Every Rose Has Its Thorn

Censorship and Surveillance on Social Video Platforms in China

Jeffrey Knockel, Masashi Crete-Nishihata, Jason Q. Ng Adam Senft, Jedidiah R. Crandall

Citizen Lab, Munk School of Global Affairs, University of Toronto Dept. of Computer Science, University of New Mexico USENIX FOCI 2015, August 10, Washington DC









4 most popular social video platforms in China Over 1 billion registered users 17,547 sensitive keywords

Research Problem

 What are the underlying mechanisms and motivations behind censorship / surveillance on Chinese social media?

Background

- Previous work shows inconsistencies in censorship / surveillance of Chinese social media (Crandall *et al.* 2013, MacKinnon 2008, Villeneuve 2008)
 - Suggest general directives may be given by government, but companies have flexibility in implementation

Overview

- We reverse engineer four of the most popular Social Video Platforms in China: YY, 9158, Sina Show, GuaGua
- Found client-side keyword censorship in each SVP, and keyword surveillance in YY
- We reveal dataset of 17,547 unique keywords, largest currently available to researchers!
- Previously largest dataset 4,256 unique keywords TOM-Skype + Sina UC (Knockel *et al.* 2012, Crandall *et al.* 2013)











日借

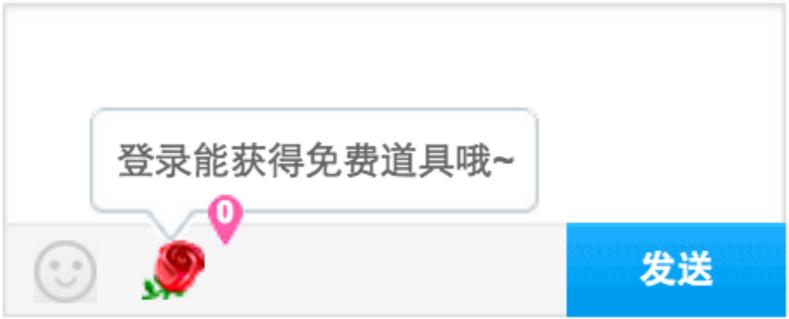
活动奖品

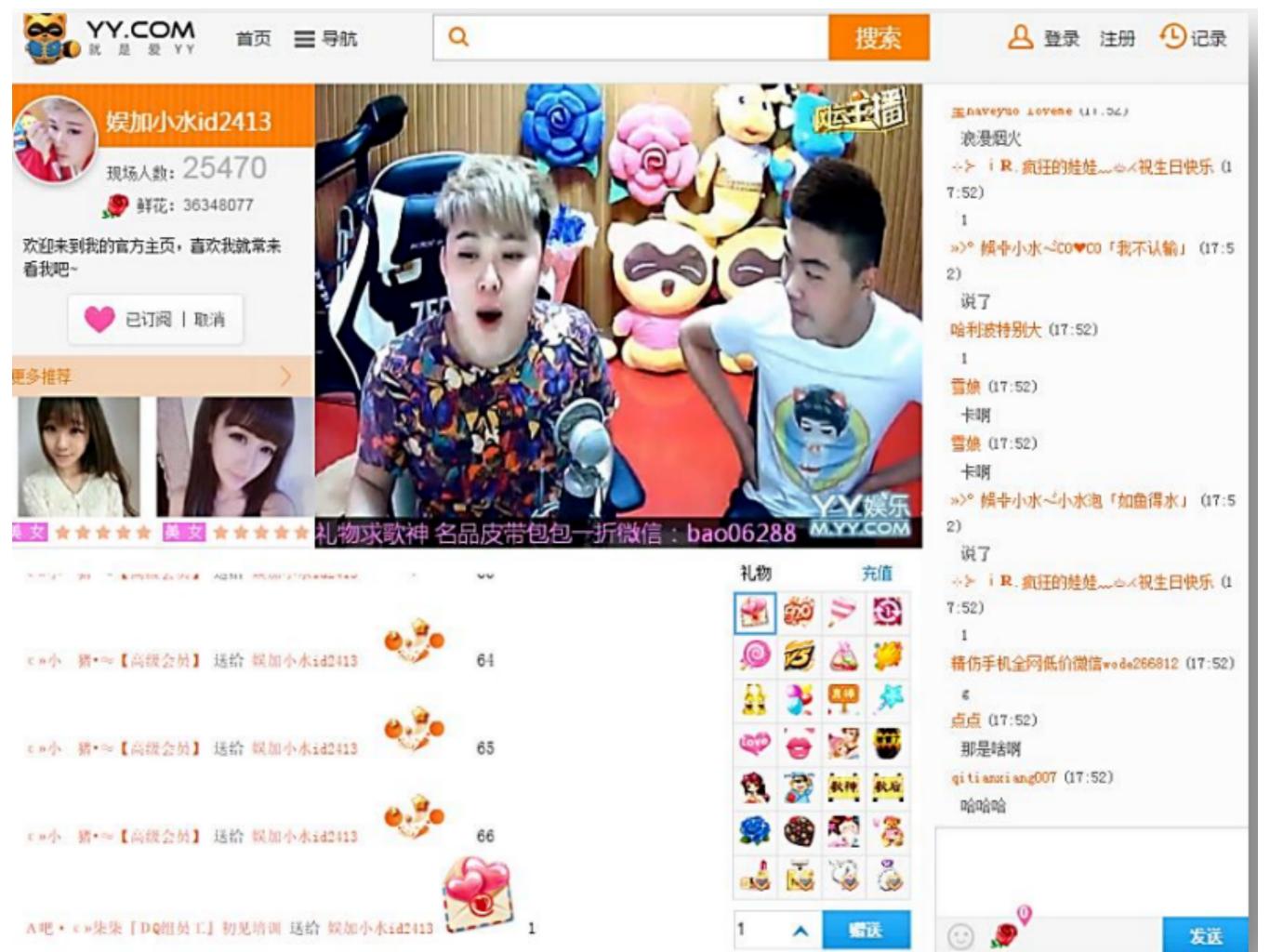
6、7短位

YY89频道开业

暨快看直播发布会









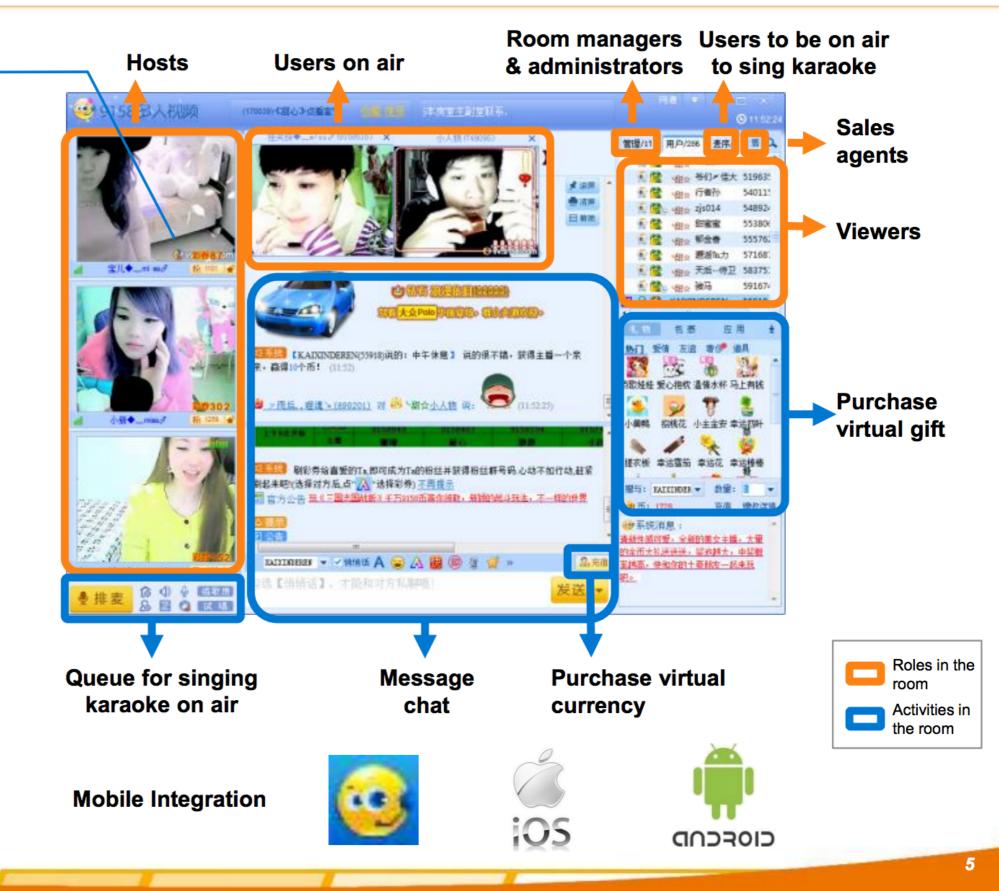
Real-time video rooms

Users sing karaoke or chat with each other in the video rooms

Users send virtual gifts to hosts and other users on air

Offer Android and IOS versions for all our major live social communities





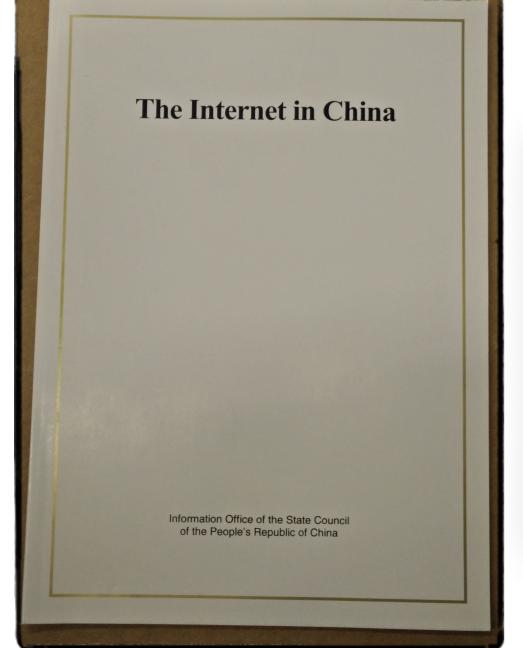
Source: TianGe Corporate Presentation

http://www.tiange.com/Upload/TiangeCorporatePPTLongVersion_en.pdf



Company	Public / Private	Product	Registered Users	Monthly Active Users
YY Inc.	Nasdaq	ΥY	861.4 mn	117.4 mn
Tian Ge	SEHK	9158, Sina Show	245 mn	14.4 mn
Jinhua Changfeng	Private	GuaGua	70 mn	N/A

