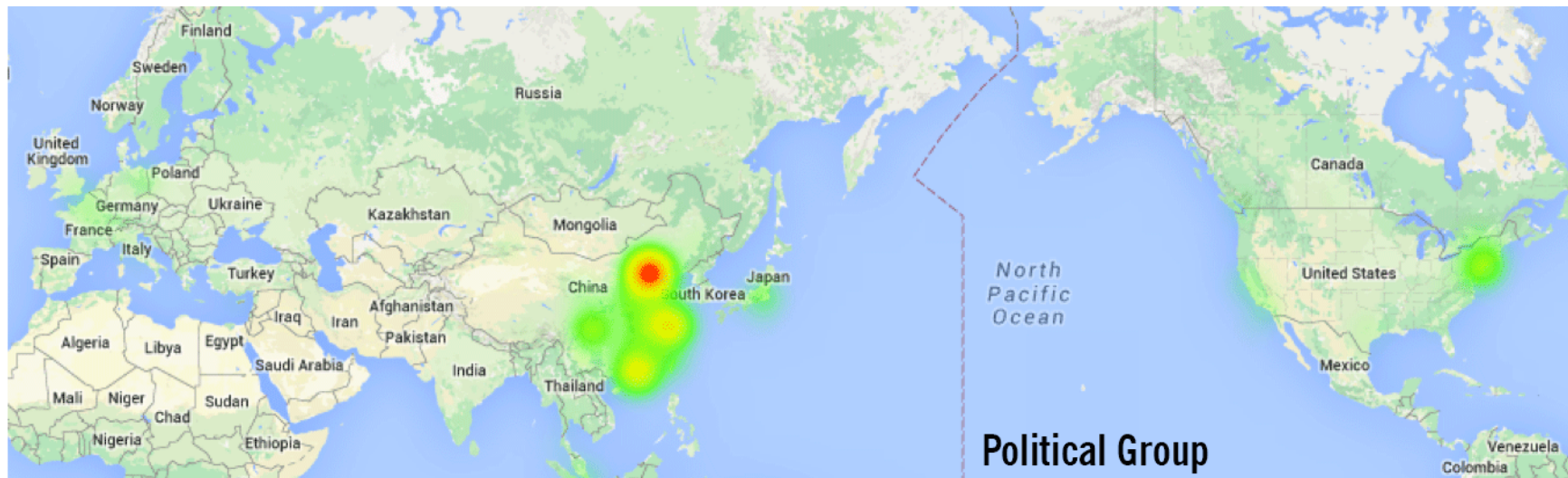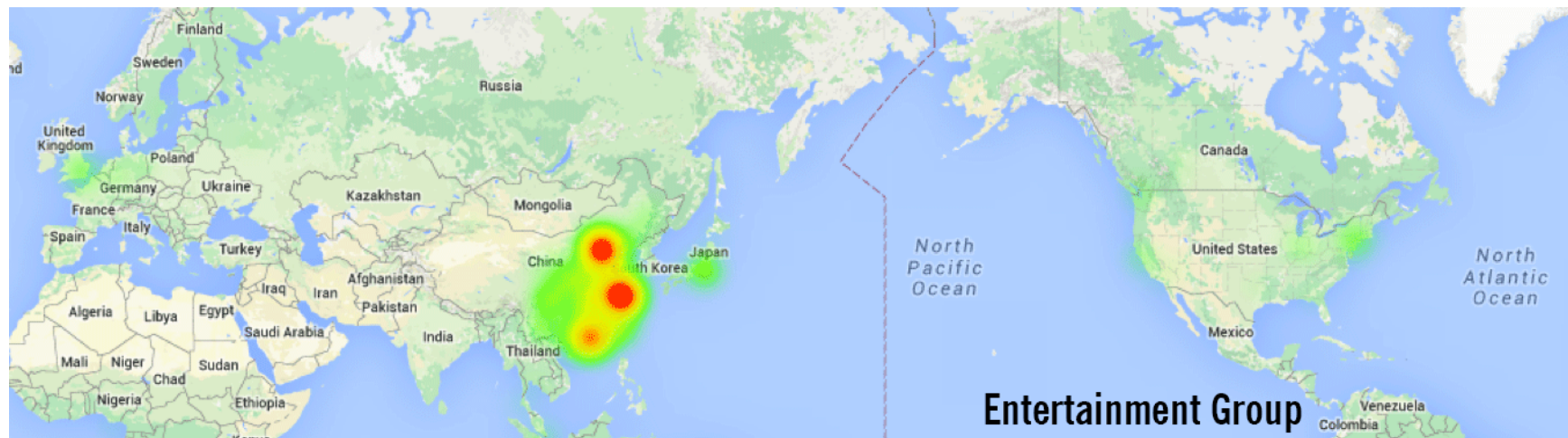

Technology Group

FIGURE 14. GEOGRAPHIC DISTRIBUTIONS OF THE POLITICAL, TECHNOLOGICAL, AND ENTERTAINMENT GROUPS

INSIDE OR OUTSIDE THE GREAT FIREWALL

As described earlier, we used these data to assess whether users are located within or outside of the Great Firewall. For this broader scale distinction, we also draw on time zone information, in addition to the user-provided location data and geographic coordinates. Drawing on these data, we are able to infer location for a larger portion of the users in the network, including some with conflicting information or dubious reliability. In using time zone information, we make two key assumptions: 1) Although some Chinese people set their time zones outside the country for ideological or technical reasons, we assumed that users outside China would not configure their time zones to locations within mainland China unless they lived there, and postulated that people with a Chinese time zone (Beijing, Chongqing, and Urumqi) were in fact residing in China. 