Five incidents, one theme:

Twitter spam as a weapon to drown voices of protest

John-Paul Verkamp
Minaxi Gupta

Indiana University
verkampj/minaxi@cs.indiana.edu

Motivation

- ~~Social media solely for contacting friends~~
- Social media as news source
- Social media as politics
- Social media as a part of life

Incidents

Five incident over two years:

Syria	April 2011	#syria
China ‘11	April 2011	#aiweiwei
Russia	December 2011	#триумфальная
China ‘12	March 2012	#freetibet
Mexico	May 2012	#marchaAntiEPN

Methodology:
Data collection

- Twitter data from the Truthy Project (http://truthy.indiana.edu/)
- Varies from 1/10 to 1/15 of all tweets
- Mostly continuous, some interruptions in data collection

Methodology:
Hashtag expansion

1. Let S = {seed hashtag} (#syria, #aiweiwei, etc)
2. Let T = {tweet | tweet contains a hash in S}
3. Let S’ = {top *n* hashtags in T}
4. If S ≠ S’, let S = S’ and goto 2

- Stabilizes after 2-4 iterations in all cases
- Tested with all user’s tweets, did not substantially change findings

Methodology:
Hashtag expansion

Syria: #syria, #bahrain, #egypt, #libya, #syria, #jan25 (Egypt), #feb14, #tahrir (Egypt), #yemen, #feb17 (Libya), #kuwait

China ‘11: #aiww, #aiweiwei, #cn417 (Jasmine), #5mao (5 May), #freeaiww, #freeaiweiwei, #cn424 (Jasmine), #tateaww, #cnjasmine

Russia: #чп (abbr of Чрезвычайное Происшествие, extraordinary incident), #6дек (Dec 6), #5дек (Dec 5), #выборы (elections), #митинг (meeting), #триумфальная (Triumphal Square), #победазанами (victory is ours), #5dec, #навальный (surname, likely Navalny), #ridus

Methodology:
Hashtag expansion

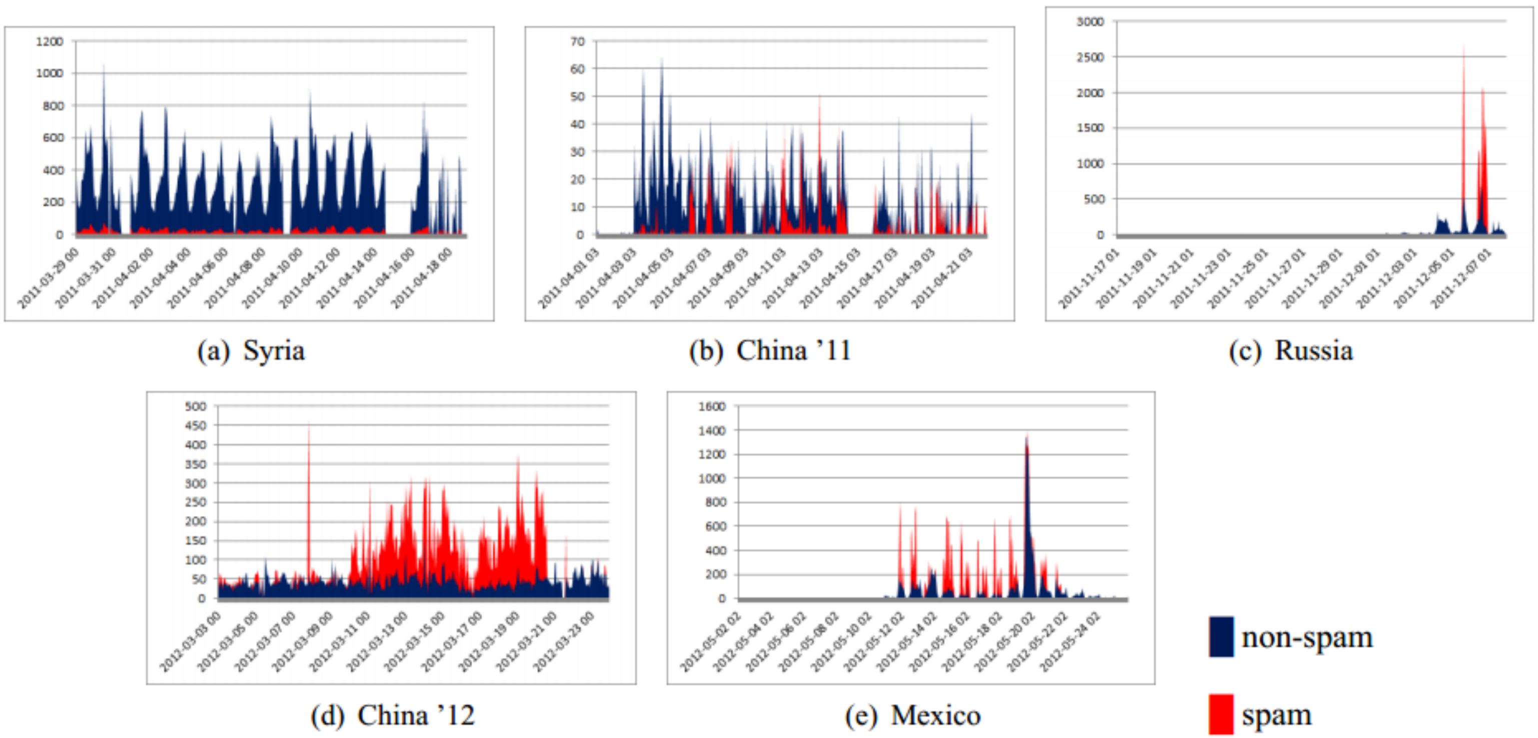
China ‘12: #tibet, #freetibet, #china, #monday, #西藏 (Tibet), #freetibet Free Tibet #tibet, #freetibet, #china, #monday, #西藏 (Tibet), #beijing, #shanghai, #india, #apple, #hongkong