Prohibited Topics on the Chinese Internet



- Endangering state security
- Divulging state secrets
- Subverting state power
- Jeopardizing state religious policy
- Spreading rumors
- Disrupting social order and stability pornography
- Gambling
- Violence
- Humiliating or slandering others

source: http://www.china.org.cn/government/whitepaper/node_7093508.htm

Sina Is Fined by Chinese Government Over 'Indecent Content'

Weibo Parent Also Says Two Internet Licenses Are Likely to Be Revoked

"We would like to... express our deep regret to the public. Sina is willing to work with other website operators to make vigorous efforts in the country's anti-porn campaign and create a clean Internet environment"

- Sina Corporation

source: http://english.cntv.cn/2014/04/27/ARTI1398608144072100.shtml

YY: "Content posted or displayed on our platform may be found objectionable by PRC regulatory authorities and may subject us to penalties and other severe consequences"

Source: http://files.shareholder.com/downloads/AMDA-NICNI/358818291x0x824505/036DB329-9796-4EED-B792-BE55BF114D54/YY-2014_Annual_Report_-_Final.pdf

Content Filtering and Management

- Strict Terms of Service for users and performers
- Violations lead to fines, account suspensions
- Content Monitoring Teams
- Automated Content Monitoring

Keyword Censorship Technical Analysis

- YY, 9158, Sina Show, GuaGua perform client-side censorship
- VV, Six Room, BoBo perform server-side censorship

 YY v7.1 downloads 3 different keyword lists: *Finance, Normal, High*

- Finance Keyword List (keywords related to phishing scams)
 - http://do.yy.duowan.com/financekwordlist
- Plain text in UTF8-encoded XML

YY安全提示:聊天中若有涉及财产的操作, 请一定要先核实好友身份,谨防受骗!

[YY Security Tip: This chat seems to involve managing assets; please be sure to verify the identity of a friend to avoid being cheated!]

- *Normal* Keyword List (sensitive keywords)
- Source: http://do.yy.duowan.com/NormalKWordlist.txt
- Base64-encoded list of UTF16-encoded keywords
- Keywords are asterisked out in the message before the message is sent

- *High* Keyword List (sensitive keywords)
- Source: http://do.yy.duowan.com/HighKWordlist.txt
- Base64-encoded list of UTF16-encoded keywords
- Keywords are never sent. If a message containing one of these keywords is somehow received, then that message will show in the chat window as a blank message.

YY Keyword Surveillance

http://sere.hiido.com/do.action? id=**<id>**&content=**<content>**

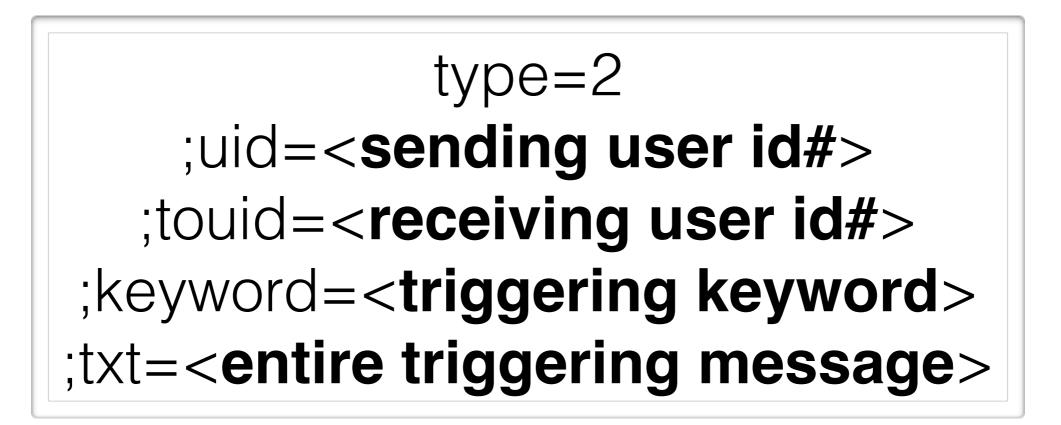
<id>is a hex encoding of a hash computed as</i>

md5(L<seconds since unix epoch>/1000_+
";username=report"+

";password=pswd@1234").

YY Keyword Surveillance

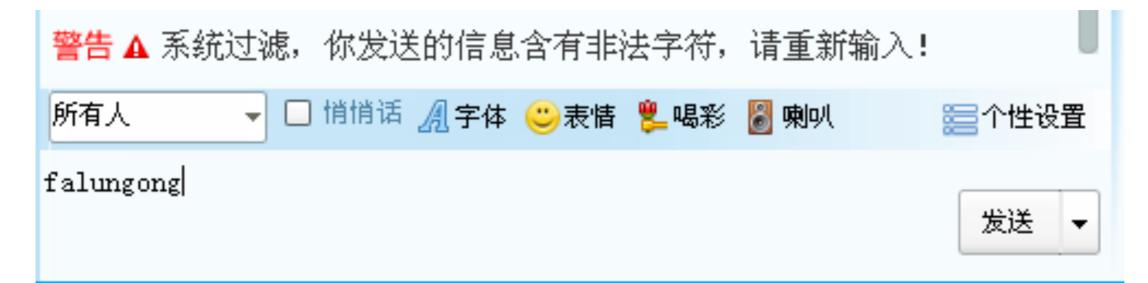
<content> is a base64 encoding of the following string:



- SinaShow v3.4 includes:
 - Built-in and pre-installed keyword lists
 - Downloaded lists

Outgoing censored message warning:

"系统过滤,你发送的信息含有非法字符,请重新输入!" [System filter: the message you sent contains illegal words; please re-enter!]



Incoming censored message warning:

"发送的消息有非法词汇,已经被自动屏蔽" [The message sent contains illegal vocabulary; it has been automatically blocked]

- Installed with binary keyword database in Word_410.ucw
- Downloads updates from http://www.51uc.com/uc_ interface/down_policy/Word_410.ucw
- Custom binary container stores GBK-encoded keywords
- Encryption: Blowfish in ECB mode with 8-byte key Dey,1b1E

- Sina Show has keywords built into many of its binary files, stored in the plain GBK-encoded
- SinaShow.exe, ChatRoom.dll, Props.dll
- UCClient.dll

- Analyzed 9158 v6.9
- Installed with two keyword lists: filnick.xml
 filter.xml
- filnick.xml replaces keywords with asterisks in usernames
- filter.xml replaces keywords with asterisks in incoming and outgoing messages. Downloads updates: http:// mimtenroom.9158.com/ web9158/filter.zip
- If chat message contains 6+ English letters all of the letters are replaced with asterisks

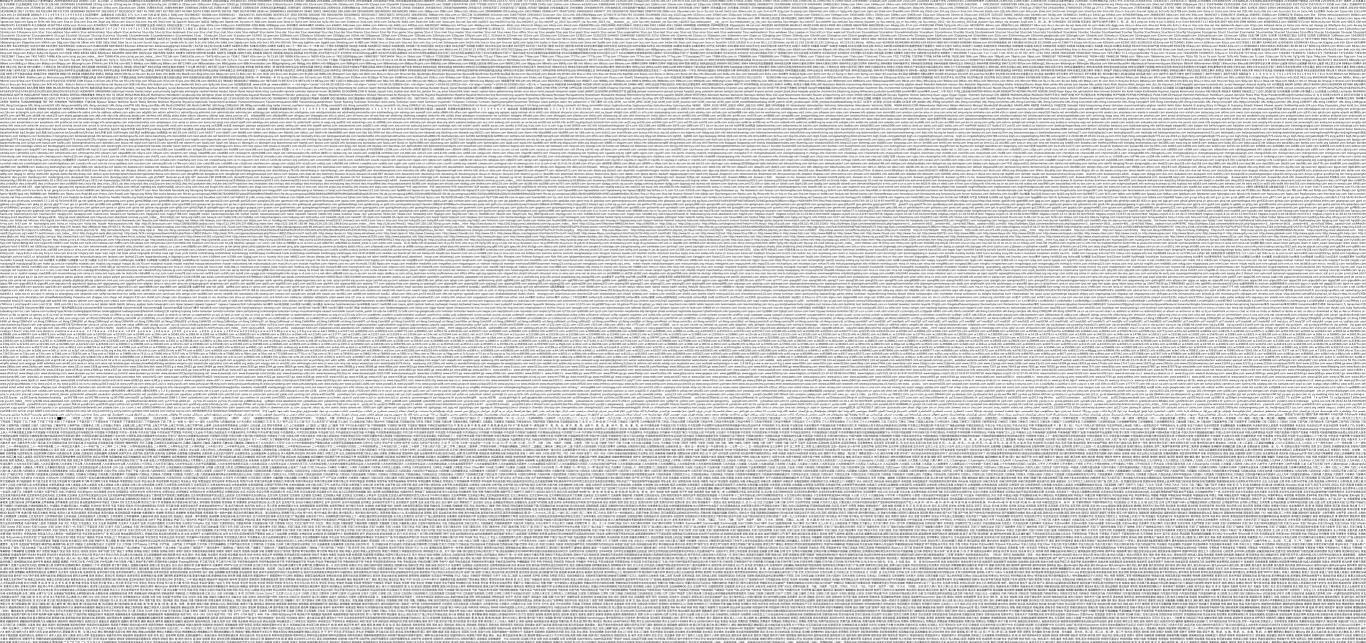
GuaGua Keyword Censorship

- Analyzed GuaGua v6.2.38
- Keywords built into RuleCenterPlug.dll
- Plain text, GBK-encoded
- Outgoing messages containing keywords are not sent, displaying warning message:

"消息发送失败,含有违法或不文明字符!"

[Failed to send message; it may contain illegal or uncivilized words!]

