

中国社会化媒体研究的已知和未知

Charting the Landscape of Chinese Social Media: What We Know and What We Don't Know

第11届全国Web信息系统及其应用科学会议
Web Information Systems and Applications (WISA2014)

Jonathan J. H. Zhu 祝建华



香港城市大學
City University
of Hong Kong



Web Mining Lab
互聯網挖掘實驗室

Mapping Landscape of Chinese Social Media



Source: <http://www.techinasia.com/china-social-media-landscape-infographic-2014/>

Mapping Research on Chinese Social Media?



We need an organizational framework!

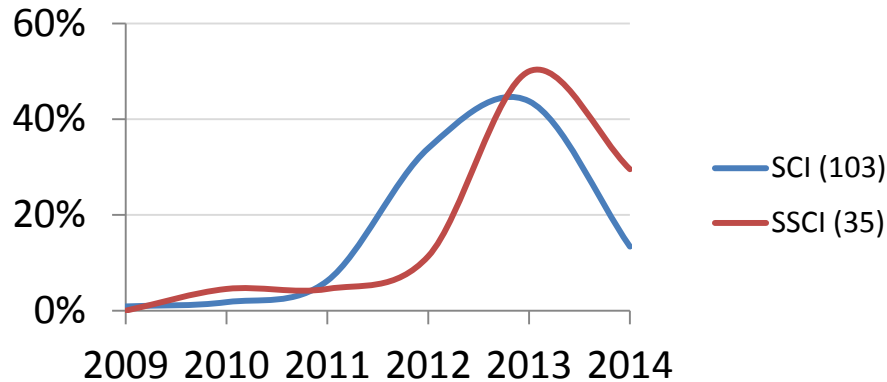
The 5W Model for Assessment

- Harold Lasswell (1948): Communication is a process of who says what to whom through which channel with what effects.

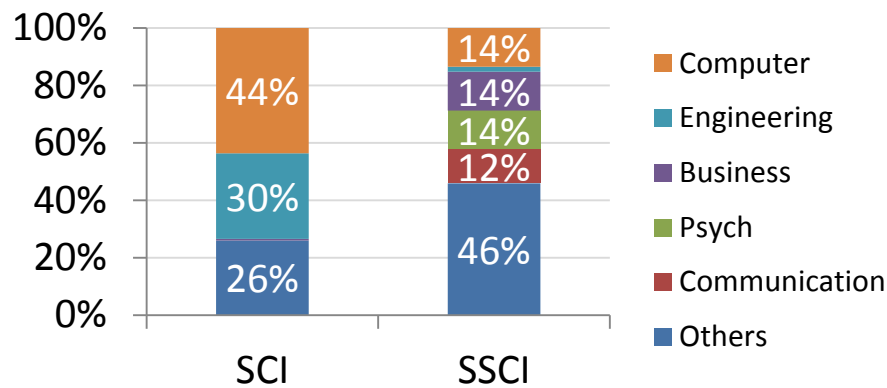


SCI/SSCI Publications on Chinese Social Media (N=138)