Mexico: #marchaantiepn, #marchaantipeña, #marchamundialantiepn, #marchayosoy132 (I am 132nd to march), #votomatacopete (vote for another), #epn, #epnveracruznotequiere (no more EPN), #pr, #amlocomp (initials of competitor), #yosoy132

Methodology:
Incident sizes

Incident	Tweets	Accounts	Comments
Syria	1,540,000 non-spam 107,000 spam	157,000 non-spam 3,000 spam	**Most overall tweets** Smallest % spam tweets
China ‘11	58,000 non-spam 15,000 spam	3,950 non-spam 550 spam	Smallest attack Relatively low % spam
Russia	151,000 non-spam 338,000 spam	12,000 non-spam 25,000 spam	Highest % spam **Highest number of spam accounts**
China ‘12	227,000 non-spam 600,000 spam	10,00 non-spam 1,700 spam	**Highest % spam** **Fewer + high volume spam accounts**
Mexico	306,000 non-spam 498,000 spam	28,800 non-spam 3,200 spam	High % spam **Fewer + high volume spam accounts**

Analysis of tweets:
Daily tweet volume

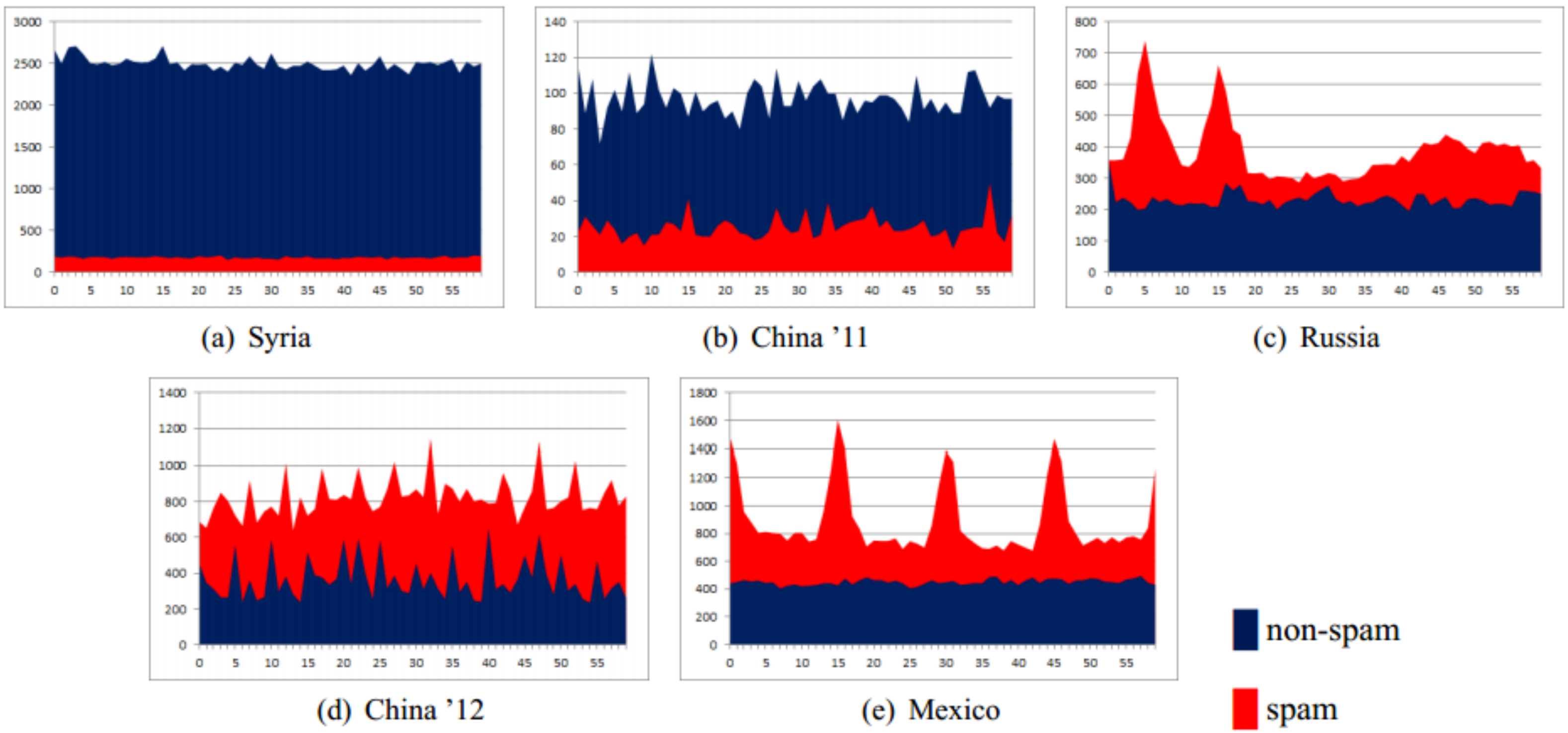


(a) Syria (b) China '11 (c) Russia

(d) China '12 (e) Mexico

- China'11, Russia, and Mexico show definite spikes of activity
- Syria, China '11, and China '12 are more sustained

Analysis of tweets:
Timing of tweets



(a) Syria (b) China '11 (c) Russia (d) China '12 (e) Mexico

- Russia and Mexico show automated (cron related?) spikes
- Diurnal activity (not pictured) generally matches that of normal usage

Analysis of tweets:
Tweet meta content

Incident	URLs (spam / non-spam)		Mentions (spam / non-spam)		Retweets (spam / non-spam)	
Syria	**41.0%**	**96.4%**	59.1%	60.4%	44.2%	45.2%
China ‘11	**58.8%**	**36.2%**	69.7%	68.3%	**3.3%**	**29.8%**
Russia	**2.8%**	**36.8%**	**4.2%**	**54.6%**	**3.1%**	**35.8%**
China ‘12	60.6%	64.5%	**81.3%**	**36.4%**	**0.2%**	**13.7%**
Mexico	**1.0%**	**32.8%**	**1.9%**	**80.7%**	**1.6%**	**68.9%**