Our Results

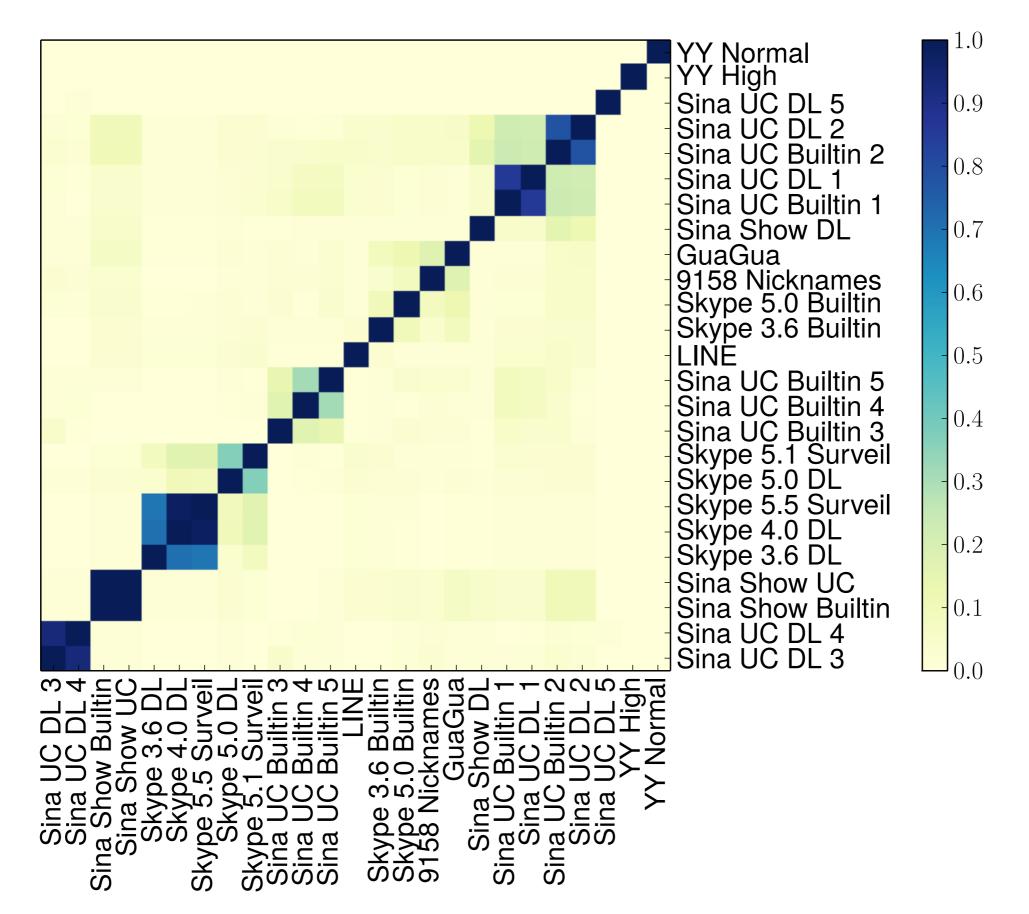


.com/ qdazam.com qdbz198.com qdbz198.com qdbet.com qdvpr.com qebriman qelbimizning qfuth qg0888.com qg2358.com qg6789.com qg6789.com qg6789.com qg6789.com qg777.com qj777.com qj686.com qg686.com qg6 lang-shop.com qiqiang.co qiqiang01.com qiqiang088.com qiqiang168.com qiqiang258.com qiqiang258.c 1 gurbon hayit gurultay guru sagchi sagci saglanglar sars sauna8.com sbmeinv.com sc.icom sc.ifeng.com/event.ng sc.ifeng.com/event.ng/Type=RawValues=&Redirect=http://www.2013viipvy.tk/vip.yv.com index . .am74.nb118.com sfte.net sgmrys sgsale8.com sgzzz.com shabakataljahad shahit shahit. shangyy.com shanhuhua.co hashow521.com sindns.info sinesun.com singapore3.com sis360.com sithoc six4 six4事件 sixfour事件 sixfour二十年历史事件 sixfour云东 sixfour讼后的中国 sixfour涵录 sixfour讼语 sixfour涵录 sixfour函读 sixfour函读 sixfour公正 six4.com singapore3.com sig code.com tyrkistan tyrkystan twayty.com twgay.com twg2pforum.com txjundao.com txjundao.com typ168.com ty21888.com ty21968.com ty21888.com ty2180.com ty21888.com ty2180.com ty21888.com ty2180.com ty218 trator ucalbum ucantivirus ucbbs ucbj ucbookmark ucceo ucchat ucclub ucclubs ucdefault ucdemo ucdoc ucdos ucdow url.cn/AibTce url.cn/B8KWae url.cn/C4QHyi url.cn/CEGqwn url.cn/Cyptw1 url.cn/D0914K url.cn/D0914K url.cn/EGN8q1 url.cn/F4y7zF url.cn/F6p51 url.cn/H8TDzr url.cn/H9TDzr url.cn/Ini57E url.cn/V149nw url.iFeng.Com/x6h8 url.iFeng. of url.ifeng.com/rR0w url.ifeng.com/clXP url.ifeng.coM/ktbT url.ifeng.coM/kbT url.ifeng.com/FBma url.ifeng.com/FBma url.ifeng.com/clXP url.ifeng.com/cWq3 url.ifeng.com/cWq3 url.ifeng.com/cQTA url.ifeng.com/FBma url.ifeng.com/clXP line uygurcha uygurcongress uygurie uygursitan uyhurkurultay uyifyf uylla v.6.cn v.9158.com/150089 v.cn v.com v221.com v223.net v3dv.com v40688. va.cn va.com va.net va harakat va sunnat vafo gilganlar vafo gilganlar vafo gilganlar vafo gilganlar vafo v221.com v221.com v223.net v3dv.com v693.c v6n v775.com v977.com v977.com v977.com v40688. va.cn va.com va.net va harakat va sunnat vafo gilganlar vafo gilganlar vafo gilganlar vafo gilganlar vafo gilganlar vafo v221.com v221.com v221.com v223.net v3dv.com v693.c v6n v775.com v977.com v40688. va.cn va.com va.net va harakat va sunnat vafo gilganlar gr.com vipdyyn.oogr.com/vip.yy.com_index_..html vipgg.com.am52.nb118. viphd1080.com vipyy82-2013-com vipyy82-2013-com vipyy.vip1.zmtu.com vipyy.vip1.zmtu.com vipyy.vip1.zmtu.com vipyy.vip1.zmtu.com vipyy92-2013-com vipyy88.w616.en. vipy m w.000773.com w.008528887.com w.0011.tw w.001dcdv.com w.003377.com w.003697.com w.0085288.com w.0085288.com w.00852885.com w.0085288.com w.00852888.com w.0085288.com w.00852888.com w.00852888.com w.00852888.com w.0085288.com w.0085288.com w.0085288.com w.00852888.com w.00852888.com w.00852888.com w.0085288.com 112529.in w.113158.com w.114x.net w.115544.com w.1151.uet w.117.104 w.117700.com w.112399.com w.123396.com w.1234000.info w.123402.com w.123402.com w.123989.com w.123989.com w.123556.in w.128861.com w.121.00 w.12122.com w.131520.com w.131520.com w.131520.com w.1234000.info w.123402.com w.12399.com w.12399 w.2558i.com w.256347.com w.258988.com w.26388.in w.266278.com w.266933.com w.266933.com w.279280.com w.271181.in w.271188.com w.271181.in w.271188.com w.272188.com w.278858.com w.278858.com w.278858.com w.279289.com w.279289.com w.279289.com w.28165.in w.266933.com w.271181.in w.271188.com w.271181.com w.275888.com w.271181.in w.279289.com w.271181.com w.279289.com w.271181.in w.271181.com w. w.3456799.com w.3474.com w.34c.cc w.353518.com w.355855.in w.35766.info w.358358.com w.3597697.com w.363385.net w.36090.com w.363385.net w.36090.com w.366668.com w.366668.com w.366668.com w.369.info w.369.info w.359855.in w.35786.com w.37092.com w.37092.com w.363385.net w.36090.info w.369.info w.36090.info w.36090.info w.36090.info w.359855.in w.35786.info w.359855.in w.35986.com w.36090.info fo w.50dy.in w.511.com.tw w.511573.com w.515635.com w.5160.info w.517379.com w.5186678.com w.5199.vs w.51bf.net w.53448.com w.53591.org n.com w.622268.com w.62229.info w.6266.info w.6266.info w.638638.com w.6636.com w.6636.com w.6636.com w.6636.com w.661698.com w.6636.com w.6636.com w.66223.com w.65257.info w.65257.info w.65257.info w.65257.info w.65257.info w.65257.info w.65265.net w.660222.com w.661698.com w.661698.com w.661698.com w.661698.com w.661698.com w.661698.com w.66223.com w.65257.info w.65257.info w.65257.info w.65257.info w.65265.net w.660222.com w.661698.com w.661698.com w.661698.com w.661698.com w.65257.info w.65257.info w.65257.info w.65257.info w.65265.net w.660222.com w.661698.com w.661698.com w.661698.com w.661698.com w.661698.com w.65257.info w.652 com w.7892.in w.79eu.com w.79kx.com w.79dg.com w.711.eu w.7p.org.tw w.7xo.info w.8004008.com w.8004008.com w.8100.cc w.813831.com w.8269.com w.83083.com w.830855.com w.838555.com w.84848.info w.850989.com w.850989.com w.850989.net w.8517.org w.91900.com w.91919.info w.9119.info w.93343.com w.939666.com w.946.cc w.94961.com w.9551.com w.9950.com w.990988.net w.99113.in w.991255.com w.990988.net w.99113.in w.991255.com w.996b.com w.996b.com w.996b.com w.990988.net in w.9950.com w.990958.com w.990958.com w.990958.com w.990988.net w.99113.in w.991255.com w.99113.in w.991255.com w.996b.com w.990988.net w.99113.in w.991255.com w.996b.com w.990988.net w.99113.in w.991255.com w.996b.com w.996 :n wa.com wa.net wadjwq.com waga7.com waga8.com waghnk wagr waikeung.info waitao2011.com waiwai.d.v.gp waiwaiid.cn.mu waiwaiid.w.gp watendaxlar waten watendaxlar waten watendaxlar waten ajia.6579.idcice. woaiyyvip.s18.23dns.net wokw.cn wonderfulmature.com wowjapan.tw wowprogress.com woz5.com wsina.net wtbhrny.512sex.tk wthmix.net wtmlei wtwlg.com wuaijiuba wuc.com wufeip6 wuhanairgun.in ww .3 9 58 8.g a ww w ww w.uc4 8.g a www www.000rr.com www.001ii.com w.com www.4453.ga www.56354.ga www.556354.ga www.556com/av/ www.556354.ga www.55610.com www.522av.com www.556354.ga www 132.ga www.a662.ga www.a663.ga www.a663.ga www.a664.ga www.a666.ga www.aeaolu2.com www.duowanyyy.tk www.dwz.cn/dKHvJ www.ef575.ga www.ef268.com www.fg556.t www.fg556.t www.fg5544.ga.ga www.ff1ff.com www.gg8888.ga www.ef268.com www.gg8888.ga www.ef268.com www.gg8888.ga www.ef268.com www.fg5544.ga.ga www.ff1ff.com www.fg554.ga www.ef268.com www.fg554.ga www.ff1ff.com www.fg554.ga www.ef268.com www.fg554.ga www.ef268.com www.fg554.ga www.ef268.com www.gg8888.ga www.gg888.ga ww vww.m338.ga www.m339.ga www.m341.ga www.manmanshe.com www.mayalike.com:8000 www.menglei.cc/vip.yy.com www.mpidao.cn/pert/ www.ntdtv.com www.pi39i.com www.pi39i.com www.ppp36.net www.ppp36.net www.ppp36.net www.npigia.cn/pert/ www.ntdtv.com www.pi39i.com www.pi39i.com www.ppp36.net www.ppp36.net www.npgia.cn/pert/ www.ntdtv.com www.nortdtv.com www.pi39i.com www.pi39i.com www.ppp36.net www.npgia.cn/pert/ www.ntdtv.com www.pol30.net www.nortdtv.com www.nortdtv.com www.ntdtv.com www.pi39i.com www.pi39i.com www.nortdtv.com www.ntdtv.com www.ntdtv.com www.ntdtv.com www.pi39i.com www.pi39i.com www.ntdtv.com www.ntdtv.com www.ntdtv.com www.ntdtv.com www.pi39i.com www.ntdtv.com www.ntdtv.com www.ntdtv.com www.pi30i.com www.ntdtv.com www.pi30i.com www.ntdtv.com www.ntdtv.com www.ntdtv.com www.ntdtv.com www.pi30i.com www.pi30i.com www.ntdtv.com www.ntdtv.com www.ntdtv.com www.ntdtv.com www.pi30i.com www.ntdtv.com www.pi30i.com www.ntdtv.com www.ntdtv.co w.vipyy2011.com www.wse55.org www.wai-wai.co.cc www.waiwaiid.ch.mu www.waiwaiid.ch.mu www.waiwaiit.com/id www.waiwaiit.com/id www.ws5959.ga www.y445.ga x81.tk www.yyvip87.rr.nu www.yyvip1.ss06.com www.zyvipy.tk www.yyvipy.2013.tk www.zyvipy.2013.tk www.zyvip.cr.cx www.zht520.com www.zkt520.com ww com xina.net xincan.net xindacz.com xindns5.info xing-ai.us xing8ke xinlangshow.com xinlangzhigujzu.com xinlangzijingujzu.com xintongqq.com xinxin2013.com.am155.x59.cn