2) If the Alaska time zone was chosen, but other information indicated a Chinese user, we assumed that access was from inside China. When these two rules failed to apply, we marked the data with missing values.

APPENDIX 2: INDEGREE CENTRALITY

There are several measures of the centrality of nodes in networks that serve as metrics for assessing the prominence of these nodes. Classical centrality measures are based on degree, closeness, betweenness, and eigenvector (Freeman 1979; Wasserman & Faust 1994).[28] Here we show the distribution of indegree centrality for the nodes in each cluster. Indegree centralization is based on the number of followers a particular node has within a cluster, or indegree. To account for differences in the size of clusters, the indegree centrality scores were normalized.[Q] The black bars in the boxes mark the median normalized indegree centralities, ranging from 0 to 1. The average or median indegree score is highly correlated with the density metrics presented earlier.
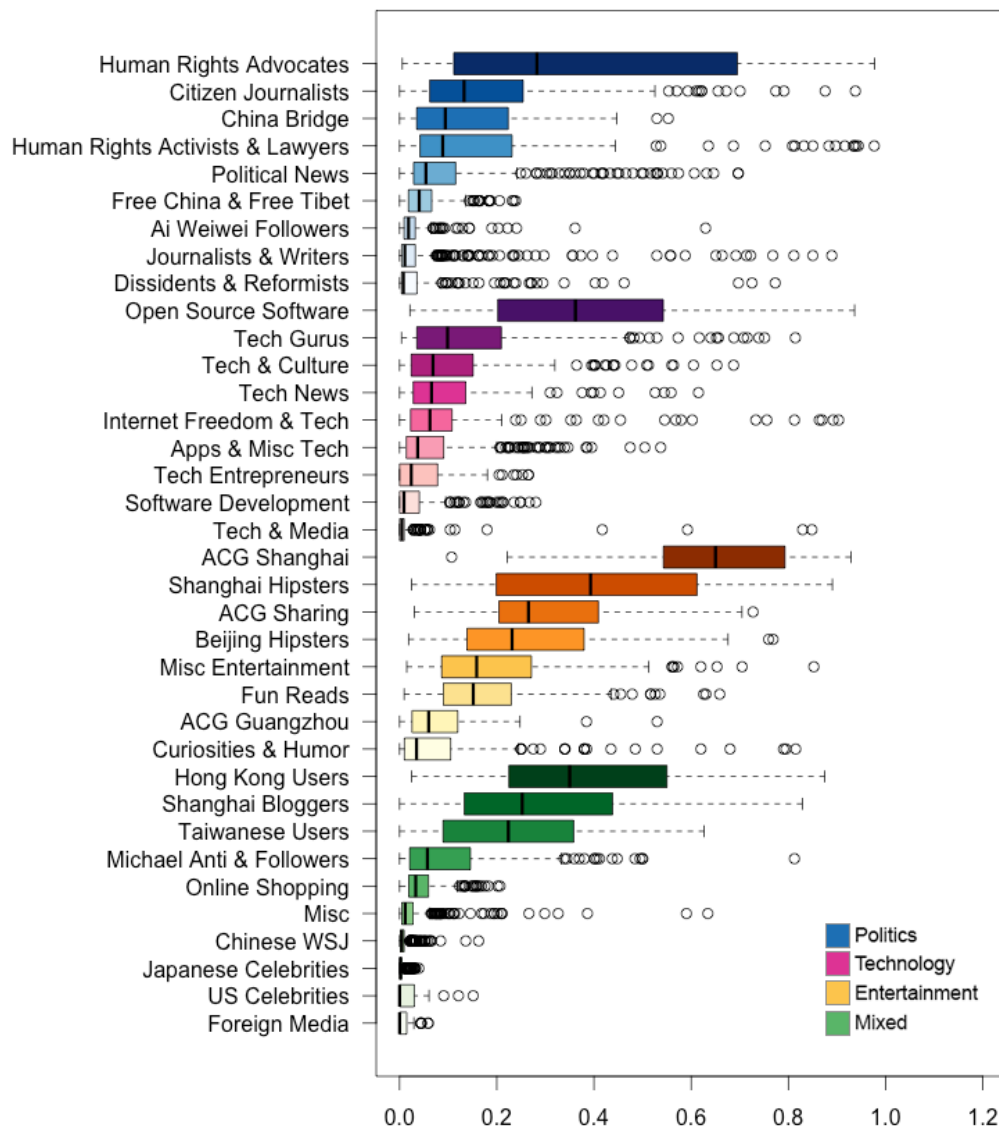

