Analysis of Tweets and Tweeters during the 2014 Umbrella Revolution

Zhouhan Chen[+] (zc12@rice.edu), Yun Zhou[+], Rima Tanash[+], Jieya Wen[+], Dan Wallach[+], Richard Stoll[*], Devika Subramanian[+]

[+]Computer Science Department, Rice University, [*]Political Science Department, Rice University

Introduction

- We collected 2,915,037 tweets about the Hong Kong Revolution from 10/05/2014 to 12/06/2014.
- **Goals**: Understand the dynamics of the revolution using tweets alone.

Challenges

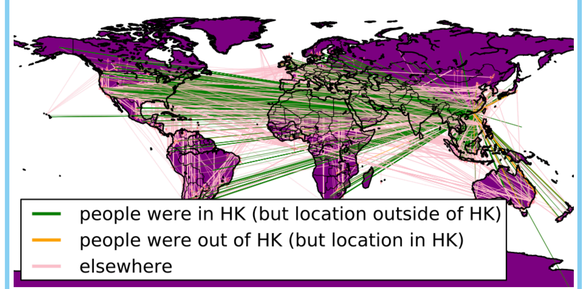
- Millions of raw tweets.
- Many top tweeters are bots, introducing noise in our analysis.
- Relationship between tweeters is complex.

Methods

- Analyze peaks in Twitter volume.
- Use Twitter API to build follower-following and tweet-retweet graphs.
- Detect bot and human groups with community detection algorithms on those graphs.
- Interpret graphs with visualization tools such as Gephi.
- Apply natural language processing (NLP) and machine learning for sentiment analysis on tweets.

Data Collection

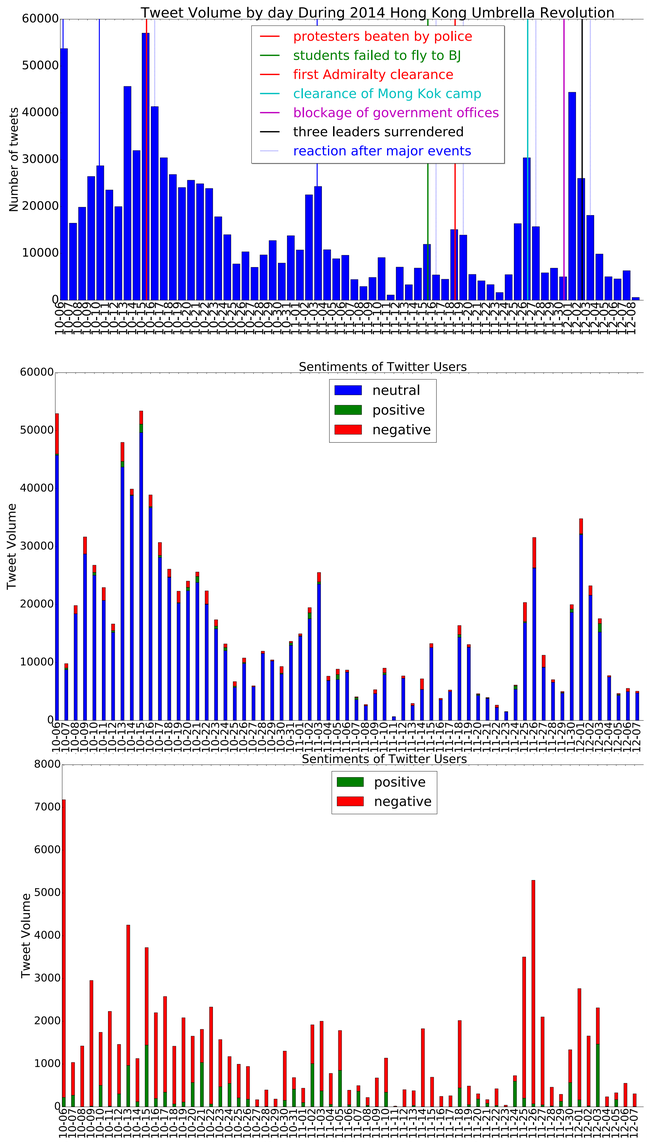
- Tweets collected from: Twitter Streaming API
- Number of protest-related tweets: 1,062,606
- Number of geo-enabled tweets: 3583



According to Twitter, ***location*** is "*The user-defined location for this account's profile*". The map shows that many people from all over the world were physically present in Hong Kong during the protest.

Analysis of Tweets: Events and Sentiment

- What are tweets about?
- What is the relationship between tweet volume and major events?
- How do people feel about the events?