xipahana xirin bulak xitay xiu.56.com xiuli.us xiuwu.name xjbenma.com xjfw.info xl51uc.xl51uc.com xlohd xlzaa xm.cn xn cn va.com ya.net ya31.com ya7.com.tw yadi.sk yagak yahxi nersiler yamgur yagmak yan-er.com/id yanwu.namE yanxiubbs.com yaxison yaxis plao.kfkf.net yueplao.nznz.net yufan.us yukumluk zukam yumingk.cn yun-yy.5c5c.info yunhyq.com yunnandiki mubarek jihadimiz yunshiyingyuan.com yusn.info yuwei46.svfree. yuzu123.com yvcfhnpbfde.com yvduowan.com yvipe.com.am154. yvvip5.cloud60-140.78host.com yvvjqmv yv ngpz.tk yy.t.bscgcxva.tk yy.vip.loufeng99. yy.vip.x8198. yy.yip.gtrh4gh.tk yy0.googlecode.com/svn/hp1/cmp.swf yy10vipus108.eoidc.net yy10vipus12aoidc.tk yy10vipus230aoidc. yy10vipus230aoidc.cn.ms yy10vipus236aoidc.tk yy10vipus236aoidc. yy10vipus236aoidc.tk yybk08.com yybk2008 yybk2009.com yybw2010.com yycjvip.ooqr.com yycvipn.ooqr. yydianying.6hkidc.com/dy yydongfeng.com.am172.x59.cn yydvowan yydw2010.com yyfdg7.tk yyfens.com yyguanvip.la2.gongyinmy. yygvipf.com.am170. yyhaohao.com.am172 yyhd.q.qiegaowang yyhdc 1588.com yyvip111.w616. yyvip19.one2.abestland.com yyvip2013.kr47. yyvip2013.kr47. yyvip2013.kr47. yyvip2013.kr47. yyvip2013.ao.free.afxq.info/yy.com yyvip5.c.qiegaowang. yyvip5.c.qiegaowang. yyvip5.c.qiegaowang. yyvip2013.kr47. yyvip201 اللها بولىدا جنهاد قىلۋاتغان مۇسۇلمانلار ۋە تۈركىستان خەلقىغە ئاللە بولىدا ھەركىتىنى ئىنلام قدارسىز شەرق تۇركىستان بىرلىشىپ ئىلىپ بارغان جىھاد ھەركىتىنى بايلاۋا ھەدلىرىمىز شەرقى تۇركىستان بىرلىشىپ ئىلىپ بارغان جىھاد ھەركىتىنىڭ يەرقى بىز ۋە گۈزەل خوراسان تىررورلۇق سىن ئىسلام قەلبىمىزنىڭ تۆرىدىن ئورۇن ئالسۇن ئەبۇ مۇھەممەد تۇركىستان بىرلىشىپ ئىلىپ بايلاۋا ھەدلىرىمىز شەرقى تۇركىستان بىرلىشىپ ئىلىپ بايغان جىھاد ھەركىتىنى بايلاۋا ھەدلىرىمىز شەرقى تۇركىستان بىرلىشىپ ئىلىپ ئورۇن ئالسۇن ئەبۇ مۇھەممەد تۇركىستان بىرلىشىپ ئىلىپ ئىلىپ بايغان جىھاد ھەركىتىنى بايلاۋا ھەدلىرىمىز شەرقى تۇرمىزى ئالىلاۋا ھەدلىرىمىز شەرقى تۇركىستان بىرلىشىپ ئىلىپ ئىلىپ ئىلىپ ئىلىپ ئىلىپ ئىلىپ ئورۇن ئالسۇن ئەبۇرىيانلىقا دىنى ئەبۇرىيە مۇسۇلمانلىرىغا نەسەئەت تۈركىيىدىن ، پارتىيەلەرگە بايانات شەرقى تۈركىستاندا ئېلىپ بىرىلغان جىھادىن شەكلەنمە مۇچاھىد ئەمىنلەرنىڭ ماياھىتى مەندىن يەتكۈزۈڭ ئى ئاتا؛ نۆرىددىن رخمە ئاللە نىڭ 6进【→【入】【手枪出售】联系电话【正悟网】一个目击者自匀广场日12 一个解放军的1989 一中一台 一久八久 一九8906零四 一九8玖年 一九八九年件 一九八九事件 一九八九六四 一九八九年六月四号 一九八九年六月四号 一九八九年六月四号 一九八九年六月四日 一九八九年30 年八九年30 年八九年30 年代 和极乐城 万宝路极乐城 万庆良 万碗东 万间南 万盛出事 万盛发生暴乱 万盛和綦江 万盛暴动 万盛雄起 万维读者论坛 万能开锁器 万豪级乐城 万达国际极乐城 三 去 车 仑 三·秒·必-射 三个代表 三优叔乐 三键弓弩 三分钟教你原子弹 三利达 三利达古爾 三利达弓弩 三利达弓弩军刀 三利达弓弩军刀 三利达弓弩直营 三利达弓弩首营 三利达弓弩有三 三利达弓弩有三 三利达弓弩和件 三利达追月 三十八军军长 三十八军军长 经勤务 点财女"妹子陈 下流财女"黄丽清 下硝化甘油的制作方法 不 黄 死 全家 不 黄 死 全家 不 黄 死 全家 不。黄 死。如 不 黄,死。参,死,黄,死。黄,死,黄,死,姜,死,姜,死,姜,死,姜,死,姜,死,姜,死,黄,死 金家 不。黄,死。专,来 不。黄,死。专,死,黄,死。参,死,封 不の黄死全家 不准提及六四 不凡军品 不凡军品 不凡军品 网络伊尔阿存异积极参与到公民运动中来 不判你死刑 不去 家乐福 不去家乐福 不反对通知书 不夜城银乐城 不容敷法》 1网 世界日报报道当时国家主席杨尚昆在接见海外华人说有六百多人死亡 世界经济导报 世纪打猎论坛 世间有净土*万吉天门今已开 丙二酸铊 丙烯醛 丙腈 业主信息 丛林开山刀 丛林直刀 东京热 东北独立 东南西北论谈 东土耳其斯坦 东方夏威夷纵乐域 东方时空 东方将猎论坛 东方星朝娱乐域 东方红时空 东社 东突 东突伊斯兰觉 东突厥系战 东突厥斯坦 搞强迫遗忘 中共江源 #當 中国封锁Twitter 中国平息反革命暴乱 中国开始乱了! 中国弓著专卖网 中国弓著府销网 中国弓著两 中国弓著两 中国改编也只是为了自己的利益 中国政府使用军警对学生运动进行武力镇压 中国政府或债的手段 中国政府只会下限价令 中国政治风波 中国改良发 中国有机构没头脑 中国毛派 中国民主党 ¹总书记赵紫韶平反 为建立民主反对魔教而进行和平请愿的学生和市民 为无辜逆去的童鞋默哀 为正道献身 为赵紫韶先生平反 为镇压评功探好也不行 主啊·衡造我 主啊愿你拉着我们的手 主席 主席终身制 主播袜子好骚 主是否也在你心歌曲中 主是我心中的歌 主是我最亲爱的伴侣 主爱万国也爱这里 主爱中华级的家 主耶稣得胜复活了 主耶稣亲远院伴我 兩人上门腦务 兩個国母 兩星邮轮操乐域 二千多人死于六四 二奶 二手弩 二氯化汞 二氯苯胂 二甲基安非他明 二甲基硫代磺酸氯 二的8次方 二盐酸盐 二碘化汞 二磷化三锌 二通 于是军人们扣动扳机 于治成六四事件 云南情蛊 云南的伟大圣战 云顶国际规乐域 云鼎银乐域 五一 家乐 五一 家乐福 五一不去家乐福 五一二七零一三 -二 五一拒绝去家乐福 五三五事件 五不 五二开席 五二开西 五人被坦克压死 五加 紀念日 今天来个APEC 从六四到法轮 从反右到六四 从天安门事件到08完章 从央视春节晚会衰落 从总书记失恋讲起 从祖国内心深处迎来的一篇大悲文 从赵紫阳看中共与良心 从逆向思维出发 他们才能躺在血锈的稳定大床上 他们输入的程序悬绝食 他妈 他母亲 付申奇 仙童下地狱 仙鹤气枪出售 代付单 代层冒身份证 代制身份语 代办 假户口 代办 假证 代办信用卡 代办全国毕业学历 四纪念活动 全球纪念六四25周年网络大会 全球网络电视棒 全球退览大潮 全球逐毛曰 全球通用户资料 全能神教 全金属仿真枪专卖 全面公布六四真相 全魔法伦 八*九*学*潮 八95hijian 八9事件 八9六4 八9动乱 八9十年代之交风波 八9学潮 八9年春夏之交 八9运动 八9风波 八jiu 八jiushijian 八jiu告端 八jiu民波 八×八事件 八七大游行几万人开进天安 法施记录片 八九年的举生运动 八九年的政治风波 八九年由操植医事件 八九年由操植压穴因事件 八九年由解植压穴因天安门 八九年由解植压穴月四日 八九年赴京执行或严任务 八九年驱逐政治风波 八九年阳学潮 八九年阳费民运 八九年陆县学潮 八九年由党组支和机枪 八九年民波 八九降发 八九牌数 八九建入设 (九)即第 八九星乱 八九星乱 八九星动 八 . 直傳集 六4榮运 六4榮运时美学生 六4环会 六4樓塞 六4團團 六6郎子 本four 本four事件 六he影 六si審件 六si屬美 六si審件 六字 勃克 六一掌 劳教所 六一掌 消息 六一掌 派出所"非法 六经事件 六经动乱 六、四学生爱国 六二總會官言 六合回影开掌 六合影 六合影公司 六合影内参 六合影拆器 六合影拆起 六合影拖耦用库 六合提 六合变机 六合渠 大和影 六和谐四 六和谐四 本种 计和型 |后邓小平奋起反左 六四和解 六四和解委员会 六四和解条件 六四四君子 四烛光悼念 六四烛光驰会 六四烛光纪念 六四烛光集会 六四照片 六四兩家武文建 六四白皮书 六四的20周年祭 六四的内墓 六四的墨客者 六四的四号是反官倒 六四的地支 六四的秦训是中国需要政治改革 六四的死难者 六四的死走者 六四的校光映会 六四的长光映会 六四的解散 立四方相与和解建议书 六四真相与和解建议书 六四真相与和解建议书 六四等近代史实 六四纪实 送用尔和平准 刘晓波同忆录 刘晓波时教胡晓浩 刘晓波 拉自四号 刘月四号 刘月四号 刘月四号 刘月四号 刘月四号 刘贞宗 刘辉波 刘叔章 刘辉 告令国人民书 刘爵 "中臣海捕龟 刘爵"于打"没宗赐了口白开水 刘辉雄 刘金宝 刘爵请扶誉 刘贵 经认为 化从上传法之路 药间烟烟 删除的面配的子 利泰提乐城 利来国际提乐域 利康提乐域 利康提乐域 身寸 包依射 包养 包袋射 包子斋 包射 包伤清意 包皮 包邮 包 - 射 匕首蝴蝶甩刀专卖 化学冰牙等 北京之春 北京198964坦克 北京64学生事件图片 (京进入或严的第四天百万人民今天下午举行示威游行 北京风声朝寝 北京风波水事记 北京风波的前前后后 北大三角地论坛 北大生穿64入校 北大等教职员挂议买了栏尾楼 北妹皇后 北方先锋军刀 北美新闻 穿破封锁 大陆新闻 北美自由论坛 北韩 医腔代指导 医腔代接号 医腔代接号 医腔化接号 医腔侧 建氯基 十九年前学期 十二名中共数坛老人呼吁为胡耀邦彻底平反 十五万群众举行烛光晚会 十五名 F台 反射微技术 反恐略 反攻 反攻大陆 反日 原日 鞭棄 反日 警察给打了 反日游行 反社会 反腐败反官倒运动 反腐败论坛 反迫害"425回顾 反道 反革命 反革命都住在中南海 发中发娘乐城 发伦 发射药 发愣 发抡 发正会 发漂代开OO 发票 发票代开OO 发票出售 发论 发轮 取消服务于一类统治 变态 叙利亚圣战 口--交 口~交 口交 口交阻务 口暴 口爆 古可叶 吉林帮 吉林省军区通讯站*陈躔 同维存连 正法极终 同城梅妻 同安 同走 同走MB上门服务 同性 同性恋 局性 同性恋 局性态服务无函由影下载 同苏雅提乐城 后六四文化祭 后勤新政委 后喜苏 后姆叔时代纳公民美德与公民责任 时鲁希鲁幸込圣战 向使惊速惊掉狠杂物 向六四事件中生去寝子 向六四请维者走支 向前军委主席邓小平谨报北支发生了 ·没期功陆建 吉底情 赴女 赴女屋号提乐城 赴女揽 复杂脱颖的家 处交与方脑 外交论坛 外地演来的家队在长安大街对峙 外提提六四问题 外提提六四问题 中提新教,王确山 大花逼 大赢家真人娱乐 大逆转中国的私有化 大道 大都会级乐城 大陆七位学人 大陆国家党 大陆媒体被禁声 大陆当局 大陆民运 大马士革刀 大麻烟 大黑鹰弓弩 大(乳) 大」B 天AN门果彼 天A门事件20周年 天A门事件20周年 天XX事件 天anmen事件 天anmen事集 天an门1989 天an门1989 天an门1事件 天an门自提清场内幕 天a门事件 天x门事件 好刀网 好葉坞袍照门种子 好货 如何制稿化甘油 如何自制燃烧弹 如意搜索"世界之门 如意搜索"世界之门 如意搜索"世界之门 如是接面了停止终者政府不会因此把对话的门关起来 如果叶剑笨不死在64以前 如果新闻可以换来自由 妇女和子女的判决 钨 结构的复数故事 與提 報道 枝 枝女 核核可得加可上門 妻子 東奥 绝记娱乐城 差离云 美洪祥 ;大的成功 巨星線乐域 巨額军費 己巳国務 巴克刀 巴克折刀 巴格达迪 巴比伦提乐域 巴登提乐域 巴赫 巴雷特 市府上访 市府门口静坐 市政府 市民挡住锅甲坦克 市民的栖种为代价的运动 布亚镇"圣战"运动 布加迪提乐域 倚父 希尔德提乐域 倚乎尔吾 希望之声"报导"纪元 希望之声"日报导 希望之声"日报导 希望之声国际广播"采访报道 希望之声广播电台 希望你们早些结束绝會谢谢同学们 帝国TW '大上访 废情娱乐域 废除中共一觉专制 康硕傅VS天线宝 康硕傅大战天线宝 康莱德提乐域 廉政大论坛 原尔跳尾刀 原尔跳弯刀 塵場龙 建卤史 律成一个自由尼主宗政的国家实现国人百余年 建联通 建行 廿年前 开启六四事件的真相与和解行动 开坦克直提镇压 开展纪念六四活动 开山刀砍刀出传 开放报禁 开放杂志 锤钟鼓 弹药制造方法 说。安干 调纤 强量 强行讲험北京和占强天安门户场 强迫条国家机关对1989年事件所做出的条酒结谋决议和文件 当易哭望不得 当局的理念 当年64续會 当年学生反腐败要民主的自语 当时北京市民的政治坂求县要求召开结别人代命撤消戒严命 当残共平九 形而上学的课家 庞永致承 彭爾慶 彭賺 屏蔽 彭賺 近平 彭賺右中国被屠蔽 彭惟右 成人视频 成人视频 成人视频下载 成人點图区 成人高清电影 成入电影 我为中国命运一哭 我以我自荐轩辕 我们告诉未来"从常人到修炼人 我们告诉出来。"从常人到修炼人 我们更圣战者 我们的家园呼罗珊 我们的家园呼罗珊 我们的家园呼罗珊 我们的家园呼罗珊 我们的家园呼罗珊 我们的家园呼罗珊 我们的家园呼罗珊 我们要心实 我们要心实 我们要吃饭 我们要工作 我们重返上访村的计划彻场 我做陪妻 我儿开西 我觉不是做不到 我把干你 我想爱爱 1計 年龄 年龄团未龄 年龄式折叠三田楼 年龄气* 龄语 年龄见而夺息 年級到後後進引傷器 年始定定自制年龄 年機截動器 年资 年物 年物专志 年物写表 年物定意 年物定意 手物定意 年物学者 手指告责 年常当面夺员 年常者你付给 林古传器 林古不得乐妓 打 飞 机 打做 打做中国改革开始的总设计师 打做封课 打做到课具产学 打倒卖罪 打印約減 打物 打拍己發三先倒 打招奏田楼 非 军队进城来镇压学生运动 政府害怕了怕人民的儿女怕自己的公民 政府门前 政府门口集合 政治反对派 政治局常委 政治清洗 政治职 激活碳 激活碘 激系磷 激音虫 敏感词 诺贝尔奖 救300民运人士 教你制作庆子准 教你制作大威力炸弹 教你怎么用土办法做武器 教养院 教养院 教养院 教养院 教养院 教养院 教养院 教养院 教育院 教育院 教育院 教育院 教育 医白豆属 化乙酸乙酸化乙酸乙酸化乙酸乙酸化乙酸医乙酸化乙酸 线乐域 新闻时間 新闻管制 新羅視角"文艺線乐"争鸣崩結 方li之 方動之 方號授导"李熙报导 方舟子 方觉 无二氮西 无人坐的凳 无声描 无根因 无根因 无根因 无相因 无常国 无常正 无信的军车编码我们的新体它却唤醒了亿万人 无政府主义者會讲 无数二次郎 无水胆 无界浏览器 无码 无码成人影腔 无线彻听器OO 无辜汉族 无限制 既然是风波何以要动用数十万军队 日你 日内瓦金融 日本使佗 株花色影院 塗 梁国雄 梁擎墩 梁源小区发生砍人事件 棋牌真钱游戏 森林之琅弓号 標題 植物冰 楼风信息 楼风性息 模B 橡胶子弹 欢乐博银乐城 欧美情色 欧美成人 欺骗公共舆论和国家权力机关的谎言 歌功颂德 止喷水 正品军刺出售 正品电视棒 OO 正法 正见网"中共 正规医院奕贤 正规真票 此木王令 此木玲 步枪专卖 步槍专卖 武警"大纪元 武警作战服 歪 提過务 留云山 留在天安门广场上学生们可能会流血 留小波 留疑波 畜生 番木響碱 磷酸 溴淤的人民领袖2005年1月17日 盛假宝 靈服魚 靈腐乳 靈眉的 靈家兒 靈服兒 靈家兒 夏夏前前的天 受緊 白净 白小姐特码 白萝 白牙卷 白皮带 白皮带 白皮素 白碇 白立朴 白粉 白粉抹筋玩 白粉技伤 白粉黄彩 白膀胱筋 白粉肤紫斑 白斑国际级乐城 白斑国际级乐城 白斑国际级乐城 白斑国际级乐城 拉的群众而玻黛不前的形兵 直刀匕首直锁网 直接看成人站 和思红镜师 看·萧·典.影 看·唐·典.影 看中国"中谢玉"张雷颖 看中国""记者奏样 看中国记者 采访报导 看到新疆人就打 看守府"亮人"称劳教所 看片 看拢友来 看远 直人二八红 直人继续 直头眼镜网站 直義 回归的披程 直義忍 直義忍集層 直案作券场景 直索作为场子 直索是办学历 直案 神欲相约艺术团 每耳太根油 译体 整片下载 整片的姿源 整片姿源 整香 缝完第三十二代传人 缝罩干去 導风 福寿賽 城布斯提乐城 金洁成 冬蕈汽栓 秀 秀.D.对 秀思 秀b 秀b 等人的多级 4.83869 秀.B 女 0.6034.83869 秀.B 女 0.6034.83869 秀.B 女 0.6034.83869 秀.B 女 0.6034.83869 秀.B 女 0.8034.000 梁鹰领艇 冬鹰变体 冬鹰领船 冬鹰领船 冬鹰领船 冬鹰领船 冬鹰领船 紫癜 化乙酸乙酸 化乙酸乙酸