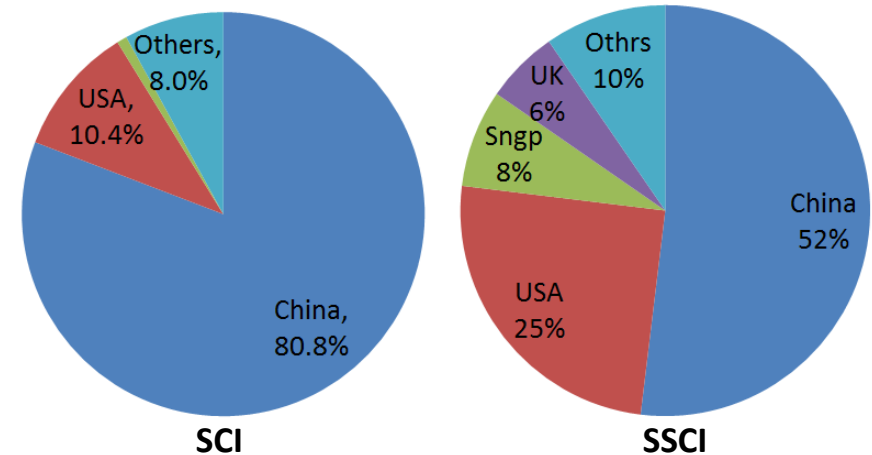
a. Annual Growth



b. Leading Disciplines



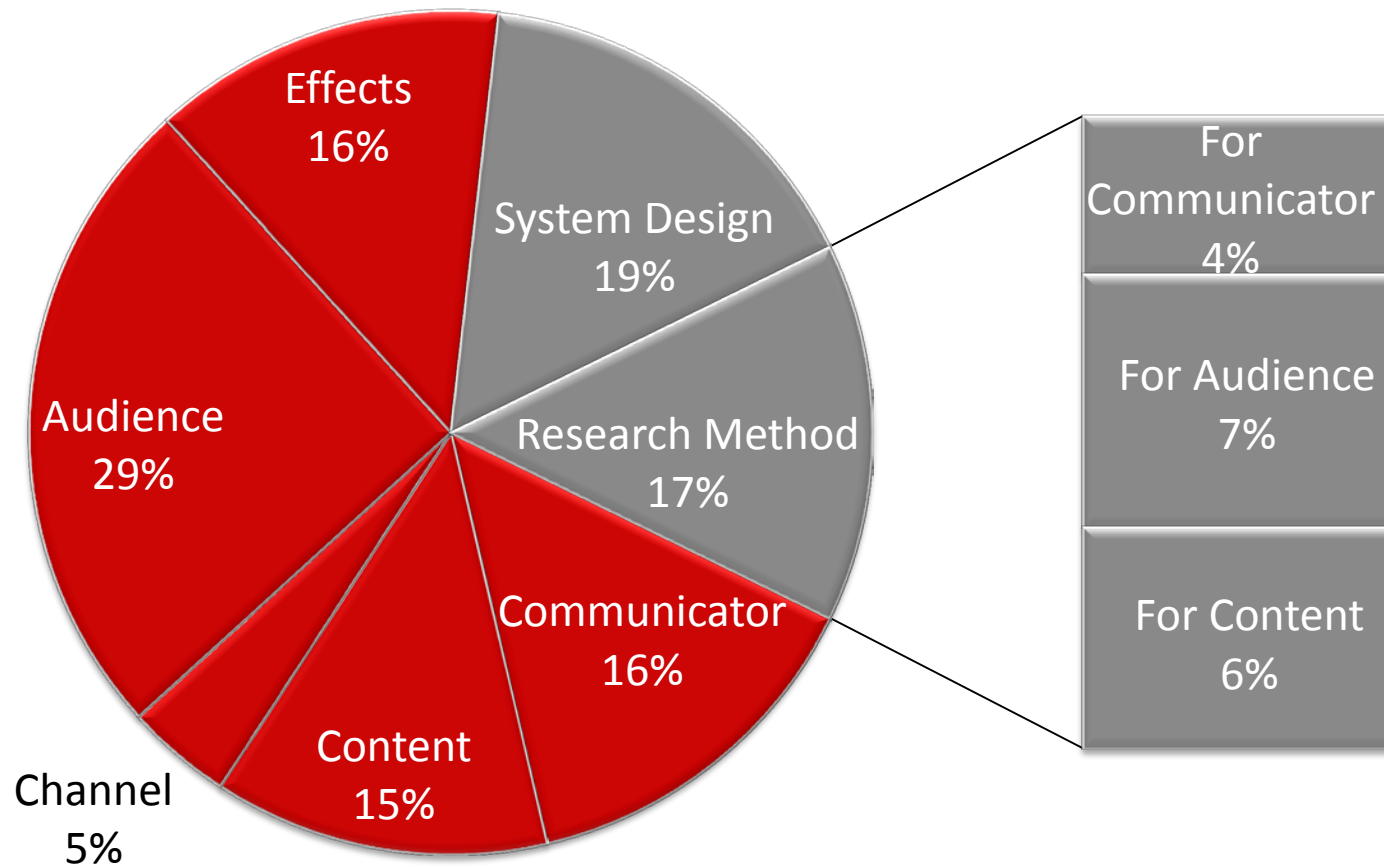
c. Leading Nationals



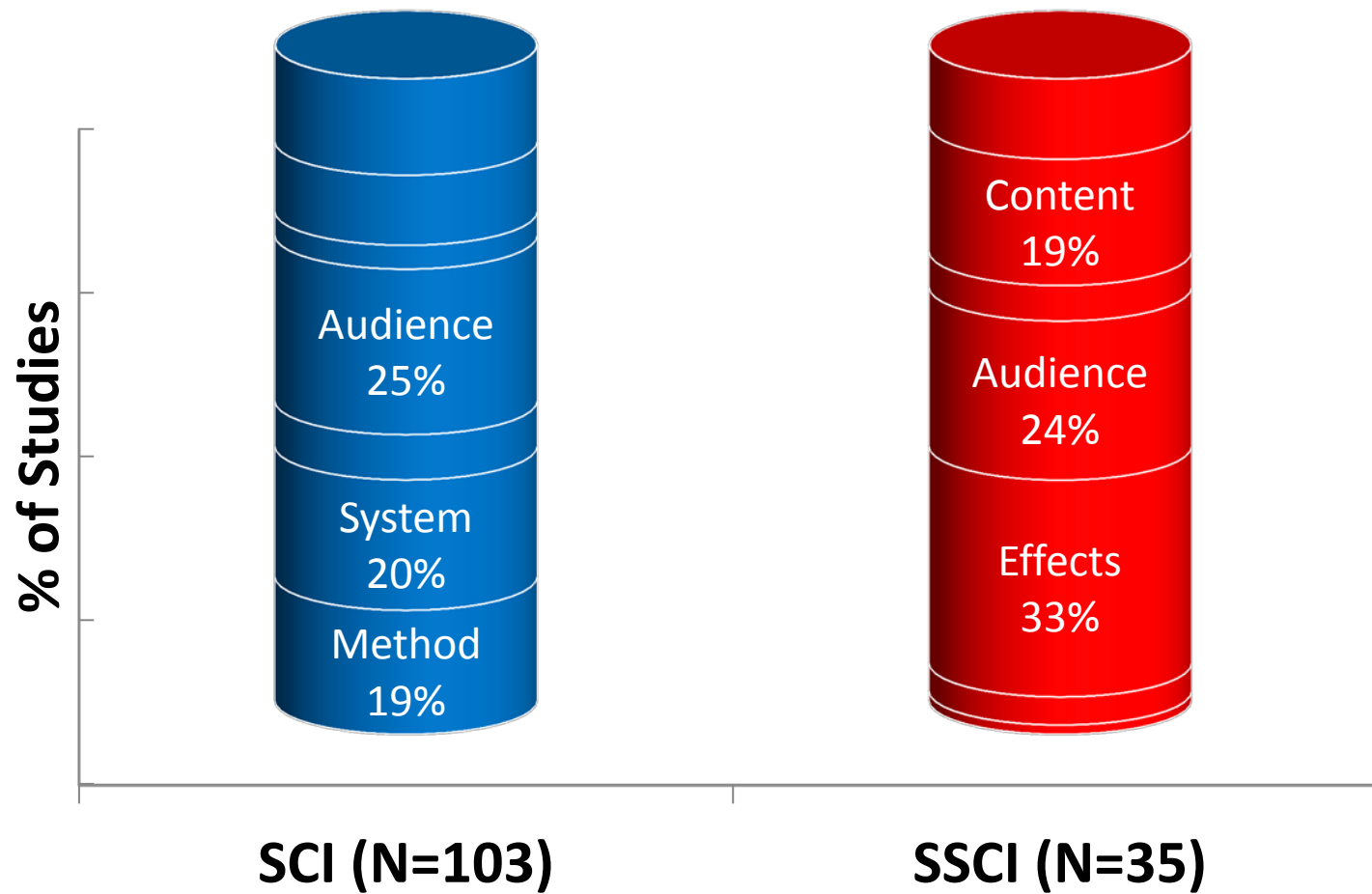
d. Leading Institutions

Rank	SCI	SSCI
1	CAS	Nanyang
2	Tsinghua	Tsinghua
3	Peking	Peking
3	Beihang	Beihang
3	BUPT	CityU HK
3	SJTU	HUST

Distribution across 5Ws



Differences between SCI and SSCI



Common Themes Emerged

1. Classification/Clustering
2. Cross-Site Comparison
3. Community /Group Detection
4. Hot/Crisis Events
5. Information Diffusion /Propagation
6. Keywords /Opinion /Sentiment Extraction
7. Opinion Leaders
8. Privacy Concerns
9. Recommender System
10. Social Capital, Social Movements, Collective Actions
11. Spammer/Zombie Detection
12. User Behavior, Usage Pattern

Common Themes by 5Ws

Communicator	Content	Audience	Channel	Effects
Opinion Leaders	Topic Classification		Information Diffusion	
Spammer /Zombie Detection	Keywords /Opinion /Sentiment Extraction	User Behavior; Usage Pattern	Cross-Site Comparison	Social Capital; Collective Actions
Recommender System	Hot/Crisis Events	Community /Group Detection		Privacy Concerns

1. WHO (Communicator)

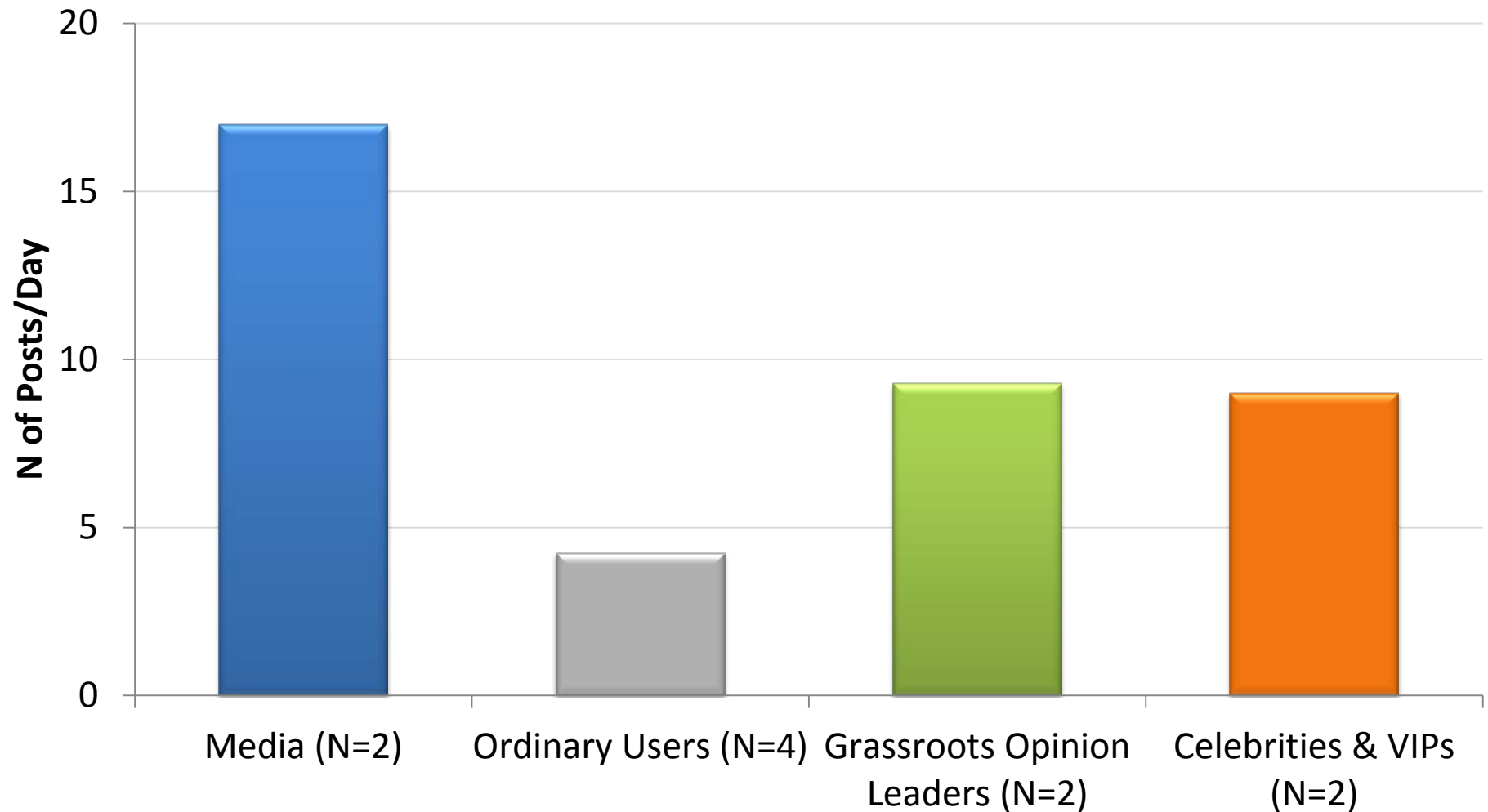
What we know:

- Media professionals
- Elite opinion leaders
- Government mouthpieces
- Rumor mills (“50-cent Party”, “Water Army”, Spammers, etc.)