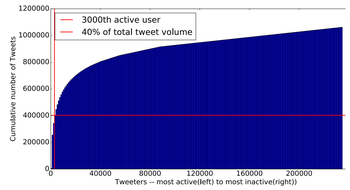


Example positive tweet*: "Something about this just gives me a little more faith in the world. Hong Kong protesters are so freaking nice."*

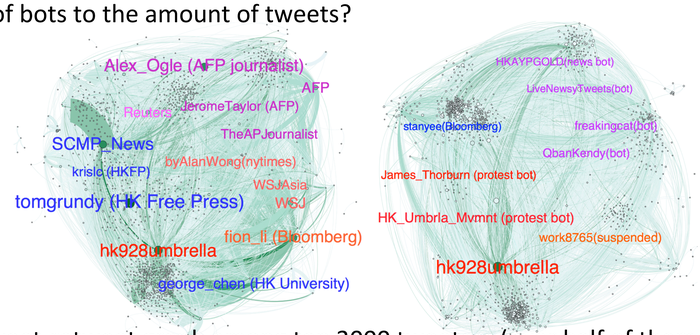
Example negative tweet*: "Everytime I hear about the Umbrella Revolution I want to groan for a few years."*

Analysis of Tweeters: Bots and Humans

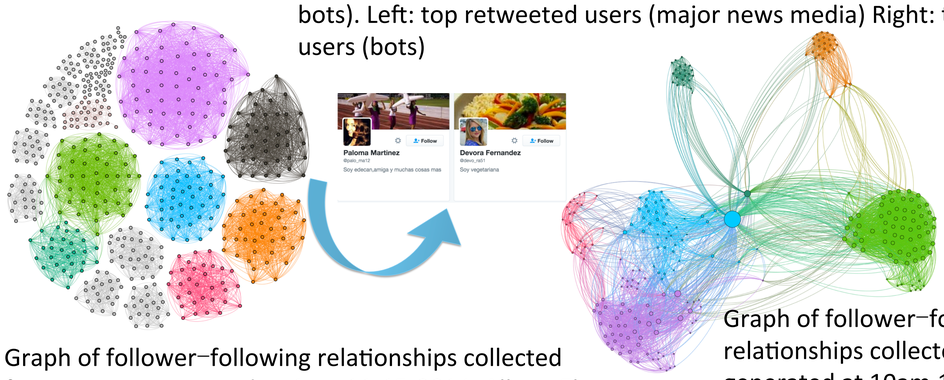
- Who are the top tweeters?
- What is the relationship between the top tweeters?
- What is the contribution of bots to the amount of tweets?



Top 3000 tweeters generate 40% of tweets

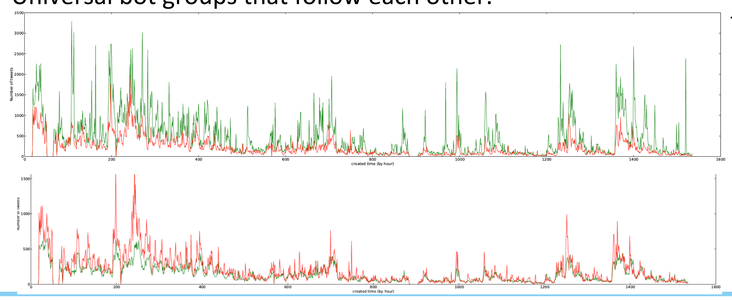


Tweet-retweet graph among top 3000 tweeters (over half of them are bots). Left: top retweeted users (major news media) Right: top retweeting users (bots)



Graph of follower–following relationships collected from tweets generated at 4pm,11-15-2014. All are El Universal bot groups that follow each other.

Graph of follower–following relationships collected from tweets generated at 10am,10-23-2014. All are Falcons News bot groups that follow each other.



Tweet-retweet volume before and after bot removal. Note the significant reduction in noise peaks.

Conclusions

1. 90% - 95% of tweets are about news events or are retweets of news events. Only 5% - 10% express opinions about the events.
2. Top 3000 tweeters (out of 240,000 users) generate 40% of tweets. Over half of them are bots. Bot groups have interesting internal structures.
3. Sentiment analysis of tweets reveals the lack of sustained positive feelings about the revolution.

References

Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2012. Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?. *IEEE Trans. Dependable Secur. Comput.* 9, 6 (November 2012), 811-824.

Zhao, Siqi, et al. "Human as real-time sensors of social and physical events: A case study of twitter and sports games." *arXiv preprint arXiv:1106.4300*(2011

Stockmann, Daniela, Big Data from China and Its Implication for the Study of the Chinese State -- A Research Report on the 2014 Hongkong Protests on Weibo (May 26, 2015).