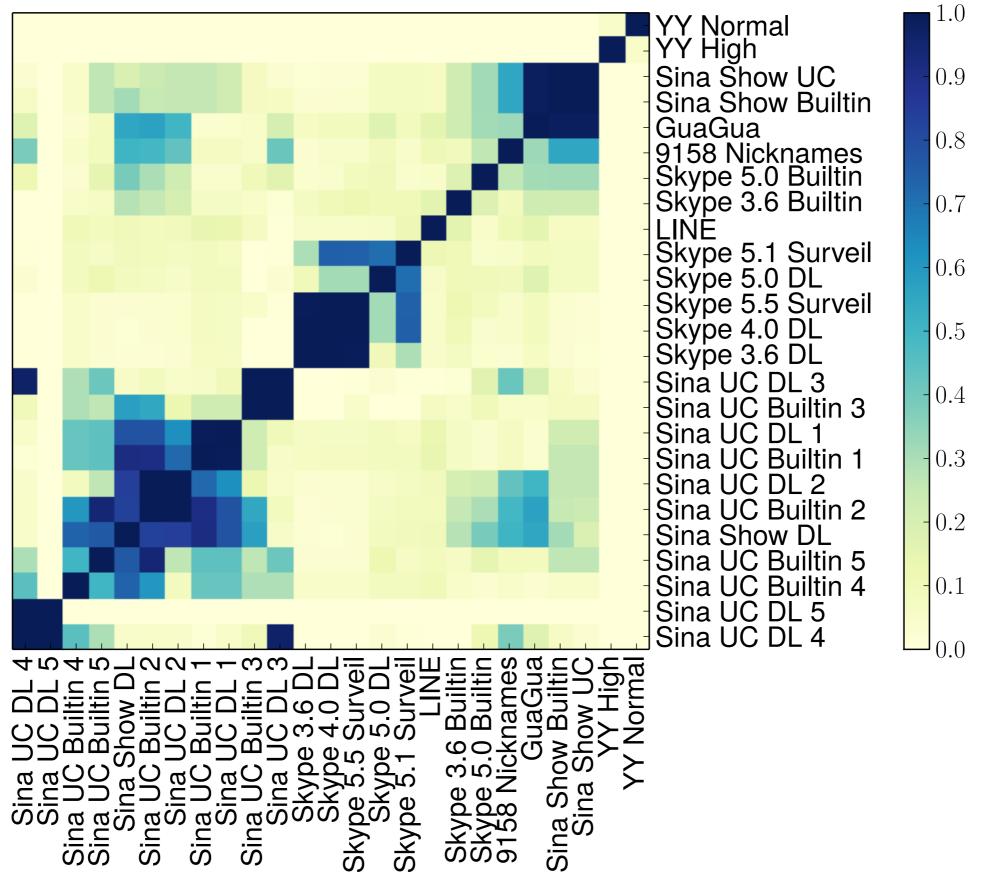
Keyword List Size (May 17 2015)