FIGURE 15. DISTRIBUTION OF INDEGREE CENTRALIZATION BY CLUSTER

[Q] The indegree centralization scores are calculated by dividing the number of followers for each node from within the cluster by the number of nodes in the cluster minus one.

INTERNET MONITOR

NOTES

[1] OpenNet Initiative, "Country Profile: China," August 9, 2012, https://opennet.net/research/profiles/china. Freedom House, "Freedom in the World 2014: China," 2014, https://freedomhouse.org/report/freedom-world/2014/china. Reporters Without Borders, "Enemies of the Internet 2014: entities at the heart of censorship and surveillance," 2014, http://12mars.rsf.org/2014-en/enemies-of-the-internet-2014-entities-at-the-heart-of-censorship-and-surveillance/.

[2] Rebecca MacKinnon, "China's Censorship 2.0: How companies censor bloggers," *First Monday* 14 (2009), doi:10.5210/fm.v14i2.2378.

[3] Gary King, Jennifer Pan, and Margaret E. Roberts, "How censorship in China allows government criticism but silences collective expression," *American Political Science Review* 107 (02): 326-343, doi:10.1017/S0003055413000014. Gary King, Jennifer Pan, and Margaret E. Roberts, "Reverse-engineering censorship in China: Randomized experimentation and participant observation," *Science* (345), doi:10.1126/science.1251722.

[4] D. Robinson, H. Yu, and A. An, "Collateral freedom: A snapshot of Chinese internet users circumventing censorship," Open Internet Tools Project Report, 2013, https://openitp.org/pdfs/CollateralFreedom.pdf.

[5] Hal Roberts, Ethan Zuckerman, John Palfrey, "2011 Circumvention Tool Evaluation," Berkman Center Research Publication, August 18, 2011, https://cyber.law.harvard.edu/publications/2011/2011_Circumvention_Tool_Evaluation.

[6] Robinson et al., "Collateral freedom."

[7] Meeyoung Cha, Hamed Haddadi, Fabríco Benevenuto, Krishna P. Gummadi, "Measuring user influence in Twitter: The million follower fallacy," Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (2010).

[8] Scott Golder and Sarita Yardi, "Structural predictors of tie formation in Twitter: Transitivity and mutuality," Proceedings of the Second IEEE International Conference on Social Computing (2010), http://research.microsoft.com/apps/pubs/default.aspx?id=133158.

[9] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng, "Why we Twitter: Understanding Microblogging Usage and Communities," Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, 2007, http://ebiquity.umbc.edu/paper/html/id/367/Why-We-Twitter-Understanding-Microblogging-Usage-and-Communities.