- Spam with URLs is often product placement; (unrelated) news stories
- Spam has significantly fewer retweets (other than in Syria)
- Number of mentions is a good indicator, but could go either way

Analysis of tweets:
Most common content

Syria	*Spam*:	rt, #bahrain, #egypt, #libya, the, in, #syria, to, في (in), of
	Non-spam:	rt, #egypt, #bahrain, #libya, the, in, #syria, في (in), to, من (of)
China '11	*Spam*:	#aiww, rt, #5mao (May 5), #cn417, 艾未未的童话涉嫌抄袭 (headline about Ai Weiwei), url_1, #cn424, url_2, #aiweiwei, #china
	Non-spam:	rt, #aiww, #aiweiwei, #cn417, ai, @aiww, #freeaiww, #5mao, the, #freeaiweiwei
Russia	*Spam*:	на (on), #победазанами (victory is ours), не (no), #чп, и (and), #выборы (elections), в (in), #6дек (Dec. 6), я (I), площади (areas)
	Non-spam:	#выборы, rt, в, на, #чп, и, не (not), за (for), с (with), #митинг (meeting)
China '12	*Spam*:	#tibet, #freetibet, @degewa, @tibet, #西藏 (#tibet), #degewa, #china, and, @sfchoi8964, #315
	Non-spam:	#china, #tibet, rt, in, #beijing, #shanghai, the, to, #hongkong, #freetibet
Mexico	*Spam*:	#marchaantiepn, marcha (march), la (the), de (of), anti, epn (initials), i, rt, #marchaantipeña, marchaantiepn
	Non-spam:	#marchaantiepn, la, rt, de, a, en (in), no, el (the), que (that), y (and)

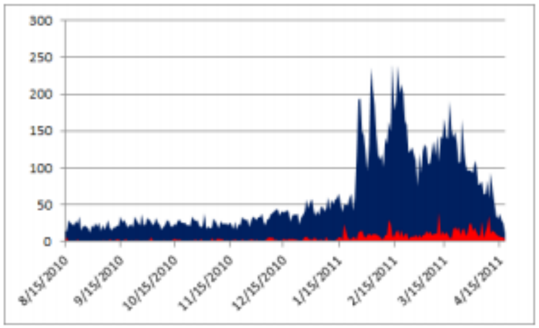
- Hashtags expected, because that's how data was collected
- China '11: Two specific URLs (for products) appeared in many spam tweets
- Russia: Stop words are much more common in non-spam
  Retweet indicators are not common in spam
- China'12: Spammers often targeted a small set of users with mentions

Analysis of tweets:
Tweet recipients

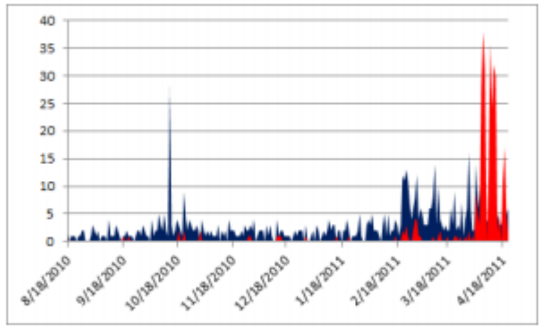
Incident	@non-spam	@spam	neither
Syria	4.7%	**78.3%**	17.0%
China ‘11	1.1%	21.5%	**77.5%**
Russia	**10.7%**	**63.8%**	25.4%
China ‘12	0.7%	**75.0%**	24.3%
Mexico	4.8%	51.6%	43.6%

- @non-spam / @spam are people that tweeted at least once in the incident
- Each incident shows spammers creating internal social mention networks
- China ‘11 and Mexico were connecting to other people

