Epidemiological Modeling of News spreading on Twitter

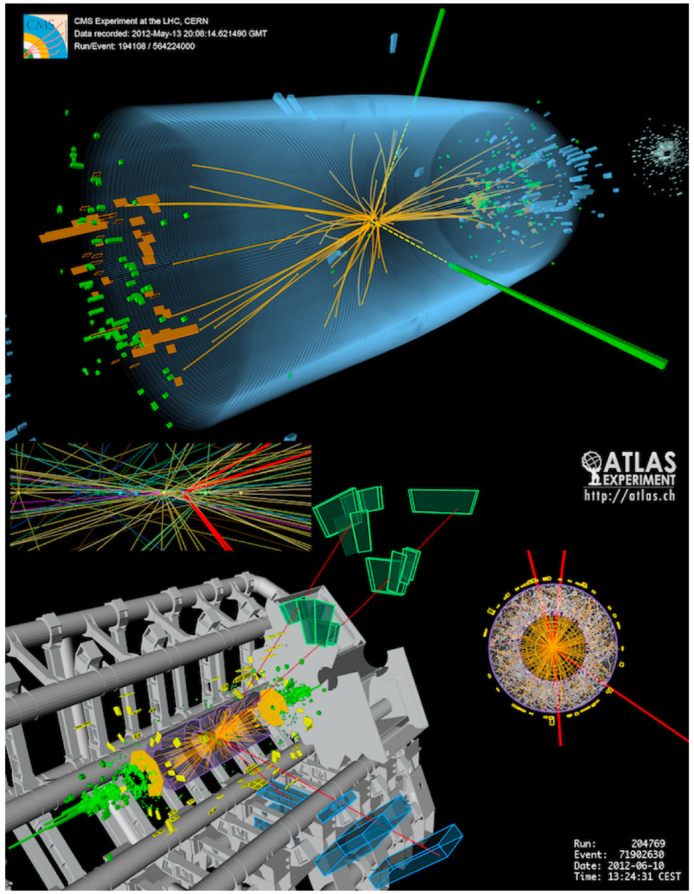
Yunran Chen, Junwen Huang

Department of Statistical Science, Duke University

Content

- Introduction
  - Dataset: Higgs Boson
  - EDA
- Modeling (Compartmental Models in Epidemiology)
  - SI Model
  - SIS Model
  - SEIZ Model
- Conclusion
- Reference

