

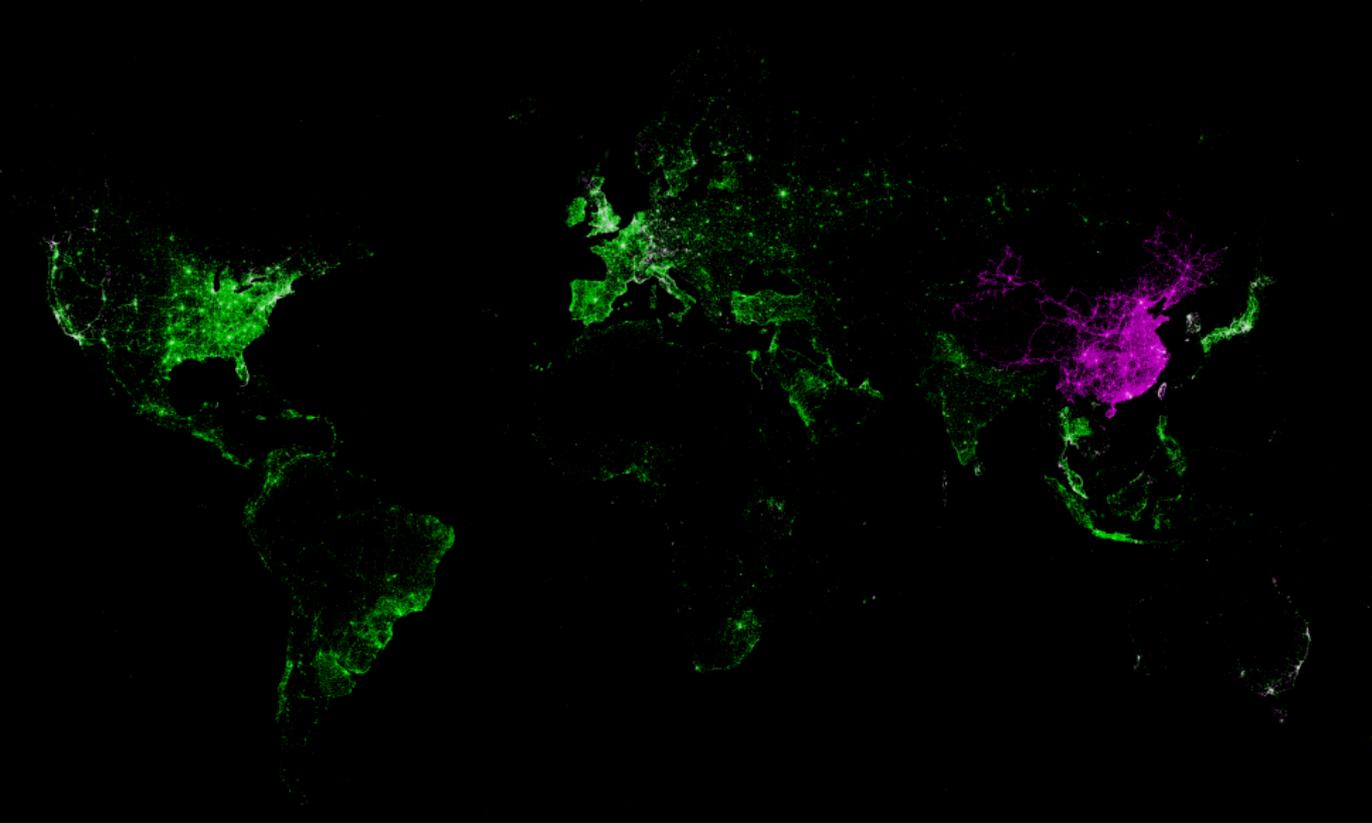
# Weibo, and a Tale of Two Worlds

Wentao Han<sup>1</sup>, Xiaowei Zhu<sup>1</sup>, Ziyang Zhu<sup>1</sup>, Wenguang Chen<sup>1</sup>, Weimin Zheng<sup>1</sup>,  
Jianguo Lu<sup>2</sup>