What we don't know:

- Grassroots opinion leaders
- E-marketers
- Foreign individuals/entities

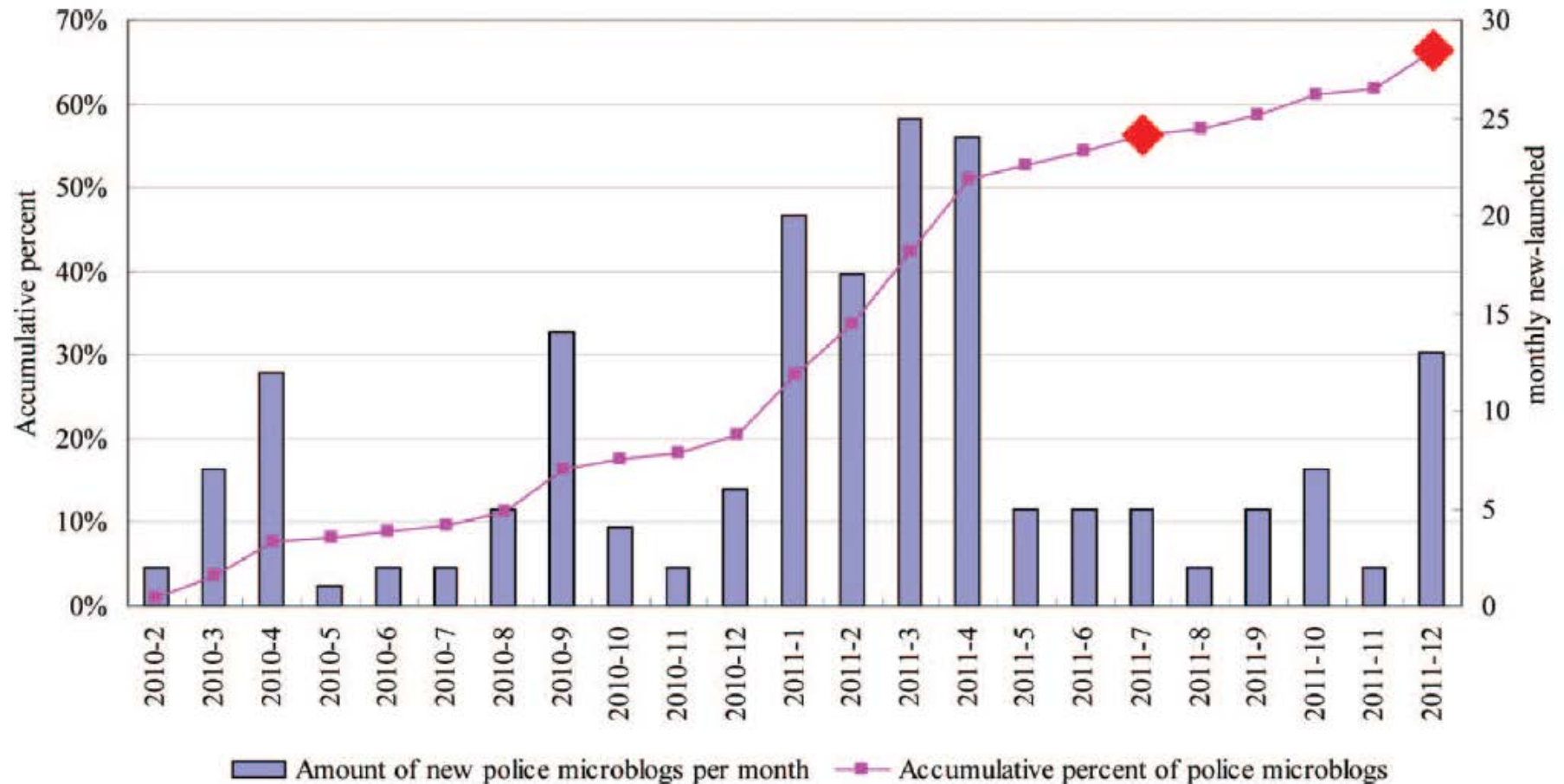
Communicators on Early Weibo (N=10)



Government Presence on Social Media

- Unique features:
 - Centralized review process
 - Widespread “Water Army”
 - Adaption of international practices
- Fundamental problem:
 - Open-networked society vs. closed-centralized government

Growth of Weibo Accounts by Police Bureaus



Predictors of Use of Weibo by Police Bureaus

Significant predictors:

- Government size
- Population size
- Adoption of neighbors
- Adoption of upper-level

Non-significant predictors:

- Revenue
- GRP
- Openness
- Safety level
- E-government efforts
- User size

2. Says WHAT (Content)

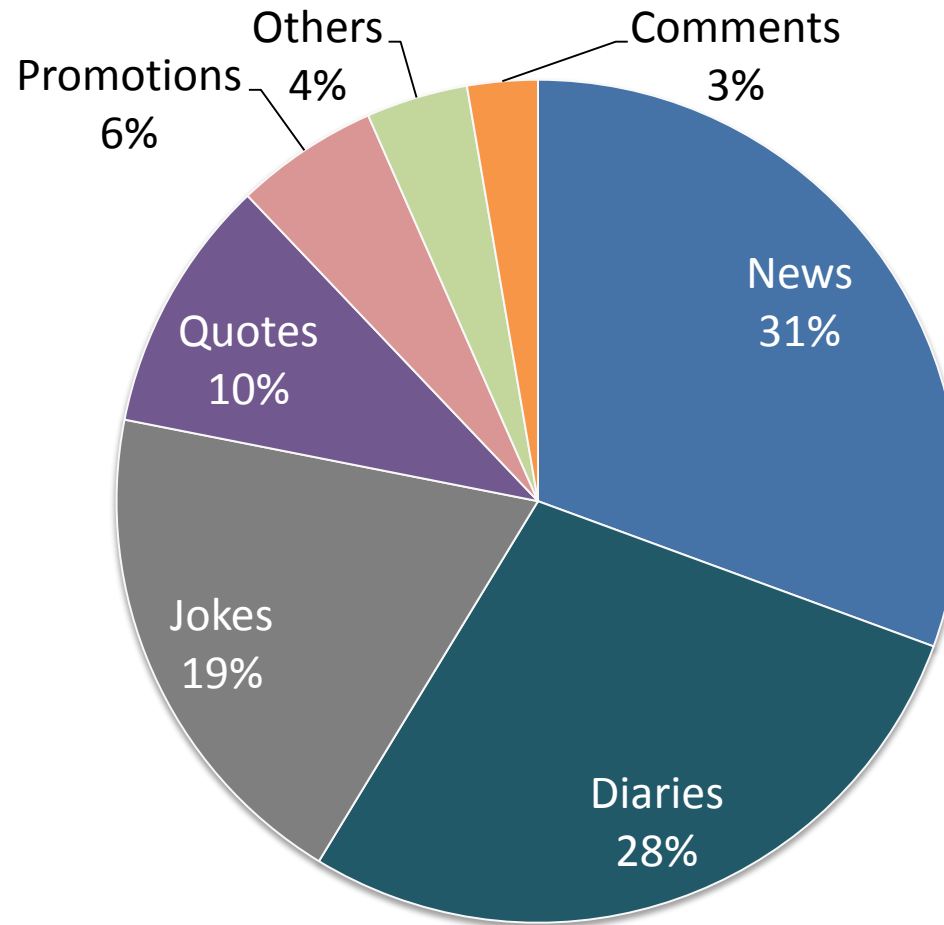
What we know:

- Multiple duplicates
- Contentious
- Censorship

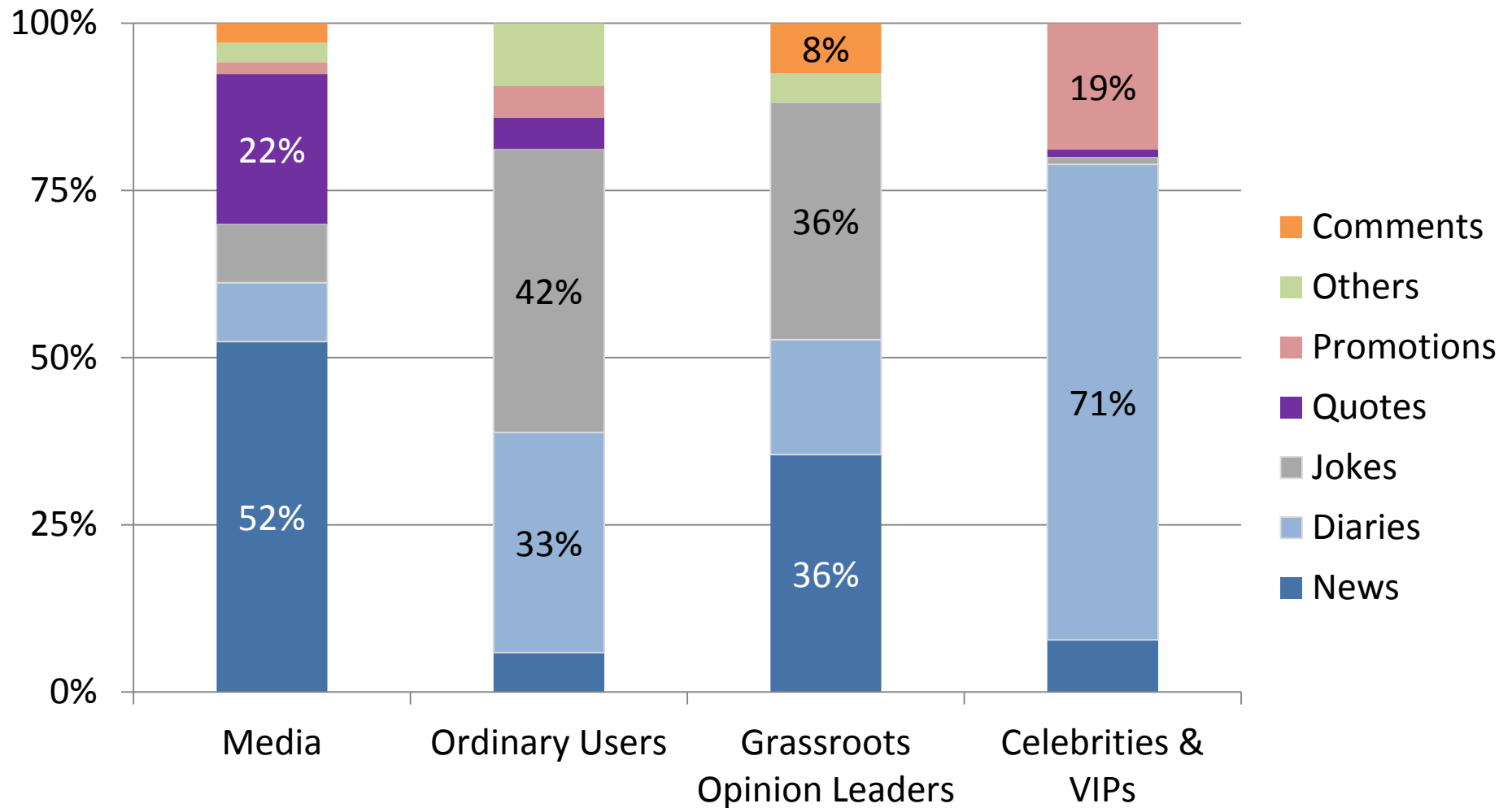
What we don't know:

- Benchmark categorization
- Non-political content (e.g., entertainment)
- Forced vs. self-censorship

Categories of Early Weibo Posts (N=438)



Who Say What on Early Weibo



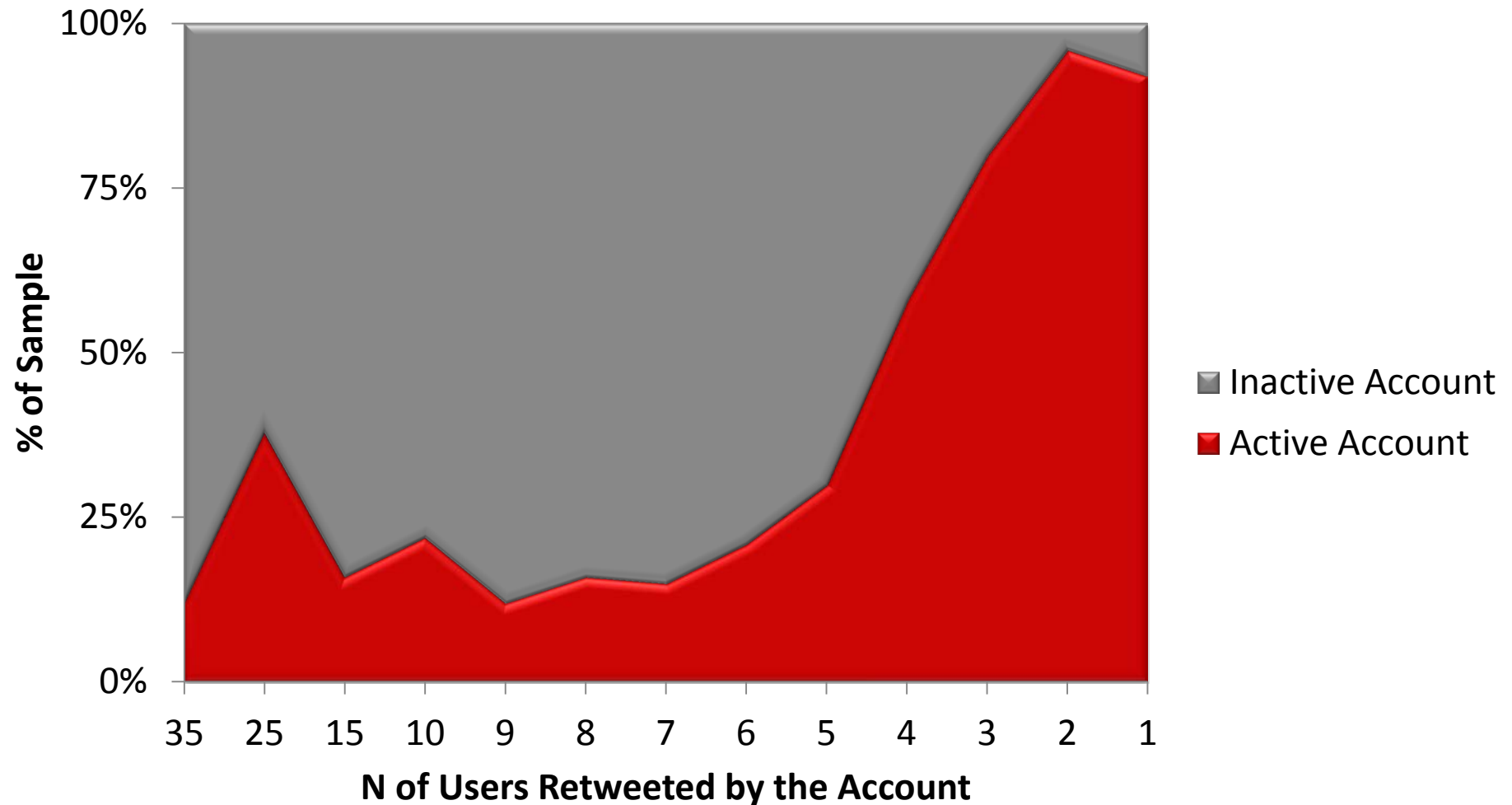
Sina Weibo vs. Twitter on Content

	Weibo	Twitter
Trending Topics	Trivial jokes	News events around the world
Trend Setters	Unverified accounts	Leading news media
Trendy Topics Retweeted	62%	31%
Presentation content	Images, videos, links	Text

What Might Cause the Differences?