Dataset: Higgs Boson

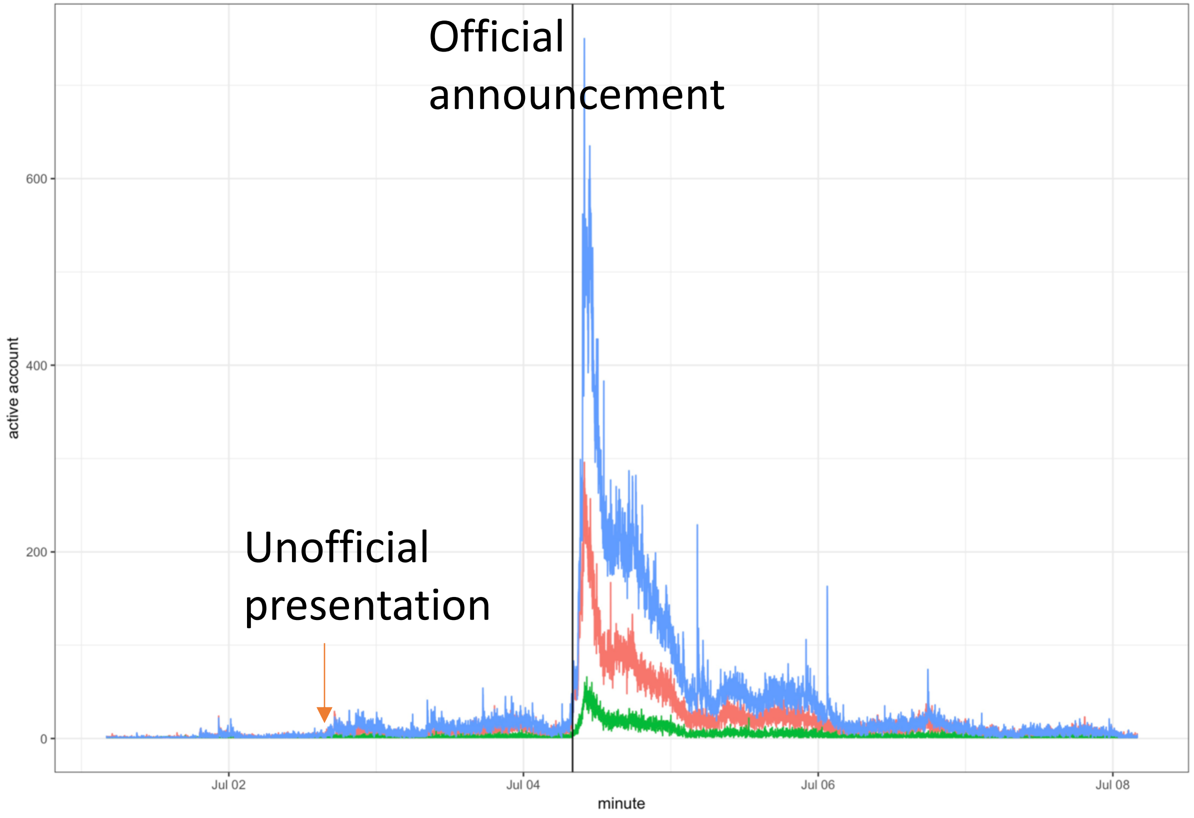


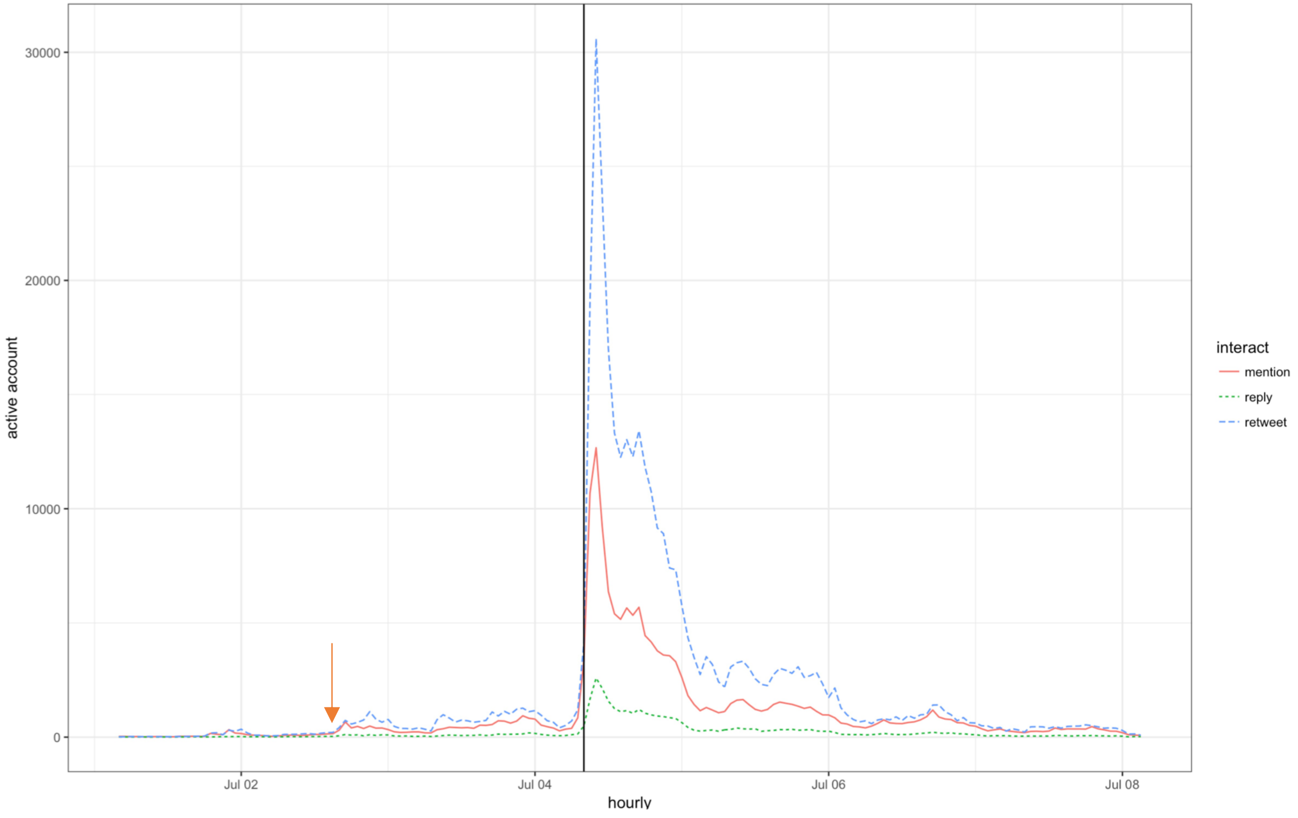
- Higgs Twitter Dataset includes messages posted in Twitter about this discovery between 1st and 7th July 2012 with at least one of the following keywords or hashtags: lhc, cern, boson and higgs.
- The corresponding social network is consist of 456,631 nodes and 14,855,875 directed edges
- The retweet network consists 256,491 nodes and 328,132 edges.