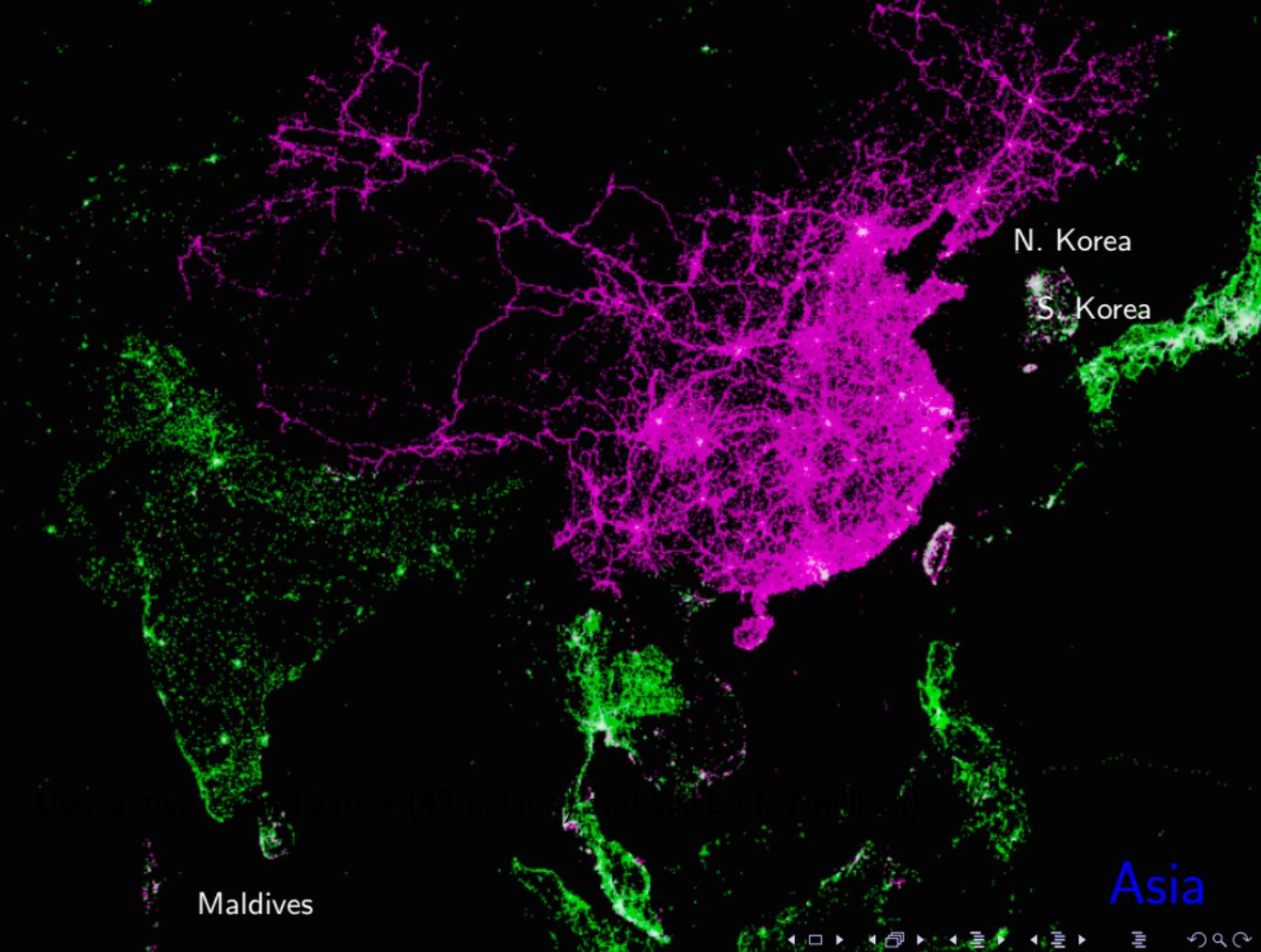
<sup>1</sup>Tsinghua University, China

<sup>2</sup>University of Windsor, Canada

# What is Weibo?



User activities of Twitter (42 million) and Weibo (37 million).