- The higher rate of retweeting is likely to be created by spammers, i.e., the accounts that were set up to retweet (promote) targeted users, which also promote the relevant topics to be trending.
- About 1% of the users are suspected spammers, who contribute 1/3 of the total posts or half of the retweeted posts.
- If the spammers are removed, the evolution patterns of trending topics on Weibo becomes similar to that on Twitter.

Detecting Spammers by User-Retweet Ratio

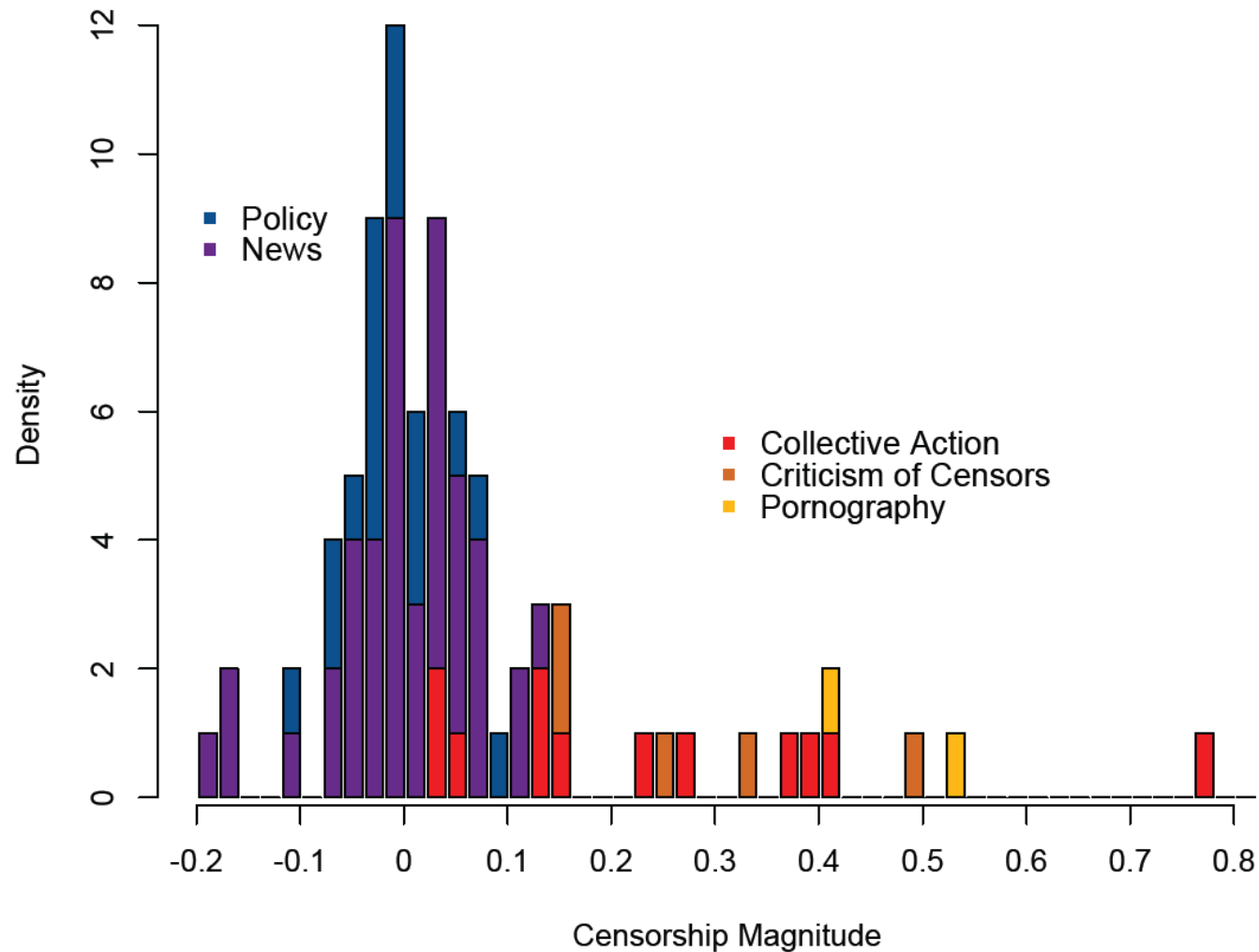


Chinese Censorship

King et al. (2013):

- The largest selective suppression of human expression in history:
 - implemented manually (within a few hours of posting)
 - by 200,000 workers
 - located in government and inside social media firms
- A huge censorship organization:
 - (obviously) designed to suppress information
 - (paradoxically) very revealing about the goals, intentions, and actions of the Chinese leadership

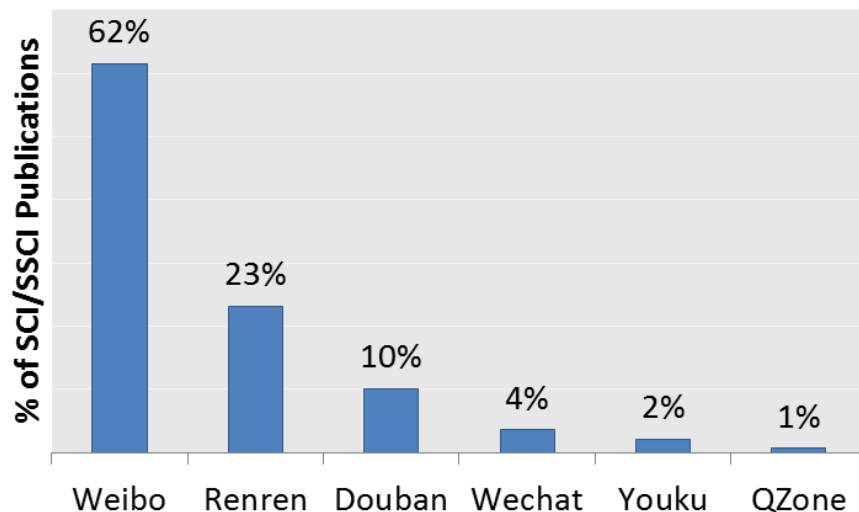
What's Censored



3. Through WHICH Channel

What we know:

- Weibo concentrated
- Compared with Twitter
- Compared with old media



What we don't know:

- Other social media, e.g.,
 - IM (WeChat)
 - BBS (Tianya)
 - SNS (QQ)
 - Blogs, dating, gaming, etc.
- Competition/cooperation among social media

Blogs vs. Microblogs

Method:

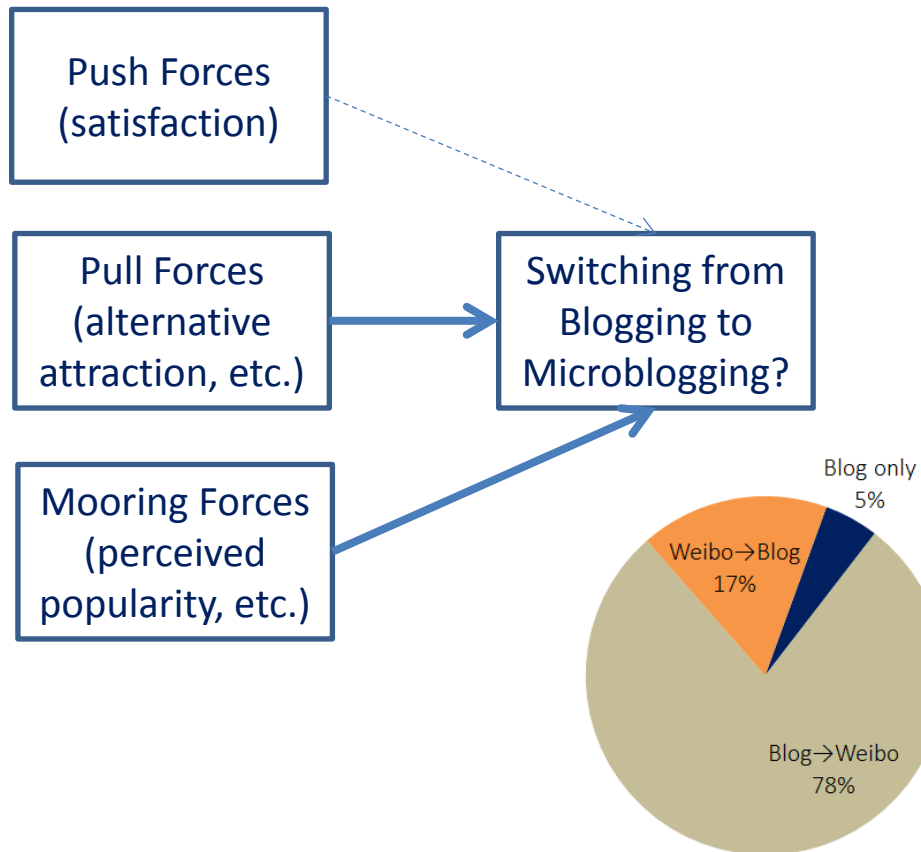
- 300 Sina blog posts on traveling to Hong Kong
- 300 Sina Weibo posts on the same topic
- Manual content analysis of the frequency, direction, and intensity of keywords, themes, visuals, etc.