https://en.wikipedia.org/wiki/Higgs_boson#/media/File:Candidate_Higgs_Events_in_ATLAS_and_CMS.png

Data Sources: https://snap.stanford.edu/data/higgs-twitter.html

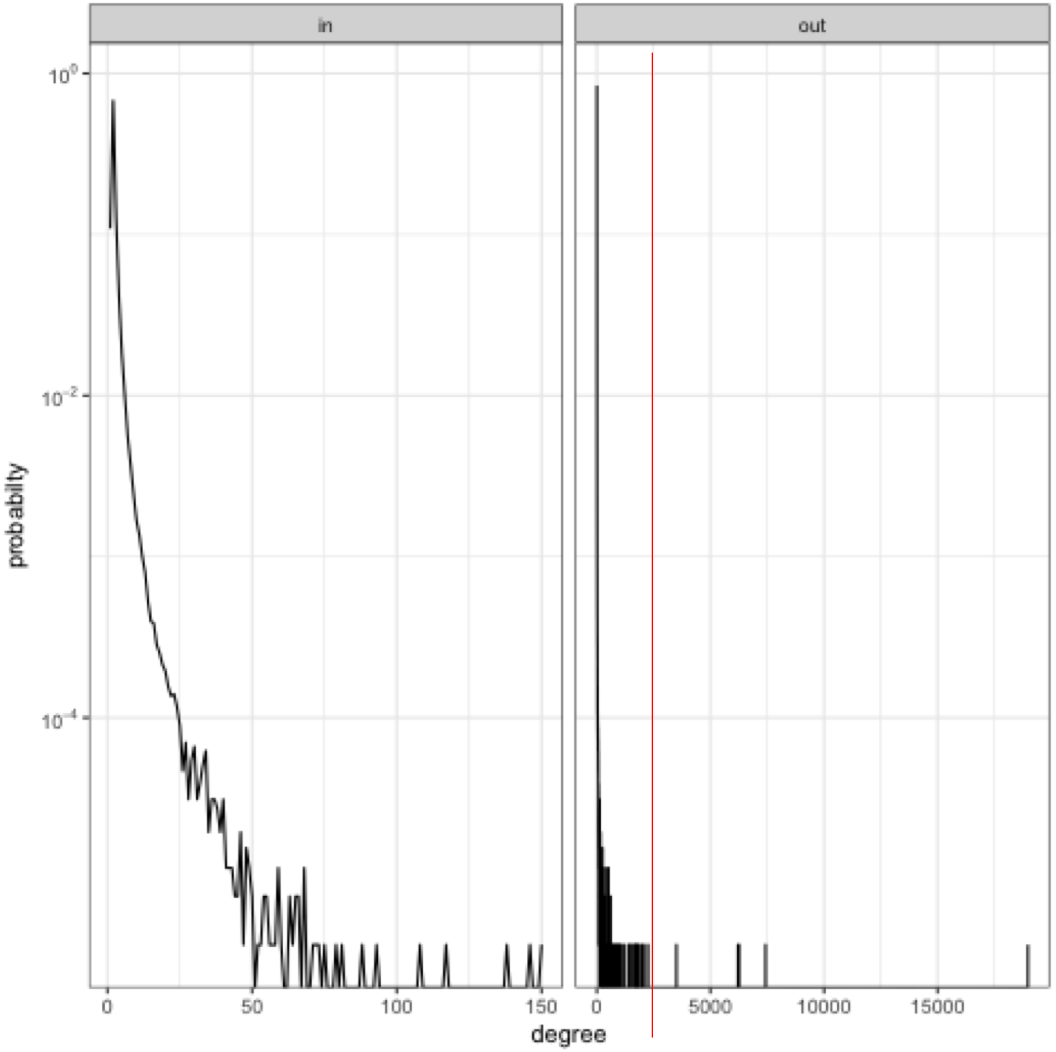
Number of Active Users

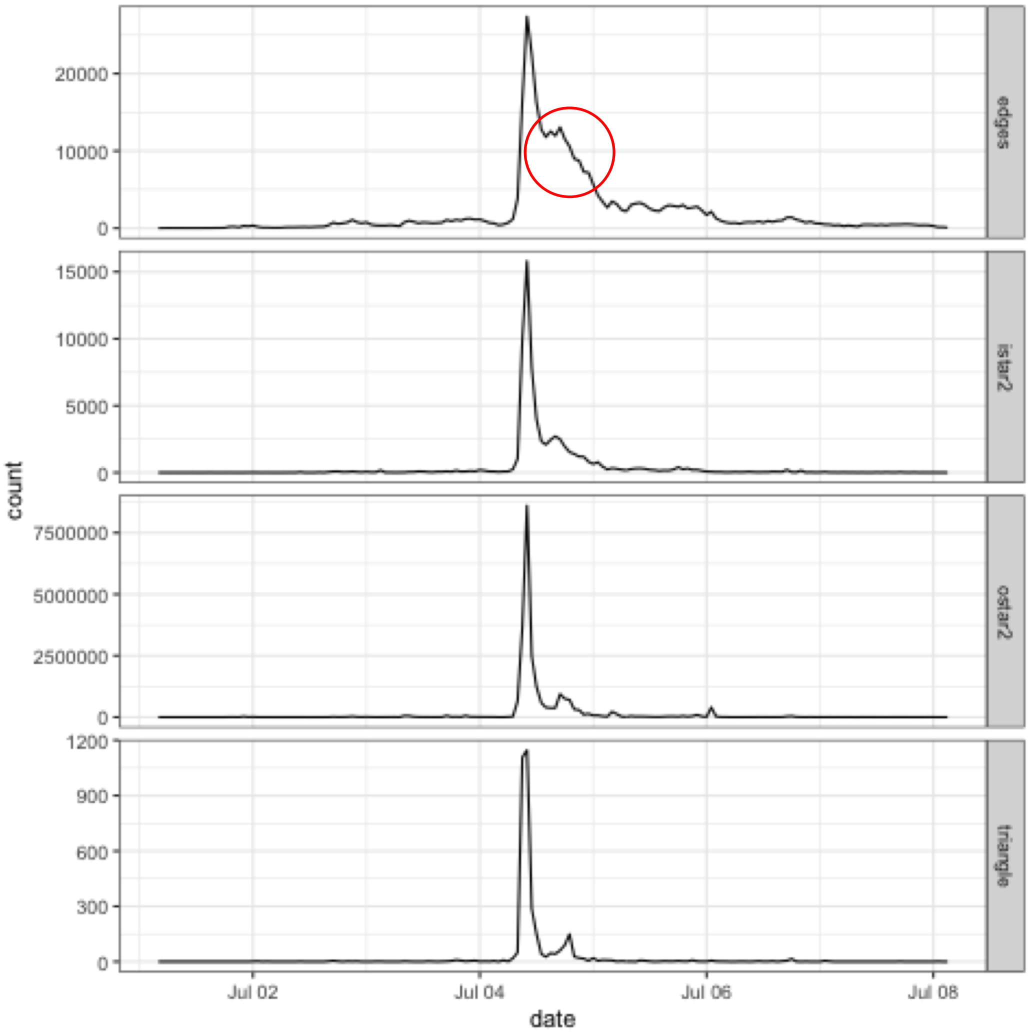