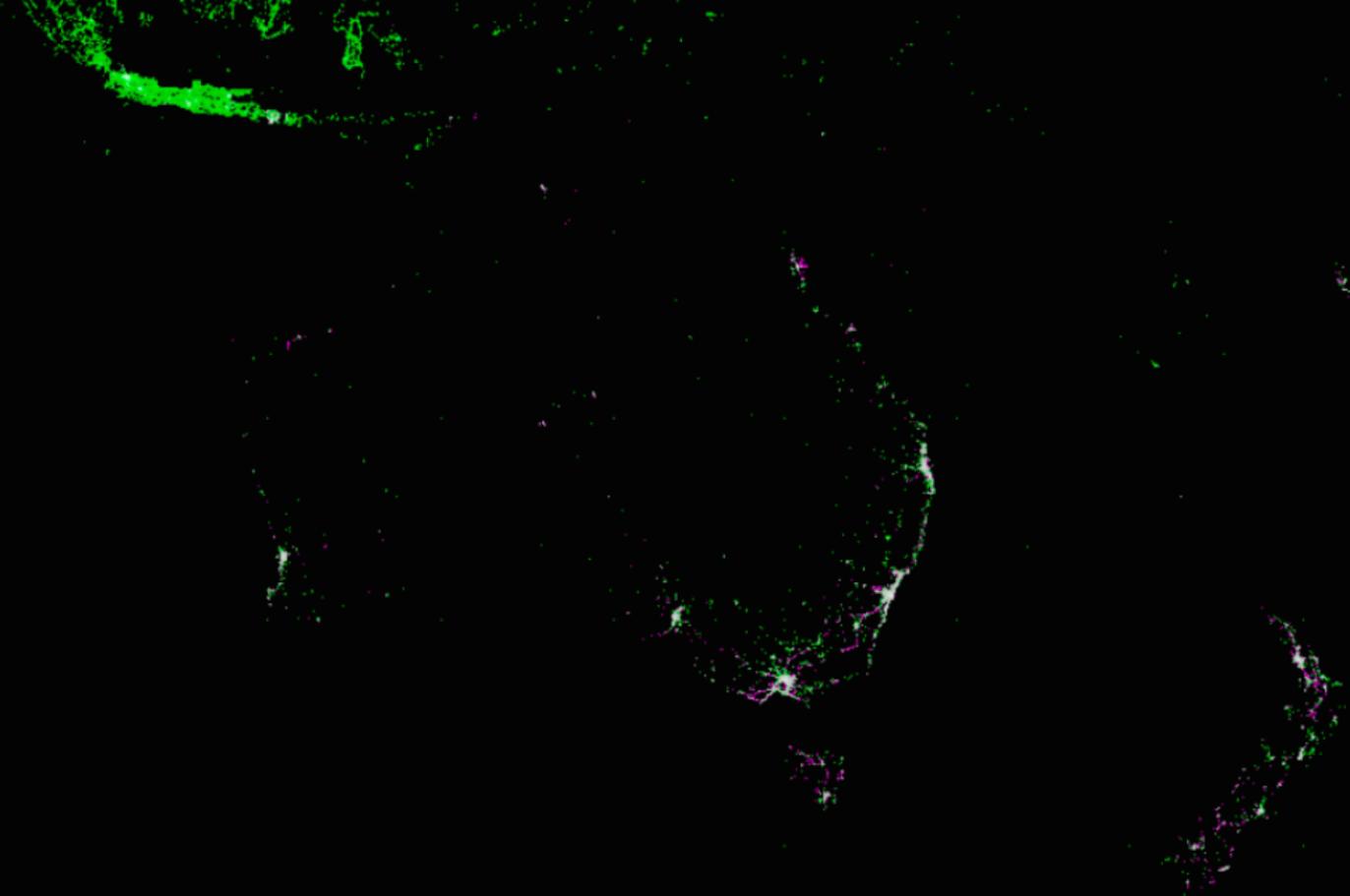
N. Korea

S. Korea

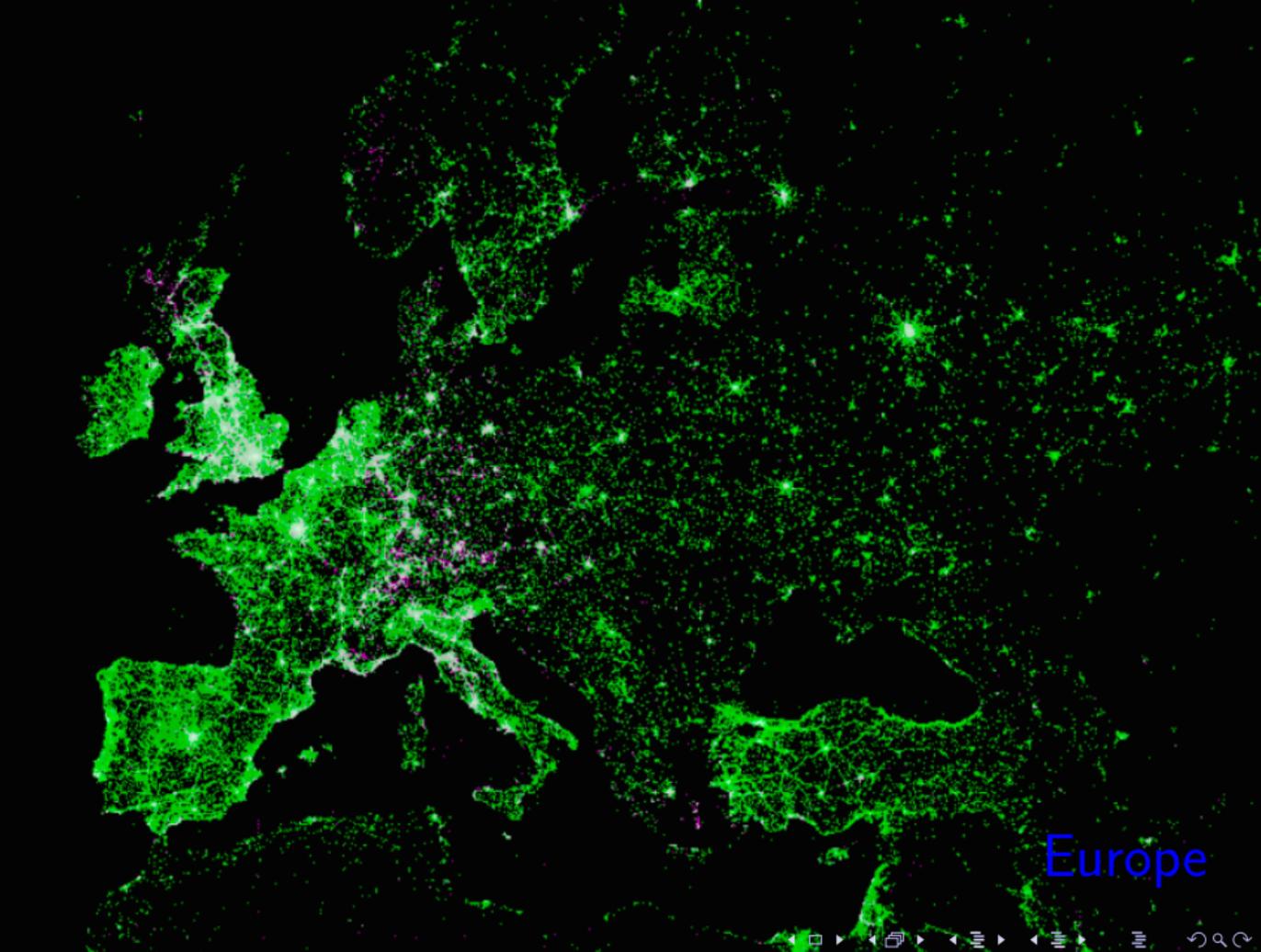
Maldives

Asia

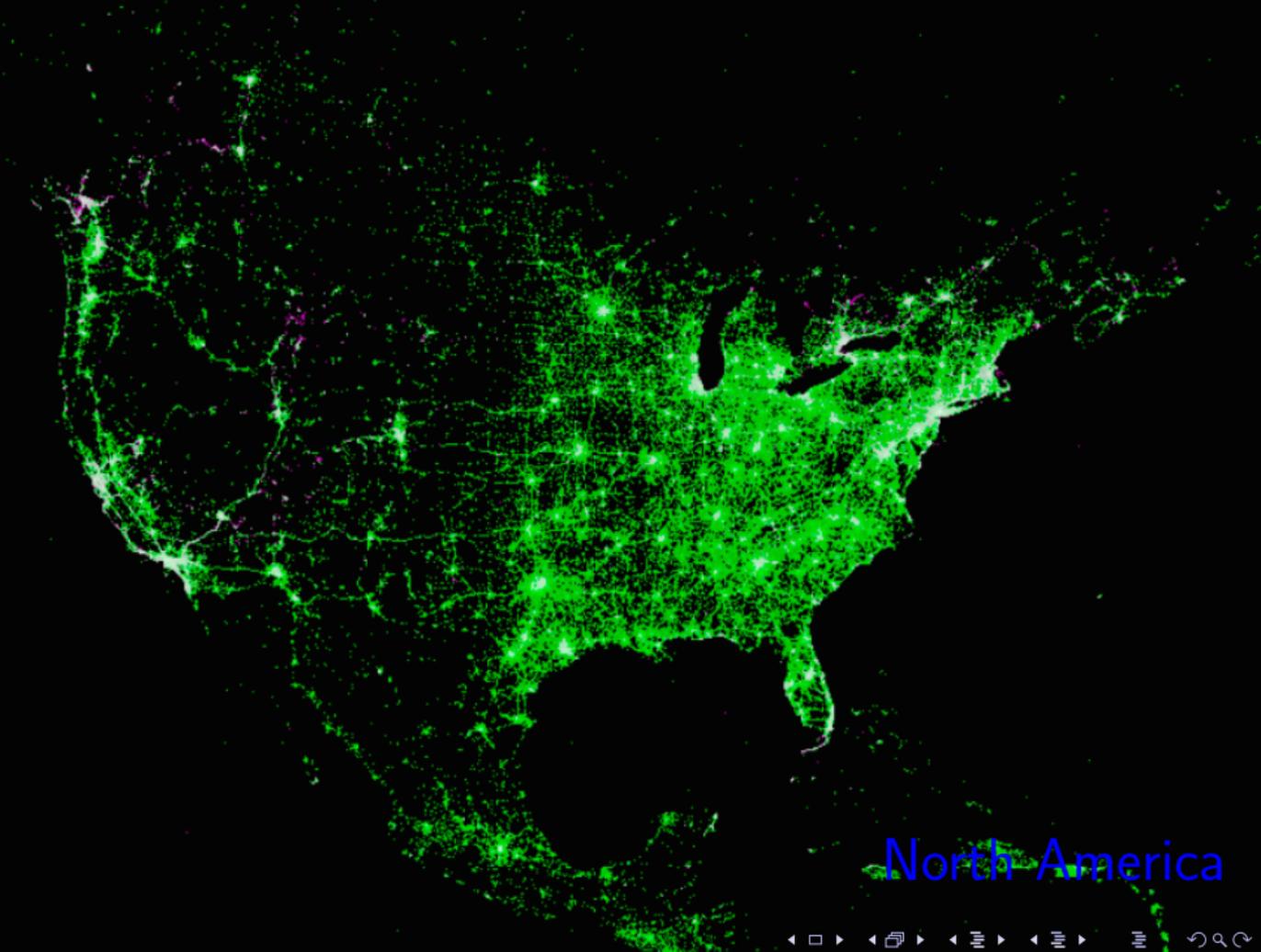




Australia



Europe



North America

# Weibo and Twitter

## Weibo and Twitter platforms

	<b>Weibo</b>	<b>Twitter</b>
Message length	140	140
Language	Chinese (mostly)	Many
Directed network	yes	yes
2000 out-link limit	yes	yes (with exceptions)

- Weibo has hundreds of millions of (registered/active) users.
- Mostly disjoint from Twitter users geographically
- Not well studied
  - Only a few studies based on small samples
  - Compare: two complete Twitter networks were studied (the 2009 network and 2014 network)

# Goals

- Whether are the two worlds similar?
  - Do the people in the two worlds interact in a similar way?
- Whether the world represented by Weibo can reflect the real world?

# Experiment

- We crawled an almost *complete* user network
  - exhaustively follow the out-links;
  - collect users who has at least one in-link
  - between November 2012 - February 2013