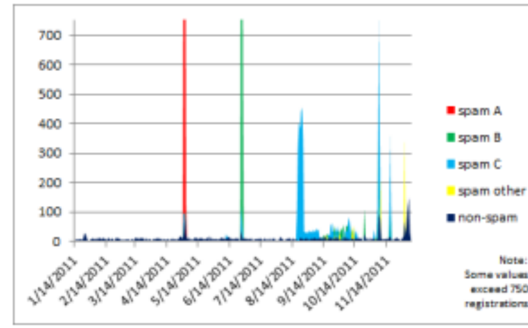
Analysis of accounts: Registration



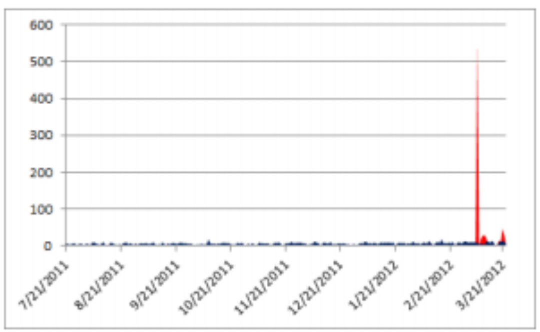
(a) Syria



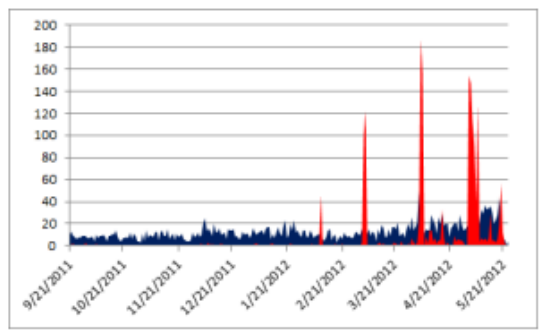
(b) China '11



(c) Russia



(d) China '12



(e) Mexico

non-spam

spam

- All but Syria have registration blocks
- Russian blocks each have their own username patterns

Analysis of accounts: Usernames

Syria — *Often end in numbers, patterns less common*

zuhair77, GC814, walidraafat, ToQiiiZ, GeorgiaKillick0, libyana1702, Bahraini61, ScottsdaleReb, Updates2424

China ’11 — *Often end in numbers, patterns less common*

cnjs2, cnjs5, cnjs10, cnjs11, cnjs12, cxbenben113, dabenben222, huashengdun111, huashengdun203

Russia — *Most are {name}{name} or {initial}{name}; vary by registration block*

SScheglov, SSchelkachev, SSchelkonogov, SSchelokov, SSchemilov, SScherbakov, SShabalin, SShabarshin

China ’12 — *Most are {name}{name}{random/number}, max length*

LanelleL4nelle6, LanieSl1dek1103, LarondaGuererro, LatanyaZummoMNS, LatarshaWeed181, LauraHelgerm1nV

Mexico — *Most are {name}{name}{number}, max length*

AnaAvil58972814, AnaAvil76571383, AnaLope95971326, AnaRive02382949, AnaSuar79305176, AnaSuar83449134

Analysis of accounts:
Default profile and image

	Default profile		Default image	
Incident	spam	non-spam	spam	non-spam
Syria	46.2%	42.9%	9.4%	6.0%
China ‘11	**89.4%**	**51.2%**	12.3%	12.6%
Russia	57.8%	34.7%	7.8%	11.1%
China ‘12	**95.1%**	**47.8%**	**59.0%**	**11.8%**
Mexico	**1.7%**	**27.0%**	0.6%	3.0%

- Earlier incidents show higher defaults among spam accounts
- Mexico reverses this trend

Summary of findings

- Spam often shows a distinct spiking pattern
- There can be indications of scheduled activity; however diurnal patterns were matched
- Non-spam tweets use more stop words; Chinese language analysis is difficult
- URLs, mentions, and retweets vary between spam and nonspam but not consistently
- Spam accounts are registered in blocks with generated usernames
- Default accounts are a good indicator of spammers in older incidents

Obligatory question slide