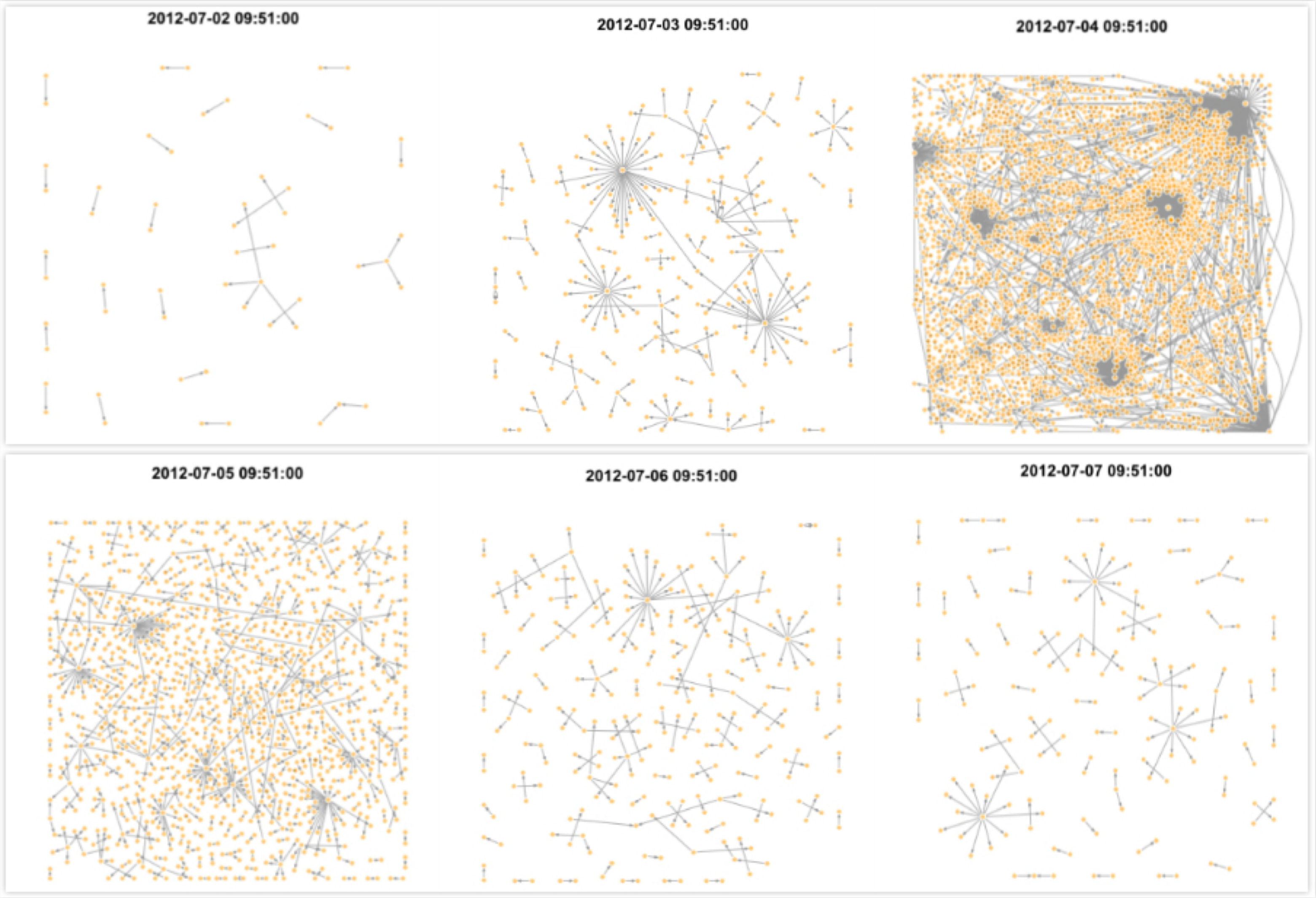
- Retweet network is more fluctuated
- Sharp linear increase
- Exponential decay