- 282 million nodes
- 27 billion links

## Why is it a complete network?

- Few new users could be crawled (689 duplicates to obtain one new user)
- Consistent with the official data on the number of registered user
  - Official registered user size is 500 million;
  - We carried out a uniform random sampling, and find that 40% of the users have no in-links;
  - 282 million are all the users who have in-links

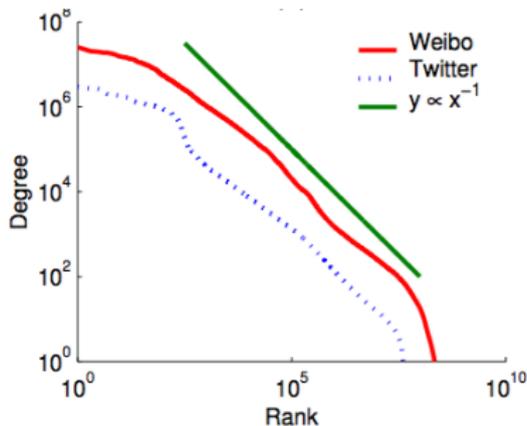
# Twitter 2009 data

- Publicly available
- Both are networks evolved after three years of inception
- H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In WWW, pages 591 - 600. ACM, 2010.

	#Nodes ( $\times 10^6$ )	#Links ( $\times 10^9$ )	Mean degree	Max ( $\times 10^6$ )	Std	CV	Simpson ( $\times 10^{-4}$ )	Gini
Weibo	222	27	121	25	9028	74	0.25	0.88
Twitter	41	1.4	35	2.9	2419	69	1.13	0.83

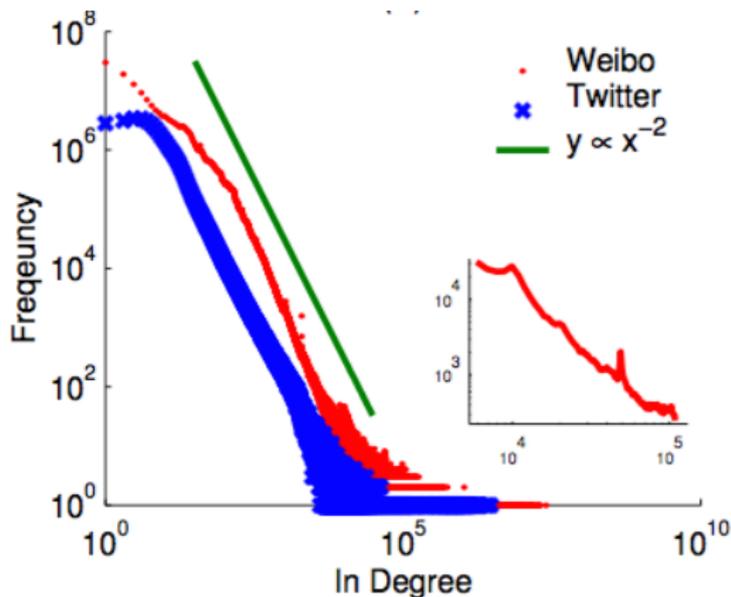
**Table:** Statistics of Weibo and Twitter 2009 data. Max, standard deviation (Std), coefficient of variation (CV), Simpson index and Gini index are for in-degrees.