Findings:

- Blog posts largely to recount past experience
- Weibo posts mainly to express desire to visit in advance
- No formal comparison made as the two samples treated as replicates rather than contrasts

Competitor or Partners

PPM Model:



Method:

- Drew a random sample of Sina users
- Mined their blog/Weibo posts to measure behavior
- Invited them to answer an online questionnaire to measure motivation
- Found most users to co-use both simultaneously

4. To WHOM (Audience)

What we know:

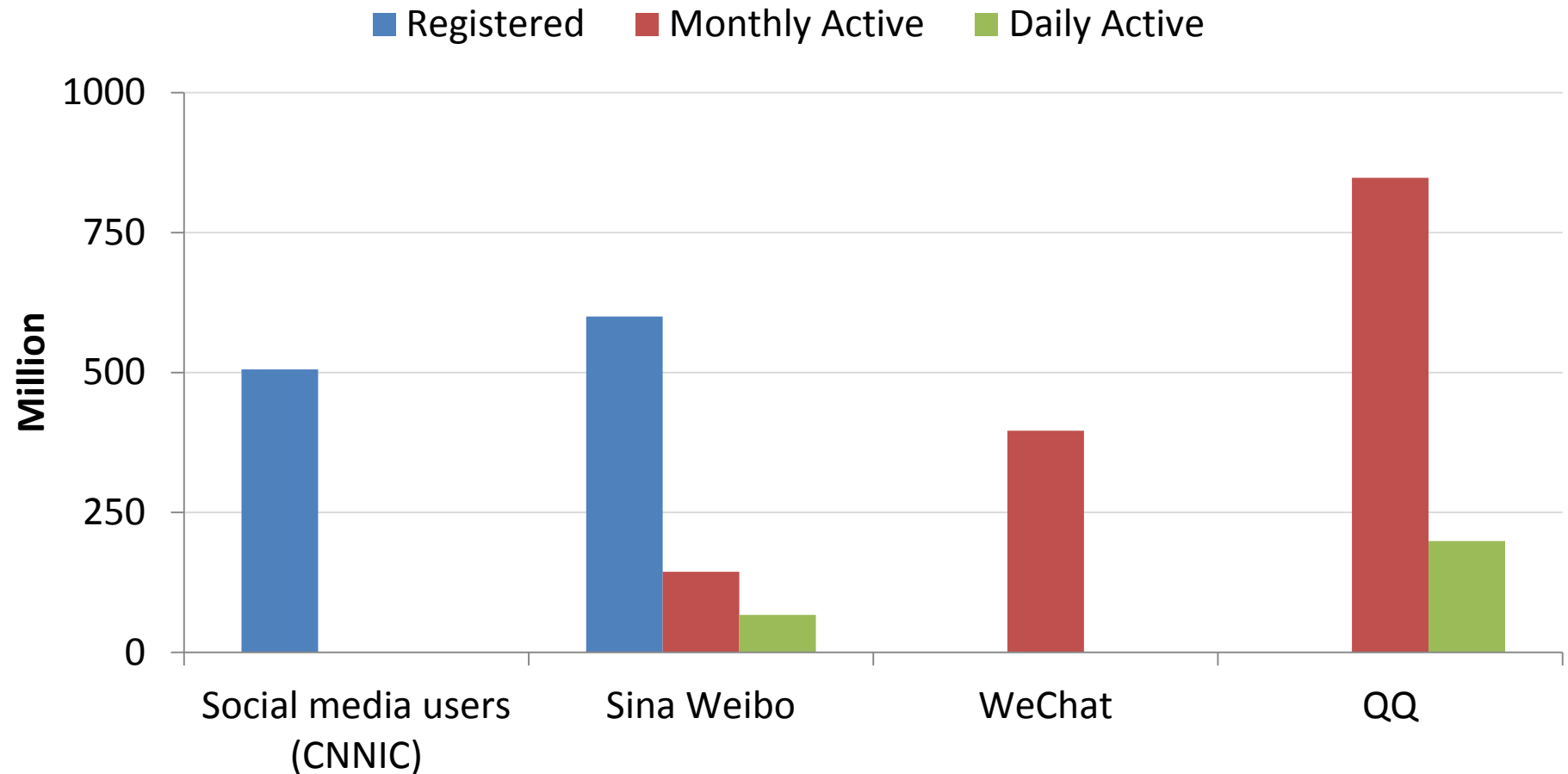
- Massive size
- Skewed participation
- Polarized views

What we don't know:

- User authenticity
- Passive audience
- Inactive audience

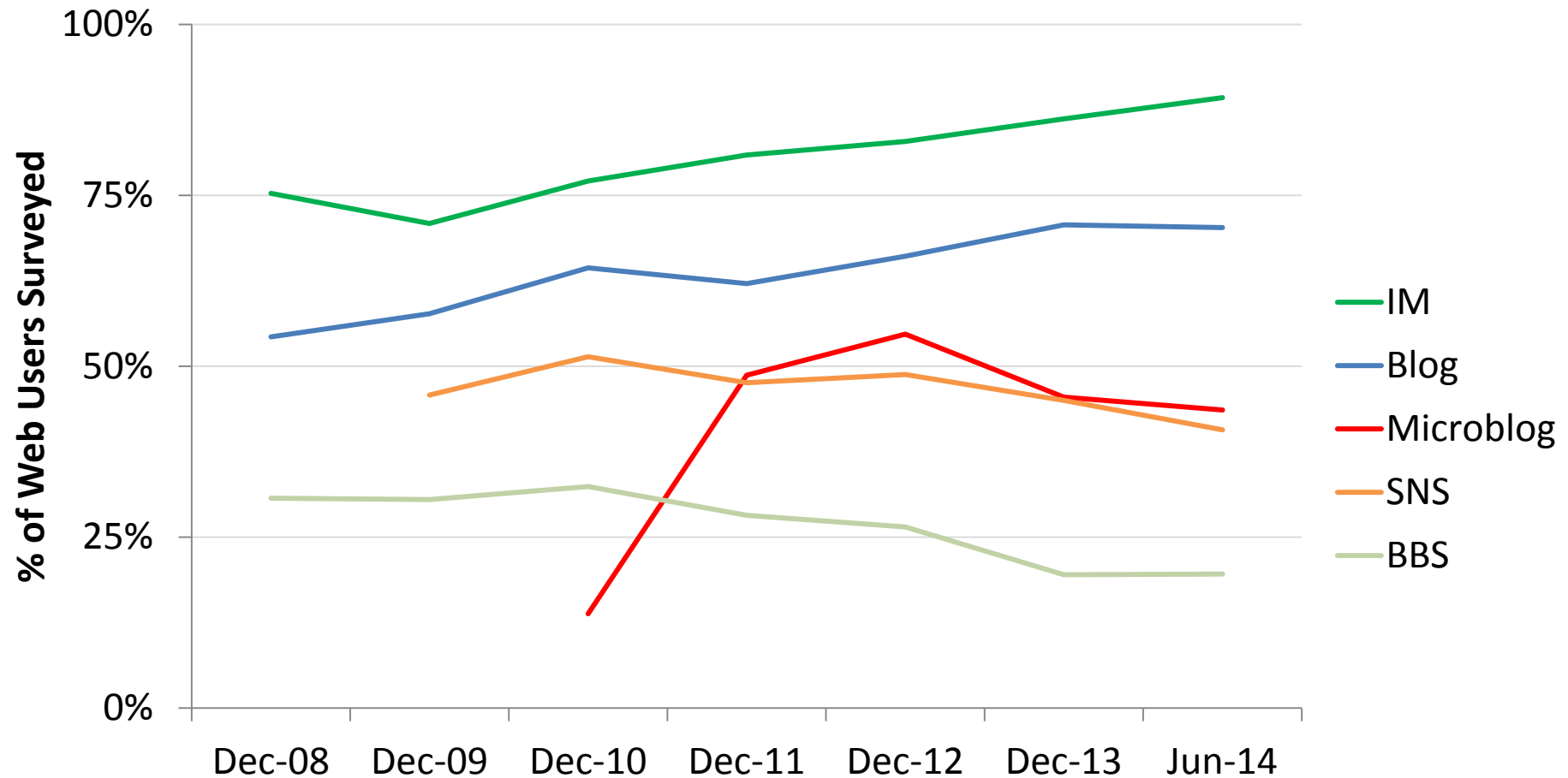


Massive Size of Social Media Users



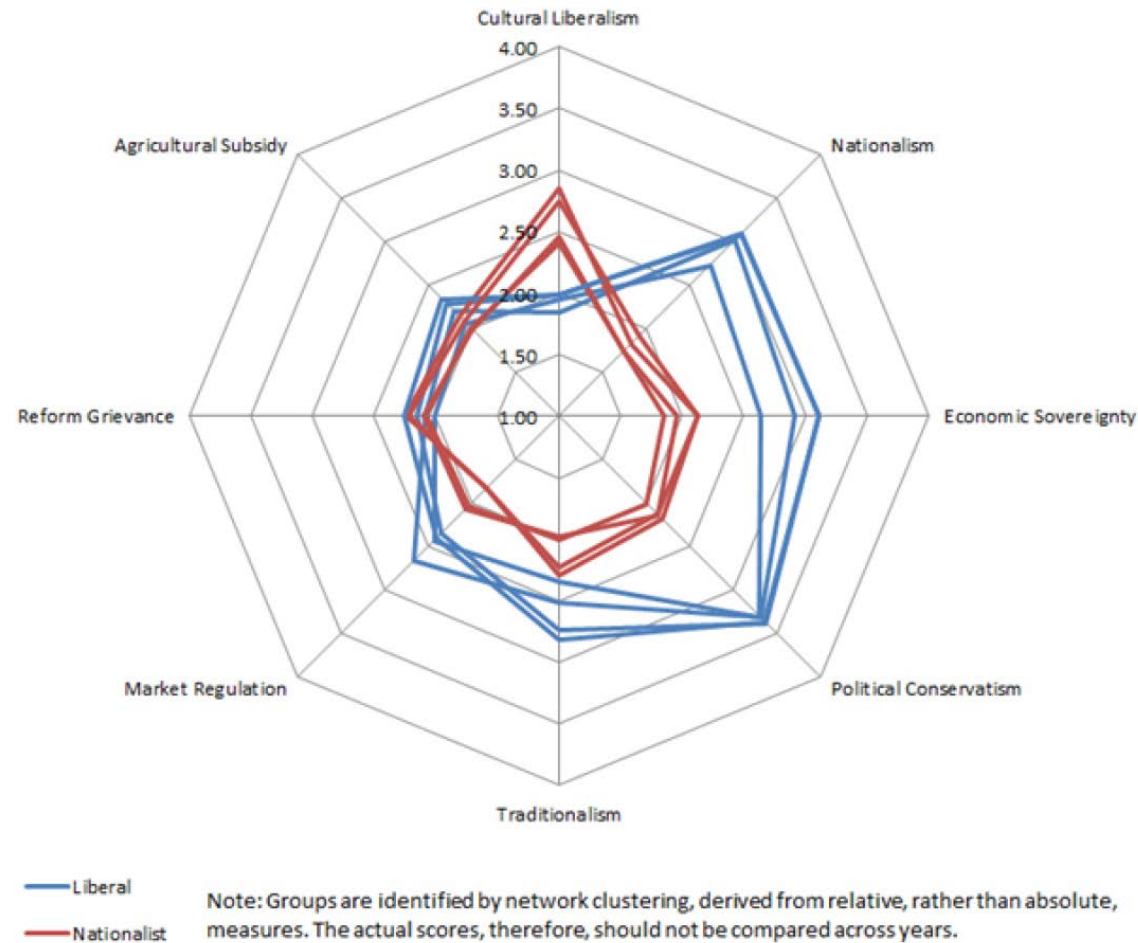
Sources: CNNIC survey report; Sina/Tencent annual reports

Changes in Popularity of Chinese Social Media

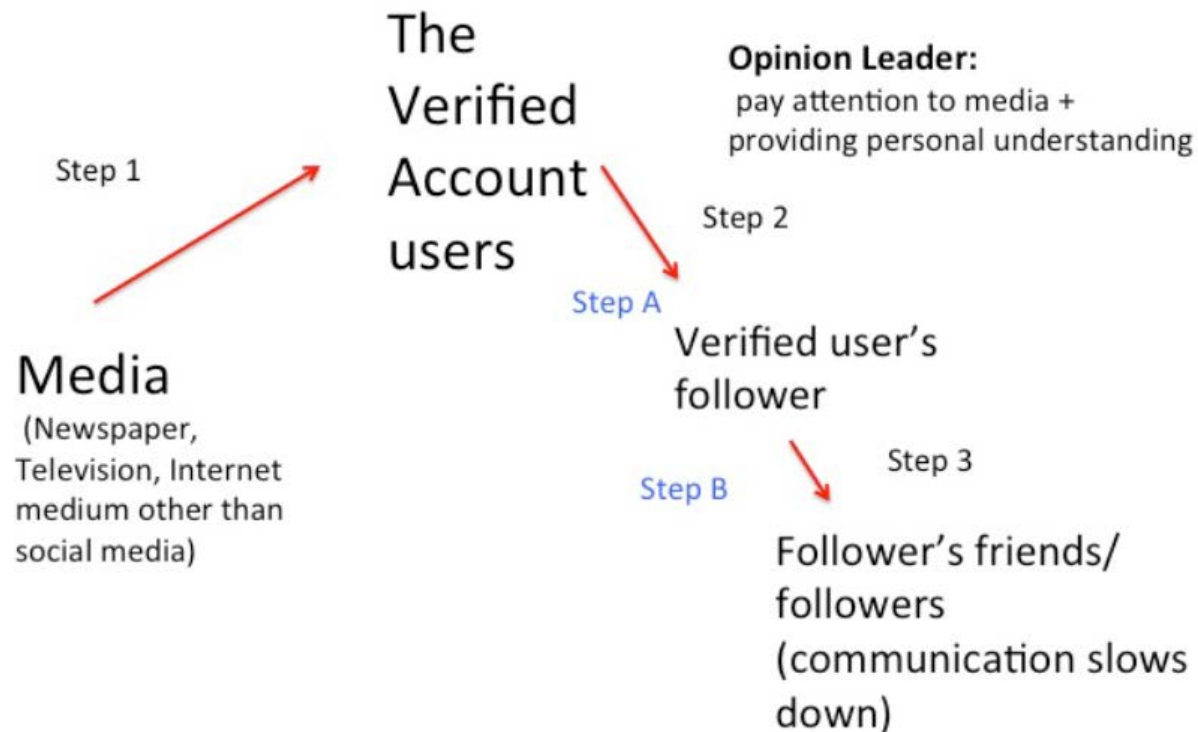


Source: CNNIC Annual Survey Reports (www.cnnic.cn)

Political Preference of Dominant Groups



A 3-step Flow Model of Information Diffusion



Distinctions between Communicator and Audience

Participation Duration	Content Creation	Content Consumption
Continuous	I. Active Communicator	III. Active Audience
Interrupted	II. Inactive Communicator (traceable)	IV. Inactive Audience (completely ignored)

5. With WHAT Effects

What we know:

- Registration = Exposure
- Exposure = Conversion
- Expression = Participation

What we don't know:

- Definition of effects
- Links to content
- Links to offline life
- Effects on individuals vs. society