Degree Distribution and Other Statistics





Network Structure



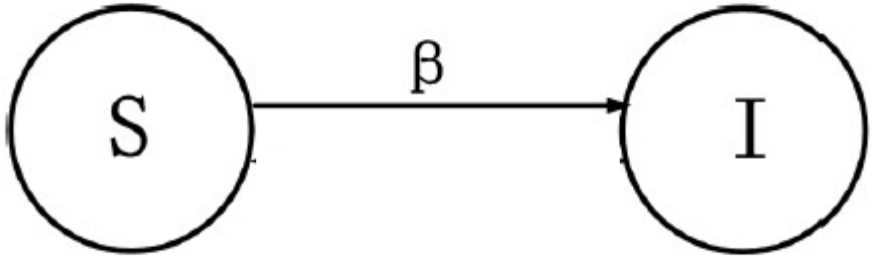
- The attendance and dropout of several large influential hubs matters; 80/20 rule
- Isolated small groups, called “node-to-node” pattern

Compartmental Models in Epidemiology

- Reasons
  - News spreading ≈ Disease spreading
  - Interpretive
- Compartmental Models in Epidemiology
  - Classical approach to study how information diffuses
  - Basic idea is to divide the total population into different compartments
  - Basically ordinary differential equations
- Parameters Estimation
  - Levenberg-Marquardt nonlinear least squares algorithm
  - ODE systems were solved with a forward Euler function.