# In-degree distribution



- In-degrees *resemble* a power law
- Exponent is around -1
- Degree vs. Rank plot
- This degree/rank plot is good for viewing the top bloggers
- Not clear for bloggers with small degrees

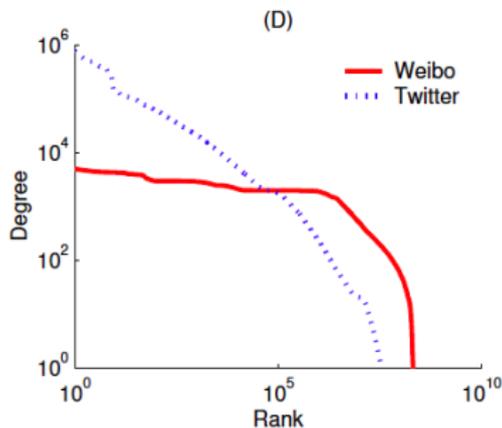
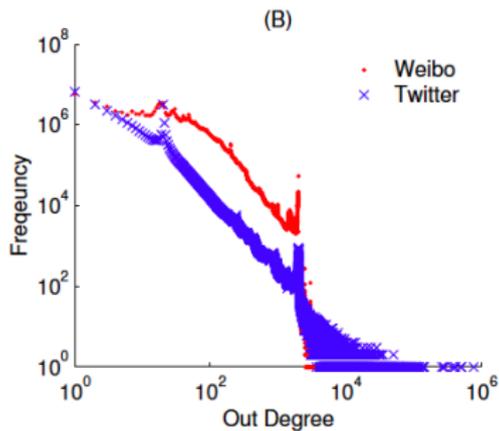
# In-degree distribution (freq. vs. degree plot)



- For frequency/degree plot, the exponent is around -2 (plus -1 of the previous plot)
- Good to view the bloggers with small degrees
- Spammers buy in-links in bulks of 10K and 50K.

# Out-degrees

- There are spikes around 2000 out-degree limit, for both Weibo and Twitter
- Weibo impose more rigid 2000 limit



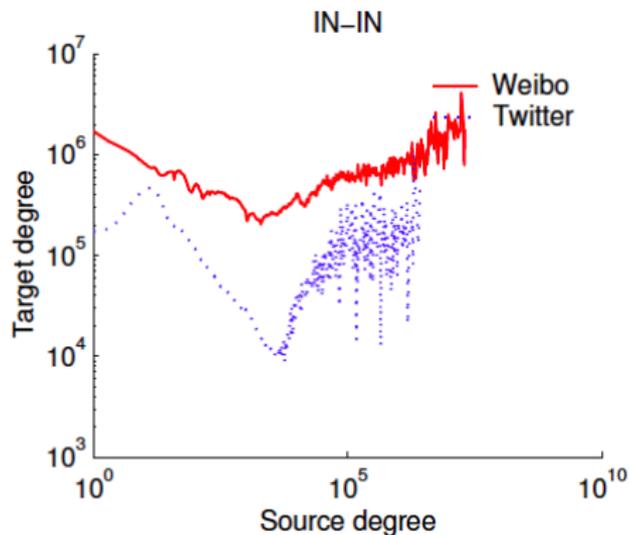
# Degrees can tell more information ...

- Your friends are richer than you are
  - Your friends have more friends than you do
  - Weibo: 5567 times more
  - Twitter: 4761 times more
  - This is the coefficient of degree variation
- Diversity index: the probability of two links pointing to the same user
  - Weibo:  $0.25 \times 10^{-4}$
  - Twitter:  $10^{-4}$
- Gini coefficient, inequality of the in-degrees
  - Weibo: 0.88
  - Twitter 0.83
  - Higher than the inequality of wealth in any economy

# Celebrities socialize with celebrities?

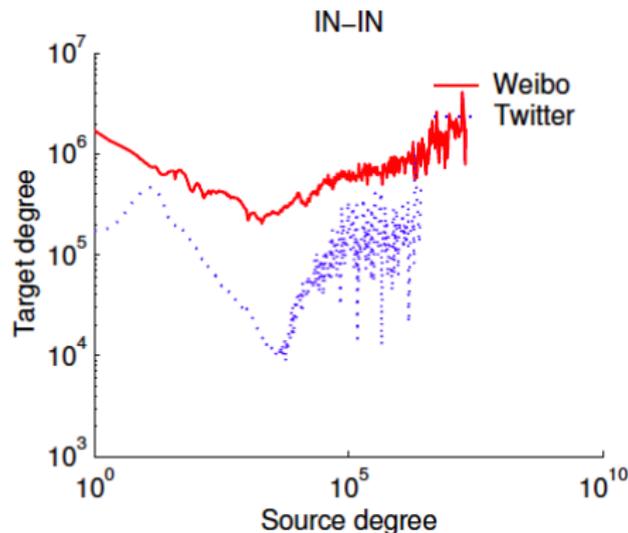
- Intuitively, yes.
- Pearson correlation coefficient between the in-degrees of the following relationship is *negative*.
  - A tempting conclusion is that popular users tend to follow unpopular users
  - Surprising result claimed by Myers et al.
    - S. A. Myers, A. Sharma, P. Gupta, and J. Lin. Information network or social network?: the structure of the twitter follow graph. In WWW, pages 493 - 498, 2014.