List	Keywords	Unique	
YY Finance	48	18	
YY Normal	20	20	
YY High	13,482	13,242	
9158 Nick	65	59	
9158 Chat	318	318	
Sina Show SinaShow.exe	1,224	910	
Sina Show UCClient.dll	910	910	
Sina Show Downloaded	3,711	3,206	
GuaGua	58	58	

Keyword lists clustered by Jaccard Similarity

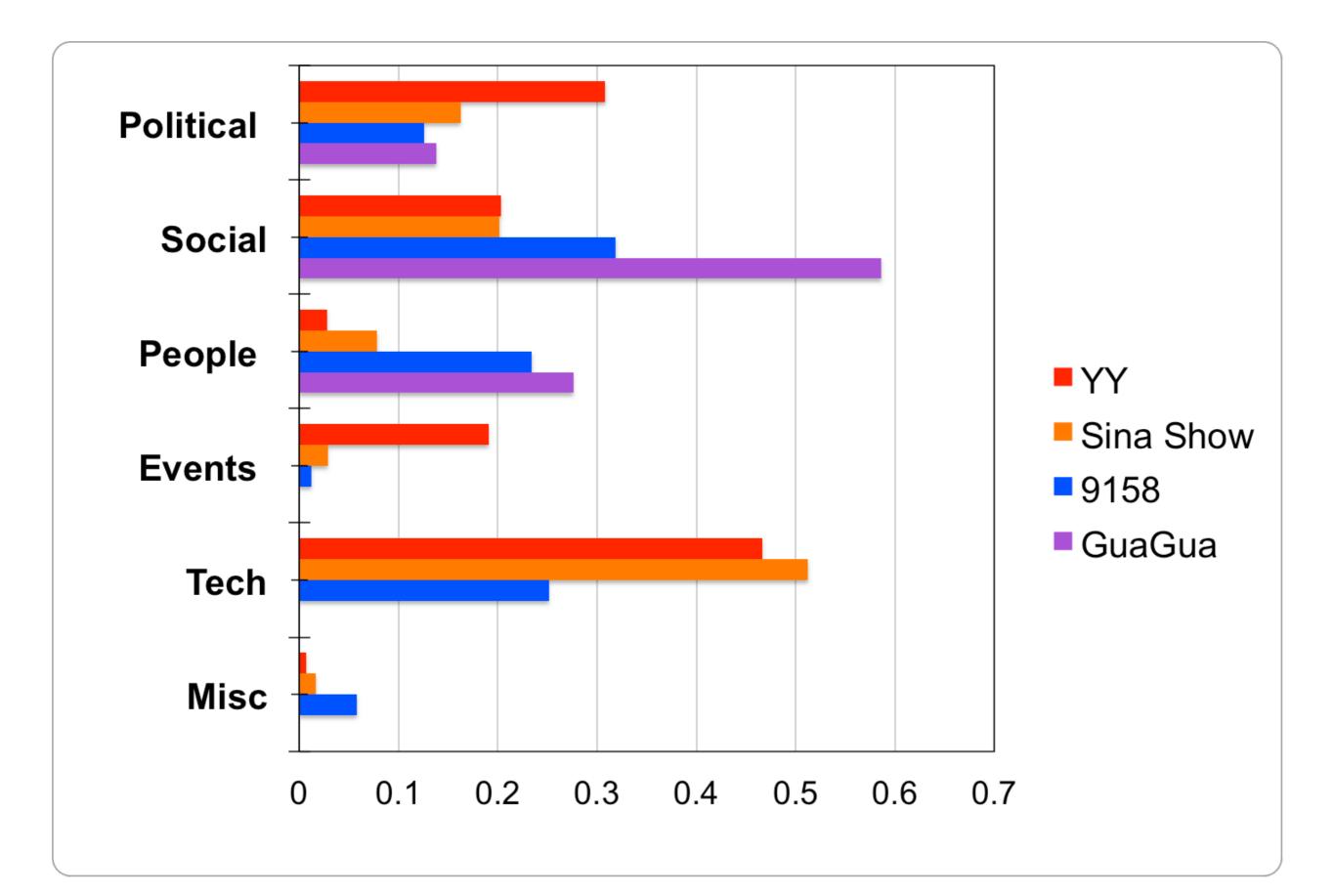


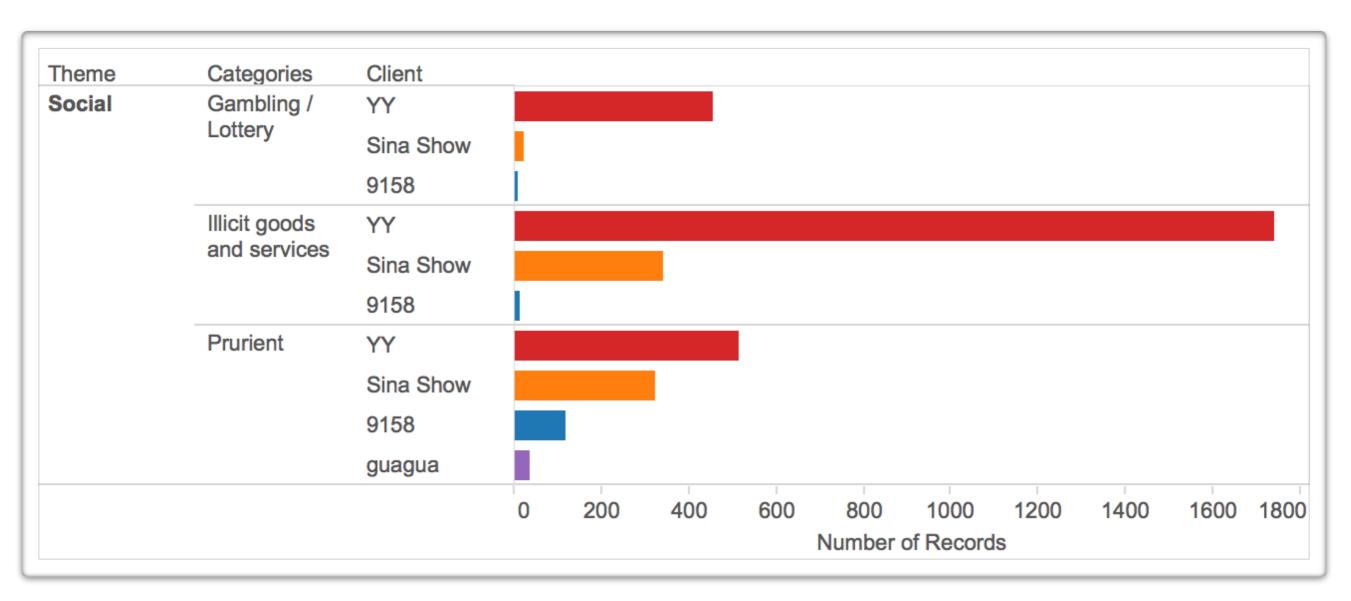
Keyword lists clustered by similarity(x, y) = max(% of x in y, % of y in x)