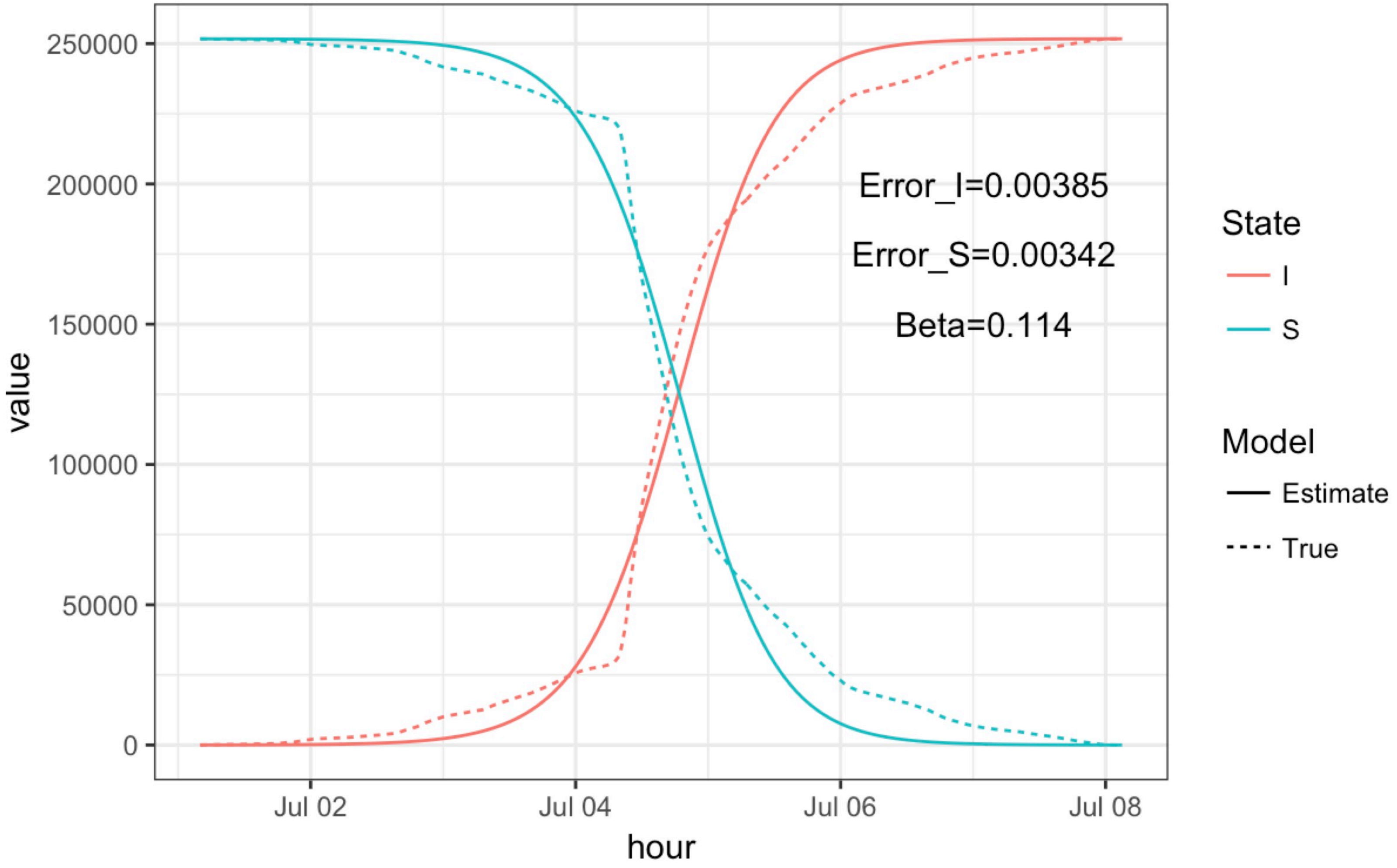
https://en.wikipedia.org/wiki/Euler_method