# Celebrities do tend to follow celebrities



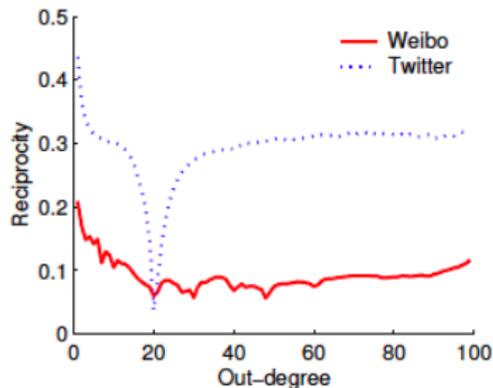
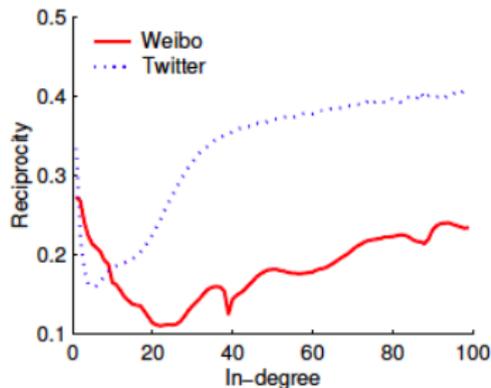
- When the relation is not linear, Pearson correlation coefficient is not enough
- 'rich' follows 'rich', but 'poor' also follows 'rich'

# Possible explanation



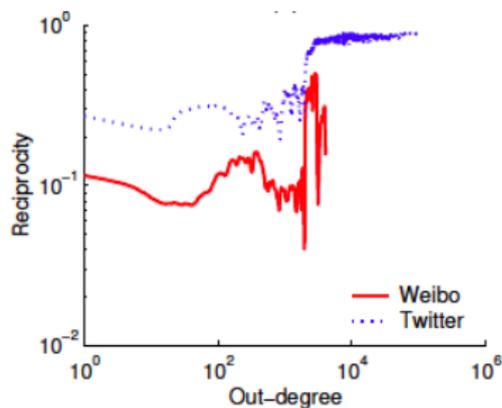
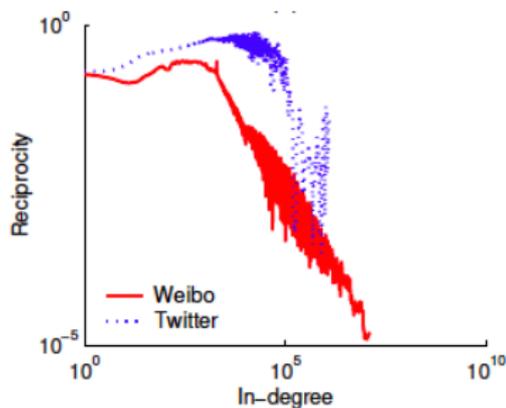
- Users start by following celebrities (high target degree)
- then normal users (target degree drops)
- Celebrities do not follow many normal users
- Why both Weibo and Twitter bottom at 2000?

# Reciprocity



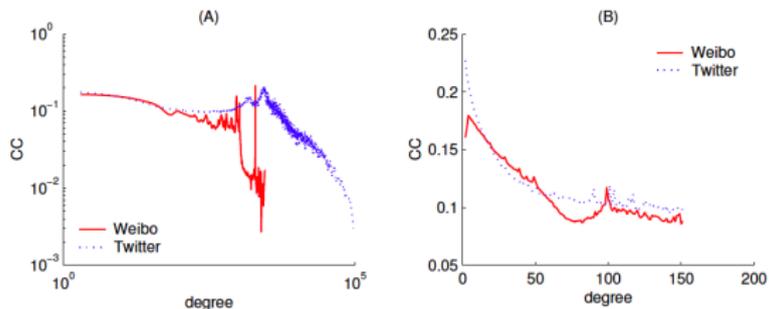
Reciprocity for most users who have less than 100 in-/out-degrees.  
Weibo is much lower than Twitter

# Reciprocity



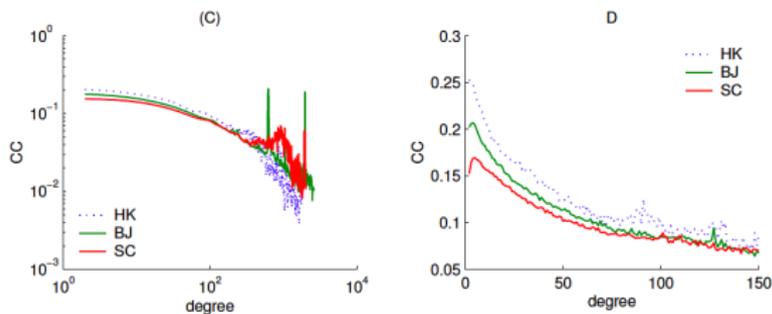
- Overall view of reciprocity for all kinds of people.
- Twitter has a much higher reciprocity;
- Their reciprocity fluctuates in a similar pattern;
- Privileged Twitter users ( $out - degree > 2000$ ) reciprocate almost every follower