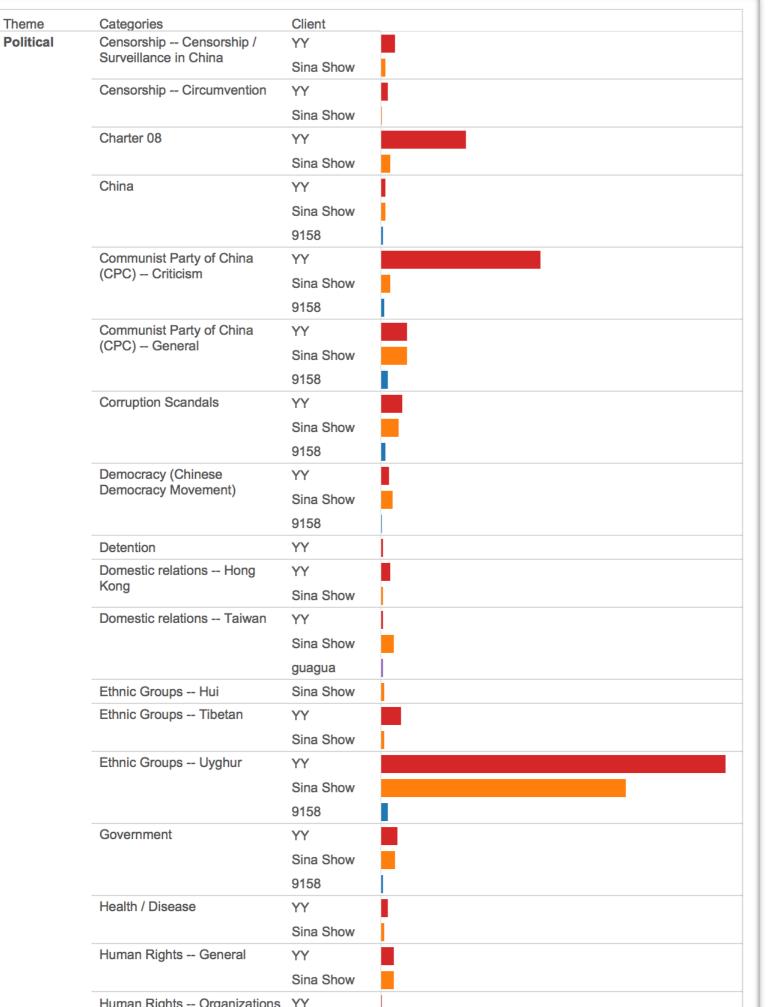
Content Analysis

- Keywords human / machine translated, contextual analysis
- Grouped into 80 content categories under six themes based on code book developed by Crandall et al. (2013)





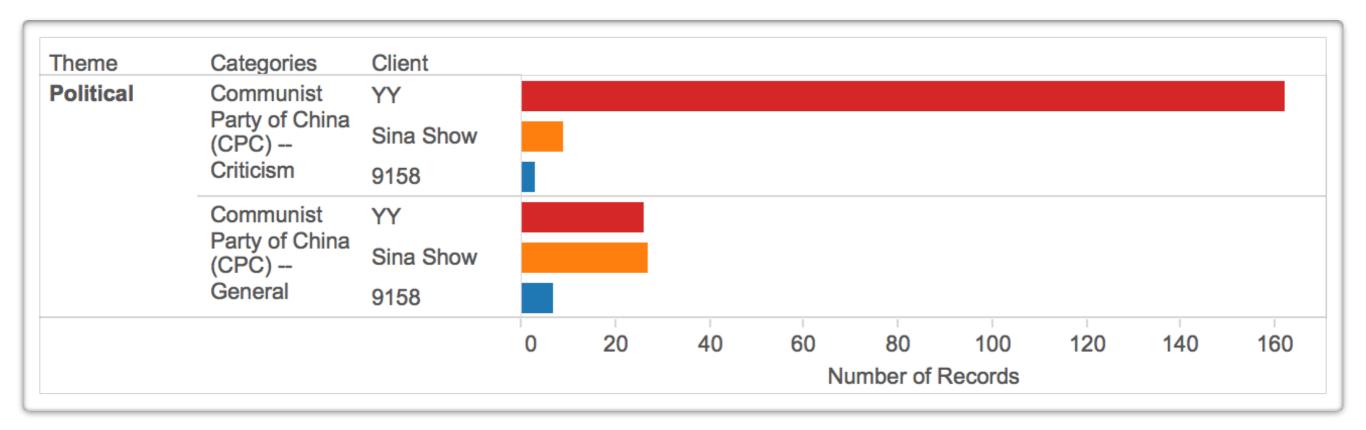
- Gambling. e.g., 多伦多娱乐城 "Toronto casino"
- Drugs, firearms, other illicit goods,
 e.g., 冰毒制作 "Methamphetamine production"
- Prostitution, sexual interests. e.g, 妓女 "Prostitute" 性伴侣 "Sexual partner"



Political Theme

• 37 content categories

e.g., CPC, religious movements, ethnic minorities, corruption, democracy groups, and terrorism

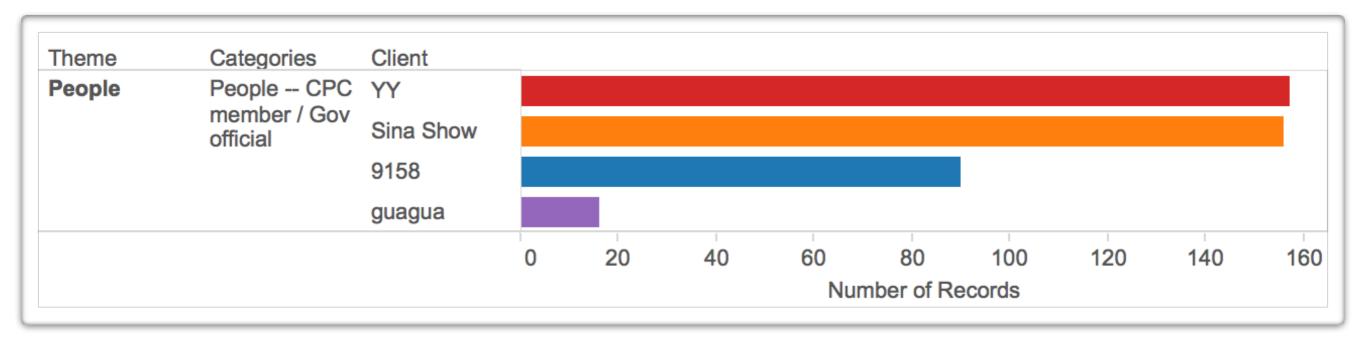


- CPC General: 共产党 "Communist Party"
- CPC Critical: 共匪 "Communist gangsters"

Uyghur Sina Show 9158	
Political Ethnic Groups YY	

- Includes references to Islam, Uyghur groups and culture, terrorism, and general terms
- English: uyghurcanadian
- Chinese: 突厥斯坦 "Turkestan"
- Uyghur: türkistan "Turkestan"
- Arabic script: پارتلىقۇچ ياساش دەرسىلىكى "instructions to make explosives"

Theme	Categories	Client										
People	People CPC	YY										
	member / Gov official	Sina Show										
		9158										
		guagua										
	People	YY										
	Dissident/Activist/ Rights defender	Sina Show										
	C C	9158										
	People Relatives	YY										
	/ associates of government	Sina Show										
	officials	9158										
	People context	YY										
	unclear	Sina Show										
	People Victims	YY										
	of Crime	Sina Show										
			0	20	40	60	80	100	120	140	160	
			Number of Records									



- CPC Leaders: 习近平 "xí jìnpíng"
- Homonyms: 习尽平 "xí jǐn píng"
- Nicknames: 习包子 "Steamed Bun Xi"

习包子 "Steamed Bun Xi"



Theme	Categories	Client				
Events	Event 16th CPC National C	Sina Show				
	Event 18th CPC National C	YY				
	Event 2008 Olympics	YY				
		Sina Show				
	Event 2012 Bloomberg Cen	YY				
	Event 2012 Kashgar Riots	YY				
	Event 2012 Scarborough Sh	YY				
	Event 2014 Kunming attack	YY				
	Event APEC 2014	Sina Show				
	Event Arab Spring	YY				
	Event Bo Xilai Scandal	YY				
		Sina Show				
		9158				
	Event Carrefour Boycott	Sina Show				
	Event Diaoyu / Senkaku	YY				
	islands protests	Sina Show				
	Event Flight MH370	YY				
	Event Forced demolitions/e	YY				
	Event Google leaving China	Sina Show				
	Event Jasmine Revolution	YY				
		Sina Show				
	Event Jiang Zemin death ru	YY				
	Event Occupy Central HK	YY				
	Event Occupy movement	YY				
	Event South Korea Yellow S	YY				
	Event Tibetan Self-immolati	Sina Show				
	Event Urumqi Riots (July	YY				
	2009)	Sina Show				
		9158				
			0	20	40	60
			-		Number of Records	

Theme	Categories	Client									
Events	Event	YY									
	Tiananmen Square (June 4 1989)	Sina Show									
			0	500	1000	1500	2000				
				Number of Records							



YY: references largest number of events (18) 90% of YY event keywords related to June 4 1989