Expression vs. Participation

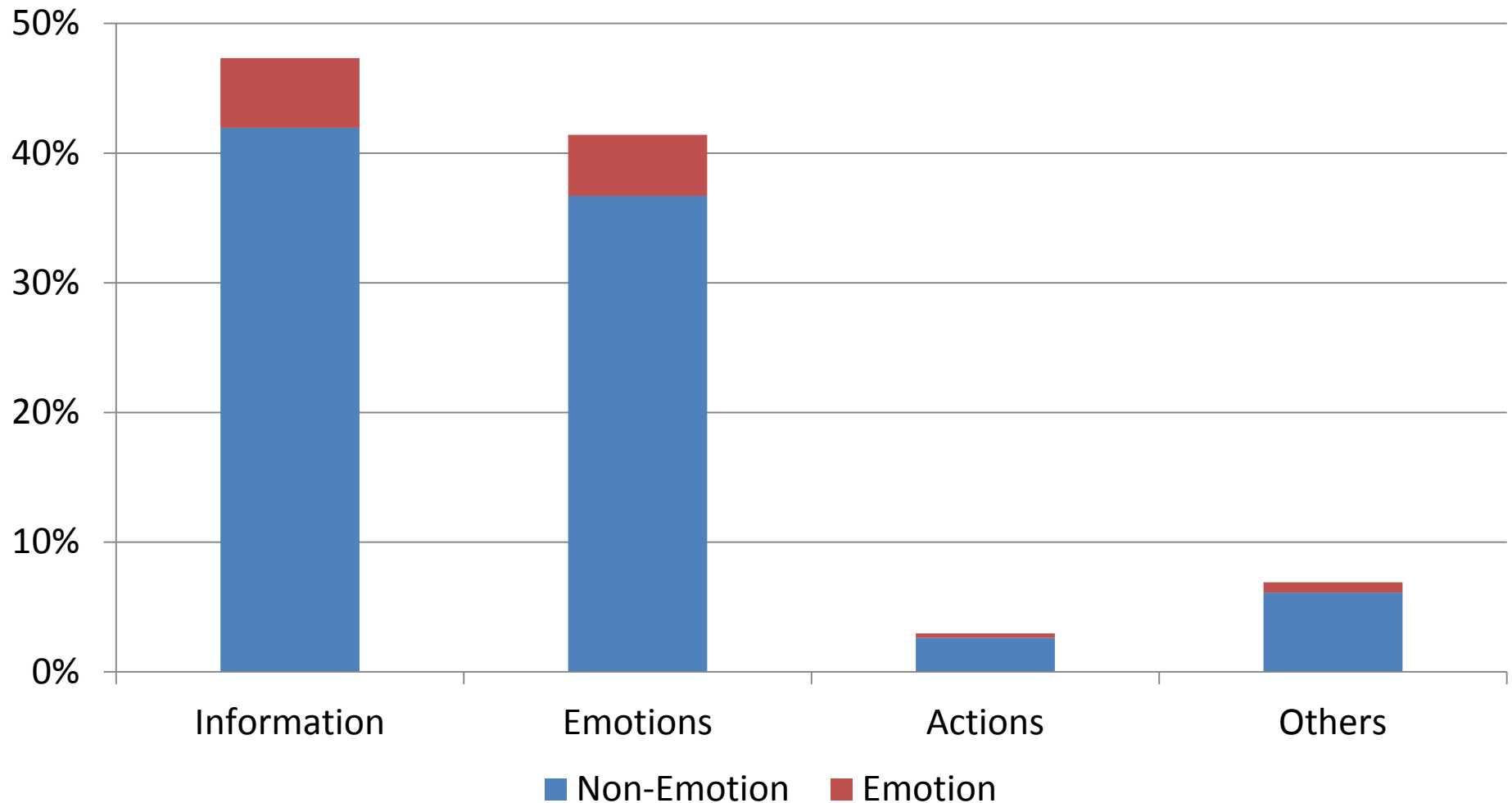
Arguments: Weibo helps rural migration workers to

- mobilize to improve their life in cities;
- release emotional dissatisfactions with reality;
- influenced by journalists, scholars, and officials (unreciprocal weak ties?).

Method:

- Searched “new generation of rural workers” (新生代农民工) to obtain 4,000+ posts;
- Content analyzed the themes of the posts;
- Results considered to support/derive the arguments.

Evidence for Arguments 1-2 (?)



Agenda-setting Effects based on N of Posts

Hypotheses:

- “Safety valve” effects: bloggers express discontent on issues initiated by the media;
- “Pressure cooker” effects: bloggers lead issues to increase social tensions.

Method:

- Randomly searched through Google 2,000+ blog posts and 4,000+ newspaper stories;
- Computerized content analysis of issue themes;
- Time series analysis to determine causal directions between the two agendas.

Safety Valve or Pressure Cooker?

Findings:

Rather than simply being a “safety valve” that reduces bloggers’ anger over political or social problems, or a pressure cooker that increases these tensions, the results of this study suggested a more nuanced perspective in which the Chinese blogosphere can act as either, depending on the topic.

- Safety value issues:
 - Energy;
 - Politics;
 - Corruption and illegality;
 - Sino-Japan relations
- Pressure cooker issues:
 - Internet;
 - Religion;
 - Rural;
 - Arts;
 - Disasters.

Summary of Empirical Evidence

What we know enough:

- Massive size of users
- Content censored or manipulated
- Active and influential opinion leaders

What we know little:

- Does the size matter?
- Is the censorship effective?
- Causality: Does social media bring social changes?

Looking into the Future

- The future lies in close collaboration between scientists (with computer scientists in particular) and social scientists.
- While social scientists know what to study (e.g., effects of social media) but don't know how to do it, computer scientists have the tools but don't know where to apply.
- The emerging “Computational Social Science” (Watts, 2014) represents a new paradigm of interdisciplinary collaboration between computing and social science.

Cited Works

1. Chen, C. (2013). Opinion leaders: *The driving force of political discussion in social media*. Master's thesis, Middle Tennessee State University.
2. Hassid, J. (2012). Safety valve or pressure cooker? Blogs in Chinese political life. *Journal of Communication*, 62, 212-230.
3. King, G., Pan, J., & Roberts, M. E. (2013). Reverse engineering Chinese censorship through randomized experimentation and participant observation. Presented at the annual meeting of American Political Science Association, Chicago.
4. Lu, H. (2013). *The sustainability of UGC: Understanding continuance and switching produsage Behavior*. Ph.D. Dissertation, City University of Hong Kong.
5. Tse, T. S. M., & Zhang, E. Y. L. (2013) Analysis of blogs and microblogs: A case study of Chinese bloggers sharing their Hong Kong travel experiences. *Asia Pacific Journal of Tourism Research*, 18(4), 314-329.
6. Wu, A.X. (In press). Ideological polarization over a China-as-superpower mindset: An exploratory charting of belief systems among Chinese Internet users, 2008-2011. *International Journal of Communication*.
7. Xie, Y. H. (2010, in Chinese). Structure and mechanism of interactions on Sina Weibo: An empirical study. *Journalism & Communication Research (Xinwen yu Chuanbo Yanjiu)*, 4, 60-69.
8. Yang, G. B. (2010). *The power of the Internet in China: Citizen activism online*. Columbia University Press.

Cited Works (cont'd)

9. Yin, L. G. (2010, in Chinese). Framing the Deng Yujiao case: How online opinion impacts offline media reports. *International Journalism (Guoji Xinwenjia)*, 2010(9), 25-31.
10. Yu, L., Asur, S., & Huberman, B. A. (2011). What trends in Chinese social media. *Proceedings of The fifth ACM workshop on Social Network Mining and Analysis*. ACM.
11. Yu, L. L., Asur, S., & Huberman, B. A. (2012). Artificial inflation: The real story of trends and trend-setters in Sina Weibo. ASE/IEEE International Conference on Social Computing.
12. Zhang, P. Y. (2013). Social inclusion or exclusion? When Weibo (Microblogging) meets the “New Generation” of rural migrant workers. *Library Trends*, 62(1), 63-80.
13. Zhang, L. (In press). Social media in Chinese government: Drivers, challenges and capabilities. *Government Information Quarterly*.
14. Zhou, Y. Q., & Moy, P. (2007). Parsing framing processes: The interplay between online public opinion and media coverage. *Journal of Communication*, 57, 79-98.
15. Zhu, J. J. H., Peng, T. Q., Liang, H., Wang, C. J., Qin, J., & Chen, H. X. (2014, in Chinese). Computational social science in communication research. *e-Science Technology and Applications (Keyan Xinxihua Jishu he Yingyong)*, 5(2), 3-13.

THANK YOU & CONTACT US @

[mailto: j.zhu@cityu.edu.hk](mailto:j.zhu@cityu.edu.hk)

<http://weblab.com.cityu.edu.hk>



Web Mining Lab
互聯網挖掘實驗室