# clustering coefficient



CC is slightly higher in Twitter

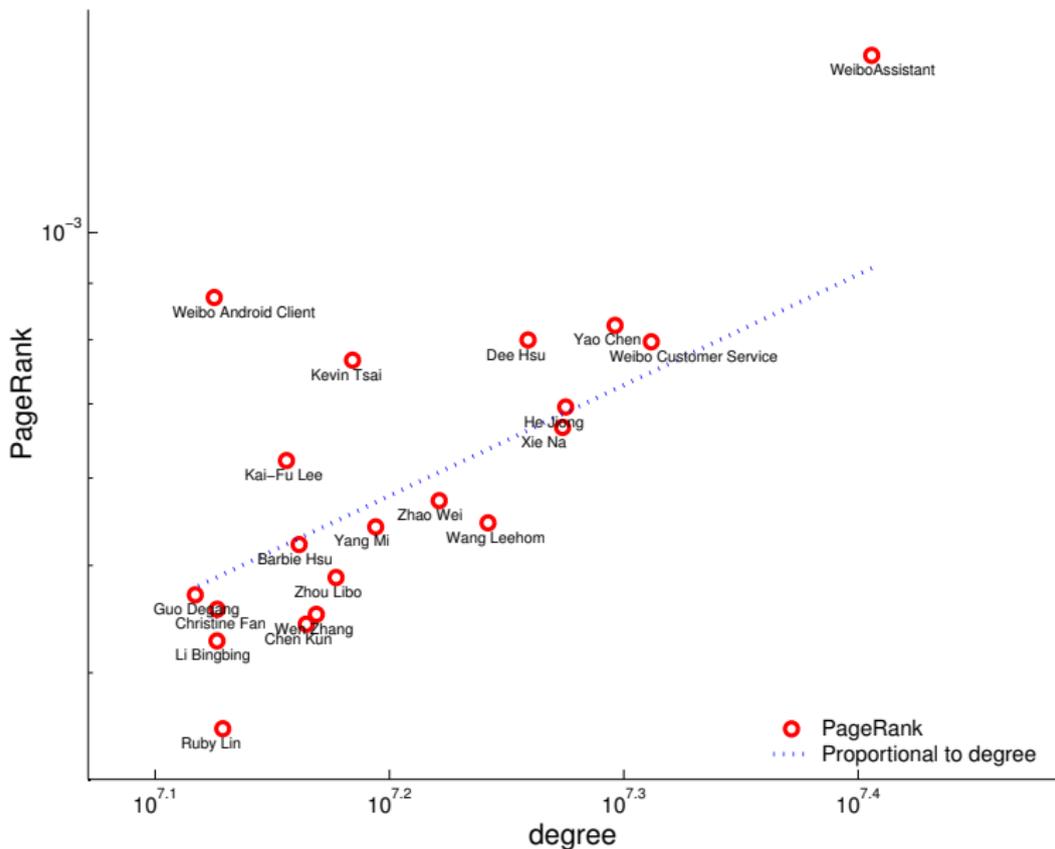
# clustering coefficient



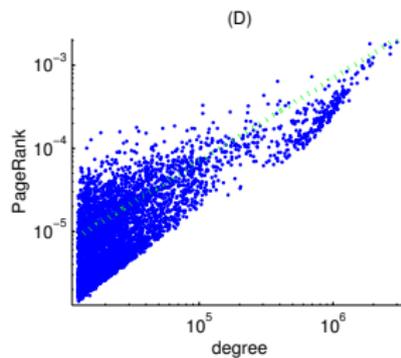
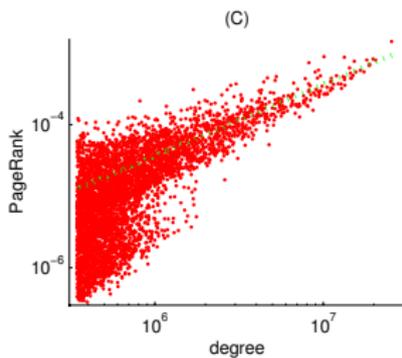
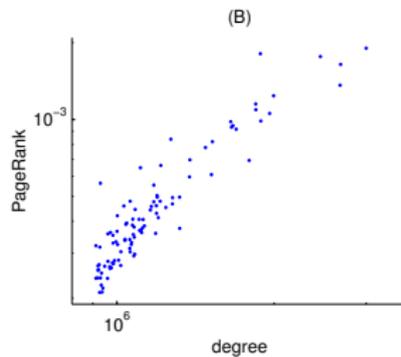
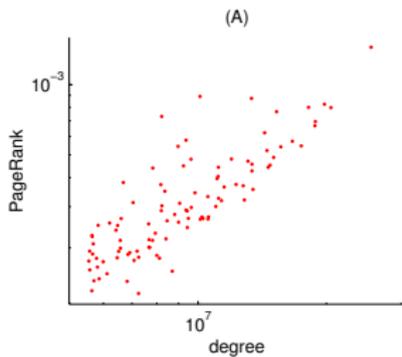
## Hypothesis