[10] Deen Freelon, " On the interpretation of digital trace data in communication and social computing research," *Journal of Broadcasting & Electronic Media* 2 (2014), 58(1):59-75, doi:10.1080/08838151.2013.875018.

[11] Bruce Etling, John Kelly, Robert Faris, and John Palfrey, "Mapping the Arabic Blogosphere: Politics and Dissent Online," *New Media & Society* 12 (2010), 1225-1243, doi:10.1177/1461444810385096. John Kelly, Vladimir Barash, Karina Alexanyan, Bruce Etling, Robert Faris, Urs Gasser, and John Palfrey, "Mapping Russian Twitter," Berkman Center Research Publication, March 20, 2012, https://cyber.law.harvard.edu/publications/2012/mapping_russian_twitter.

[12] Edward Wong, "China Admits Building Flaws in Quake," *New York Times*, September 4, 2008, http://www.nytimes.com/2008/09/05/world/asia/05china.html.

[13] Edward Wong, "China Presses Hush Money on Grieving Parents," *New York Times*, July 24, 2008, http://www.nytimes.com/2008/07/24/world/asia/24quake.html.

[14] James Reynolds, "Shutting us out?," BBC, June 13, 2008, http://www.bbc.co.uk/blogs/thereporters/jamesreynolds/2008/06/shutting_us_out.html.

[15] Gábor Csárdi and Tamás Nepusz, "The igraph software package for complex network research," *InterJournal* Complex Systems 1695**,** 1-9, http://www.necsi.edu/events/iccs6/papers/c1602a3c126ba822d0bc4293371c.pdf.

[16] James G. Webster, *The Marketplace of Attention*, (Cambridge, MA: MIT Press, 2014).

[17] Gillian Wong, "Zhai Xiaobing, Chinese Blogger, Arrested For Twitter Joke About China's Government," *Huffington Post*, November 21, 2012, http://www.huffingtonpost.com/2012/11/21/zhai-xiaobing-arrested-china-twitter-joke_n_2170462.html.

[18] "Young Chinese Twitter User Arrested for Proposing Method to Spread Truth about June 4th Massacre," *China Change*, June 9, 2014, http://chinachange.org/2014/06/09/young-chinese-twitter-user-arrested-for-teaching-a-method-to-spread-illegal-information/.

[19] Face++, http://www.faceplusplus.com/.

[20] Marc A. Smith, Lee Rainie, Ben Shneiderman, and Itai Himelboim. "Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters" Pew Research Center, February 20, 2014, http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/.

[21] Frédéric Krumbein, " Media coverage of human rights in China," *International Communication Gazette* (2014), doi:10.1177/1748048514564028.

[22] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He, "TwitterRank: finding topic-sensitive influential twitterers," Proceedings of the third ACM international conference on Web search and data mining (2010), Association for Computing Machinery, 261-270, doi:10.1145/1718487.1718520.

[23] Shaomei Wu, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts, "Who says what to whom on Twitter," Proceedings of the 20th international conference on World Wide Web pages (2011), Association for Computing Machinery, 705-714, doi:10.1145/1963405.1963504.

[24] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon, "What is Twitter, a social network or a news media?," Proceedings of the 19th international conference on World Wide Web pages (2010). Association for Computing Machinery, 591-600, doi:10.1145/1772690.1772751.

[25] Golder and Yardi, "Structural predictors of tie formation in Twitter."

[26] Leo Katz and James H. Powell, "Measurement of the tendency toward reciprocation of choice," *Sociometry* (18): 403-409, doi:10.2307/2785876.

[27] Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven, CT: Yale University Press, 2007). Bruce Etling, Hal Roberts, and Robert Faris, "Blogs as an Alternative Public Sphere: The Role of Blogs, Mainstream Media, and TV in Russia's Media Ecology," Berkman Center Research Publication, April 22, 2014, doi:10.2139/ssrn.2427932.

[28] Linton C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks* (1): 215-239, doi:10.1016/0378-8733(78)90021-7. Stanley Wasserman and Karen Faust, *Social Network Analysis: Methods and Applications* (Cambridge: Cambridge University Press, 1994).