Recent Events in Keyword Lists



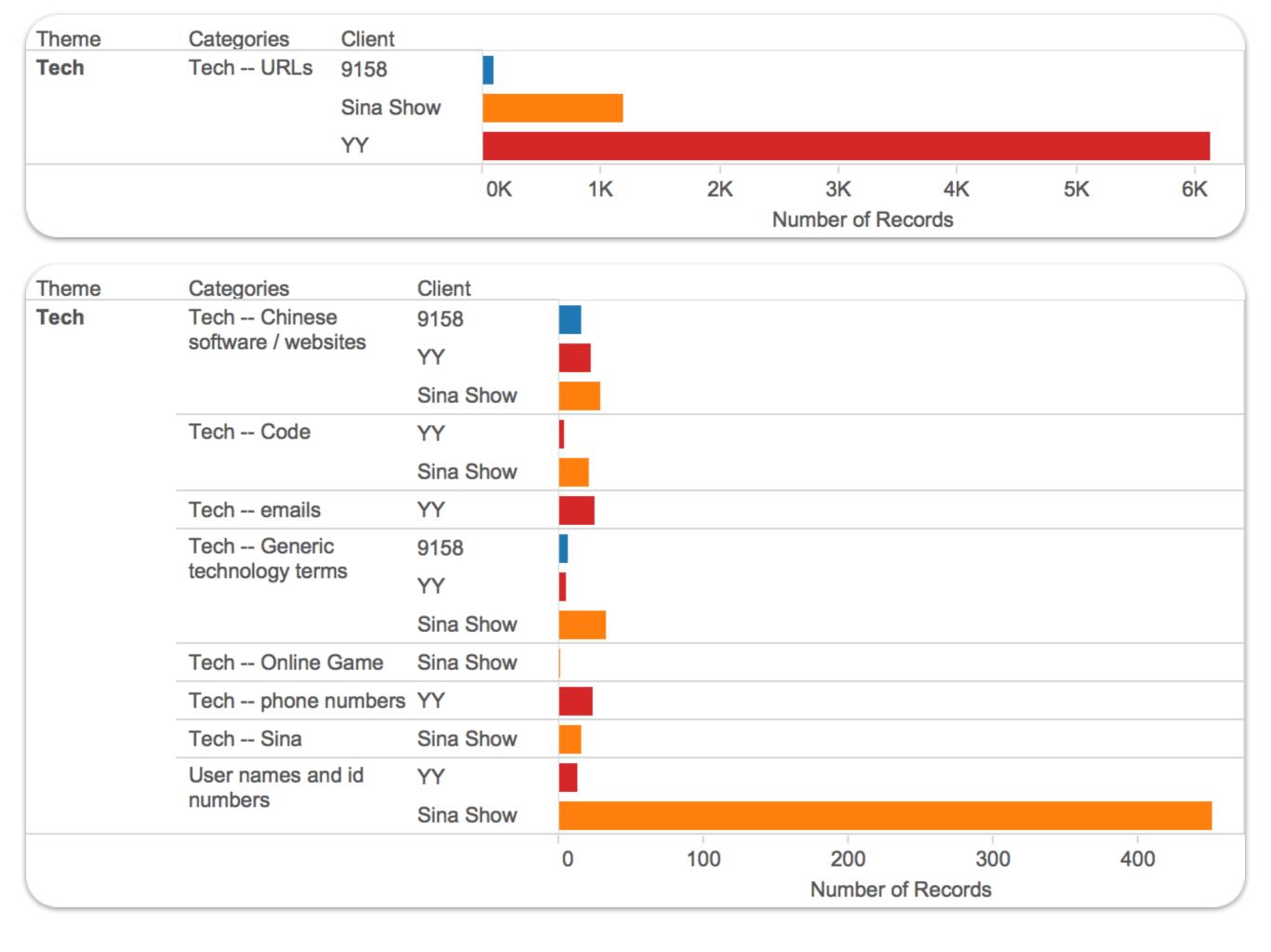
2014 - Occupy Central Protests





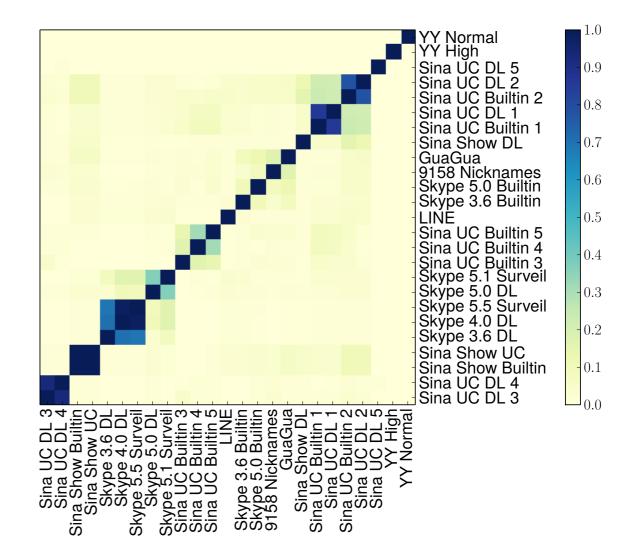
2014 - APEC Summit

2014 - Kunming Terrorist Attack



Discussion

- Inconsistencies in the content and implementation in keyword lists across companies and platforms
- Very limited overlap within SVPs and in comparison to chat clients
 - Supports theories that companies are under general pressure from the government to perform censorship and surveillance, but have flexibility in implementation



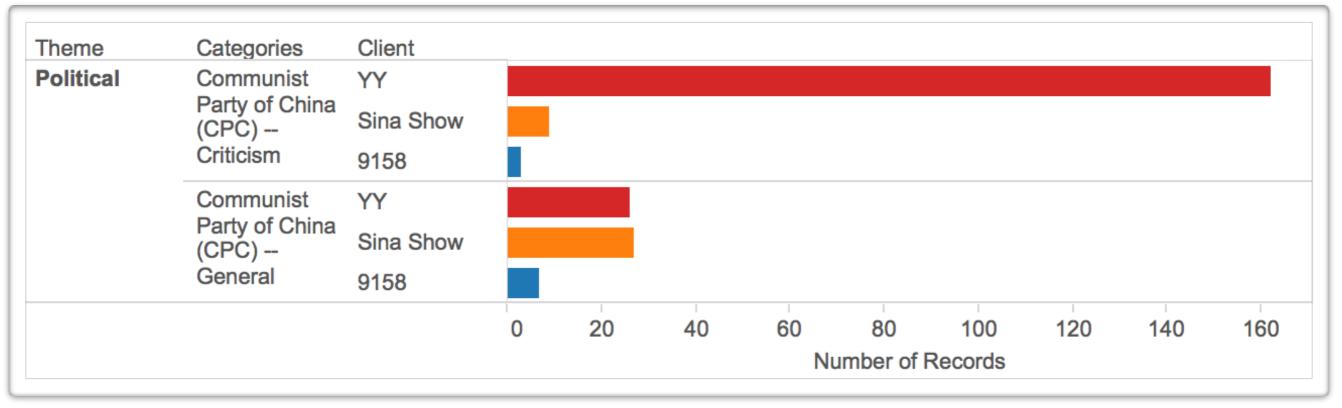
Discussion

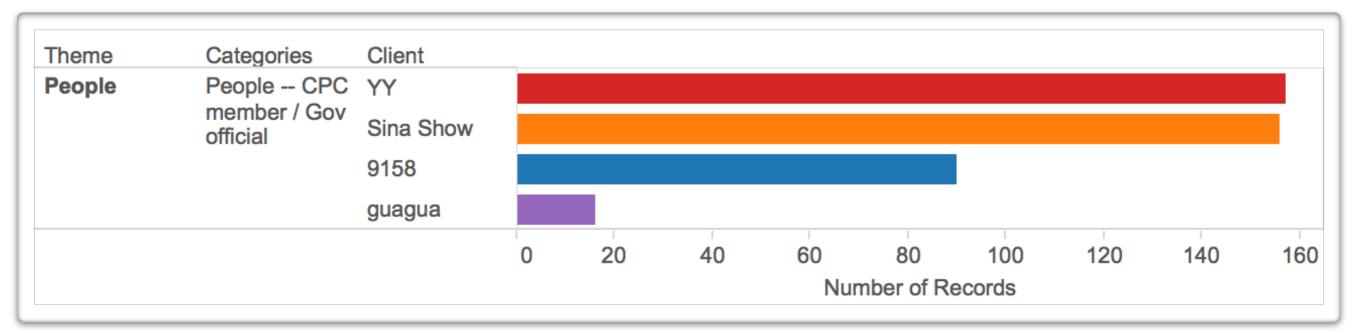
- Targeted content relates to government criticism and collective action
 - King *et al.* analyze post censorship in 1,382 Chinese social media sites.
 - Contend content related to collective action censored but content critical of government permitted
 - Our results reveal sensitive keywords in SVPs and chat clients related to collective action and government criticism
 - Calls into question generalizability of King et al. theory

Collective Action Keywords

Theme	Categories	Client								
Events	Event 2012 Kashgar Riots	YY								
	Event Arab Spring	YY								
	Event Diaoyu / Senkaku	Sina Show								
	islands protests	YY								
	Event Jasmine Revolution	Sina Show								
		YY								
	Event Occupy Central HK	YY								
	Event Tibetan Self-immolati	Sina Show								
	Event Urumqi Riots (July	9158								
	2009)	Sina Show								
		YY								
Political	Protest	9158								
		Sina Show								
		YY								
	Tuidang	Sina Show								
		YY								
			0	10	20	30	40	50	60	70
						Numbe	er of Rec	ords		

General / Critical Government Keywords





Keyword List Updates

- Hourly data collection: YY (Feb. 7, 2015), 9158 (Feb. 24 2015), Sina Show (March 11 2015)
- YY Normal (3 updates), YY High (21 updates)
- 9158 *Chat* (8 updates)
- Sina Show *Downloaded* (0 updates)
- GuaGua does not download updates

YY Updates

- 32 keywords referencing names of Christian songs added April 23, removed April 30 (YY *Normal*)
- e.g, 主耶稣永远陪伴我 "Jesus is always with me"
- May have been in response to controversy over recent church demolitions



9158 Updates

- Keywords added on May 28 2015
- 戴建勇 "Dai Jianyong" (name of Shanghai artist)
- 麻痹哥 "Paralysis brother"



Data: https://github.com/citizenlab/chat-censorship

Special Thanks:

Jakub Dalek, Mamatjan Juma, Greg Fay, Zubayra Shamseden, and Greg Wiseman

Sponsors:

Open Technology Fund Information Controls Fellowship Program John D. and Catherine T. MacArthur Foundation