CC is higher for more economically developed regions

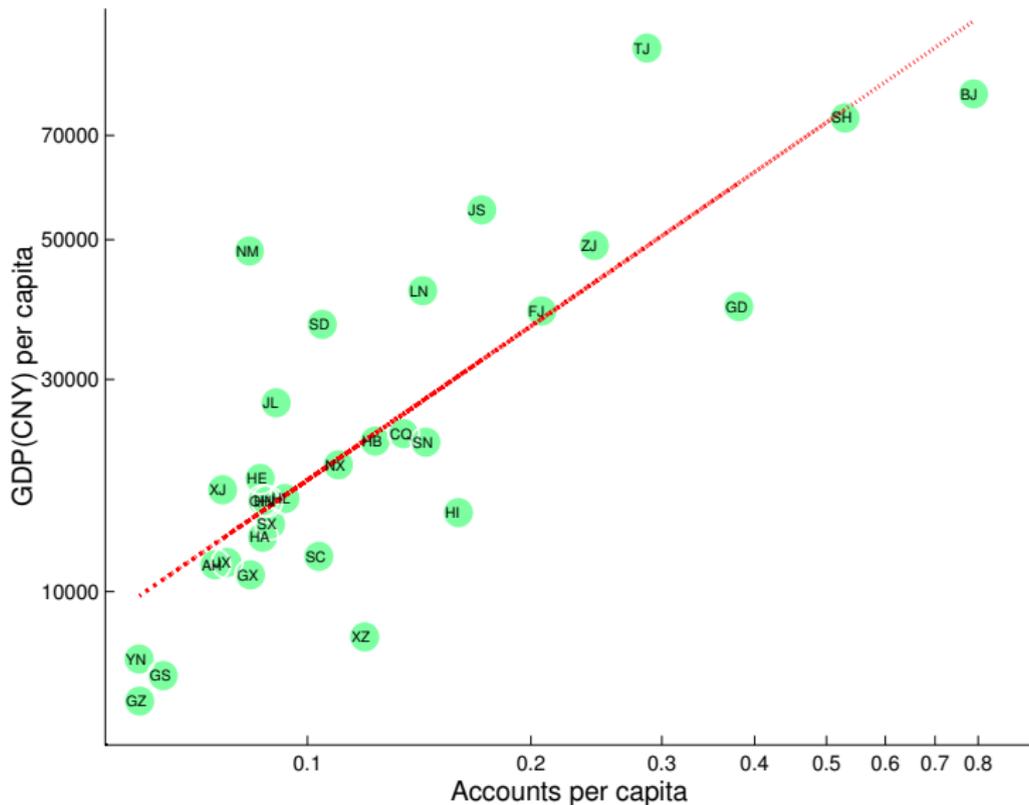
## top bloggers



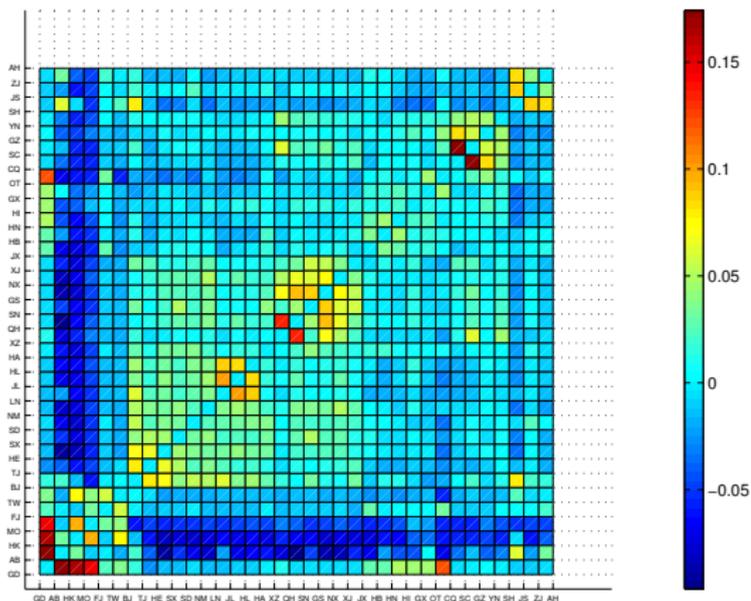
# pagerank vs. in-degree



## weibo adoption rate

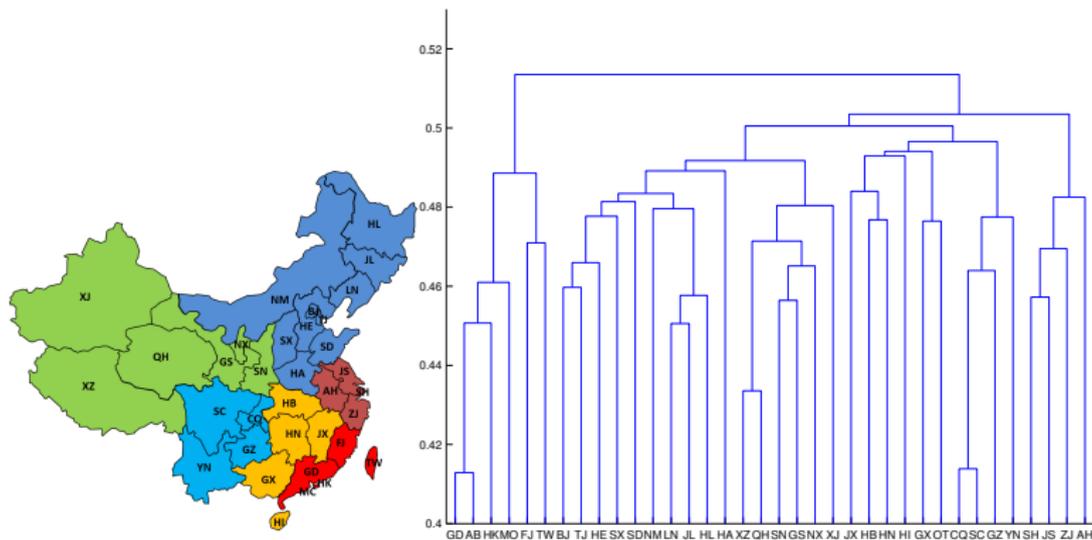


# Connections between regions



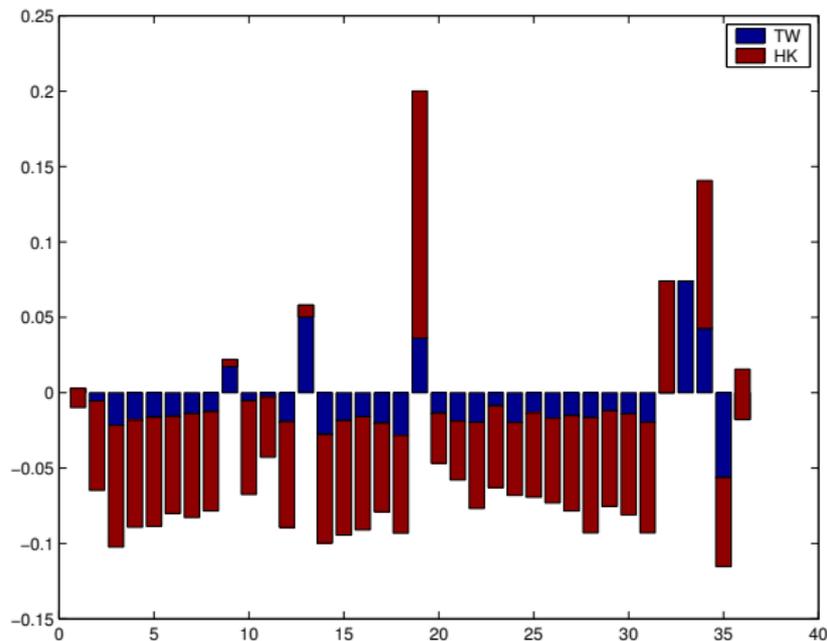
closest pairs: ChongQing vs SiChuan, Tibet vs QingHai, HongKong vs Macau vs GuangDong

# Connections between regions



result of running hierarchical agglomerative clustering

## mutual information of HK and TW



normalized point-wise MI: zero means the strength is the same as in a random graph

# Take away messages

- We have the largest OSN user network available for research;
- Similarities and differences between the two worlds:
  - degree variation is similar
  - reciprocity
- Connection to the real world
  - quantify the connection strength between regions