SI Model: Retweet Network

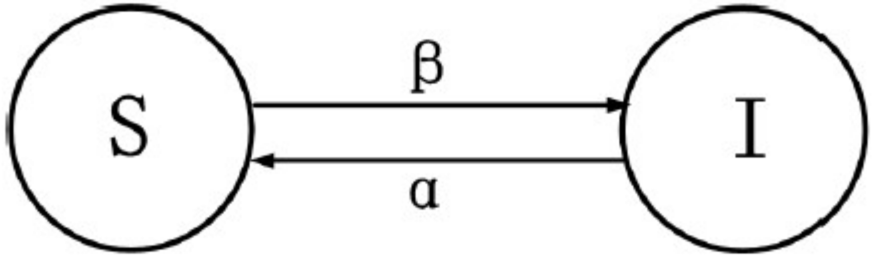


$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N}$$

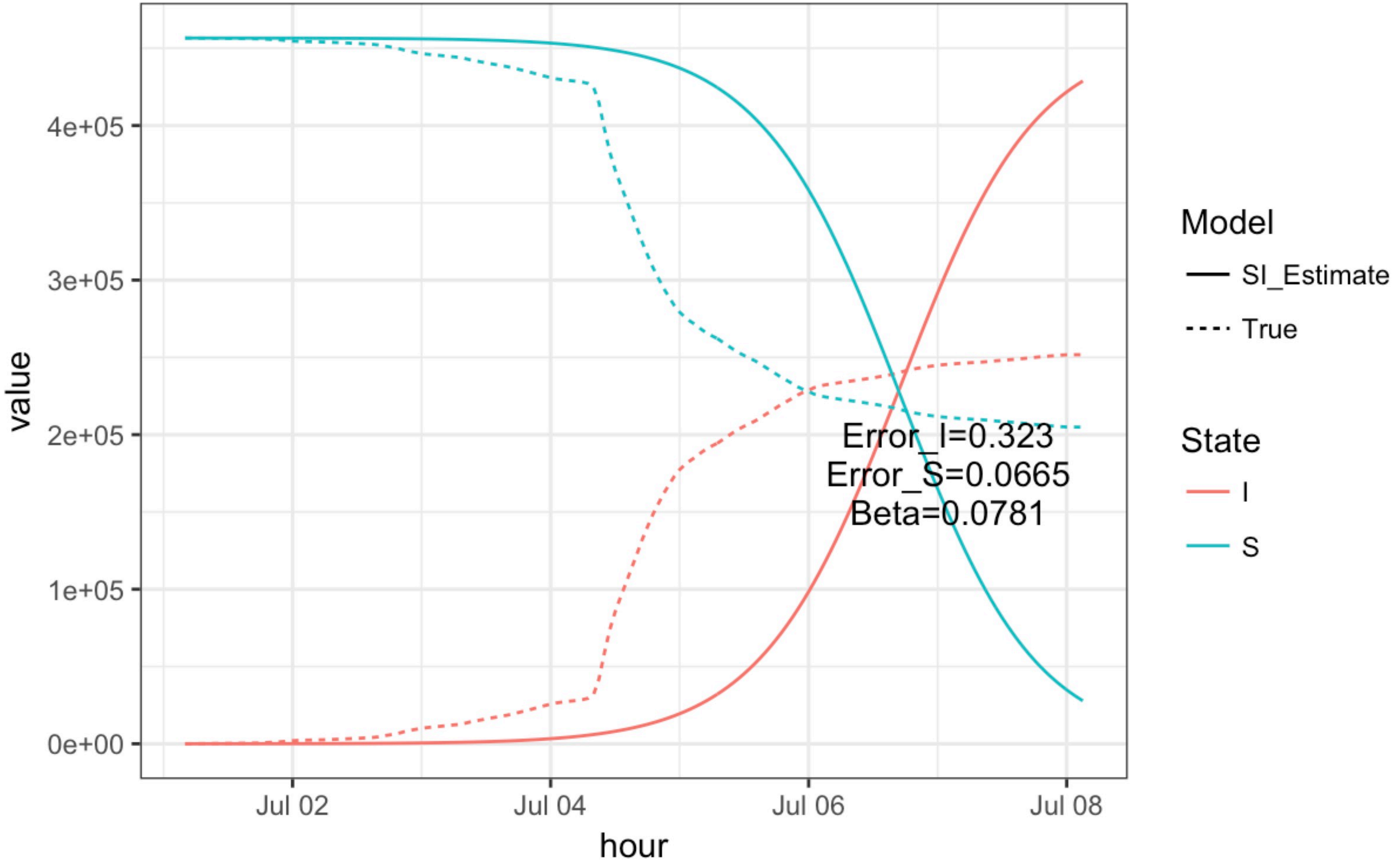


SI Model: Friendship Network

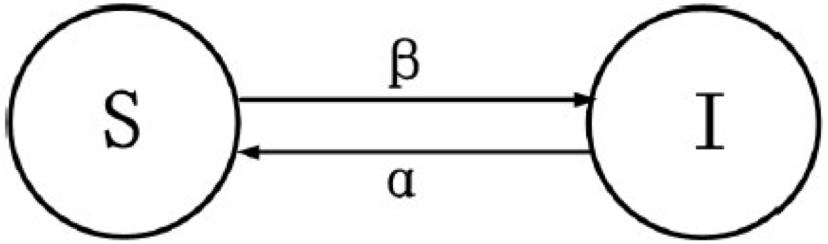


$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N}$$

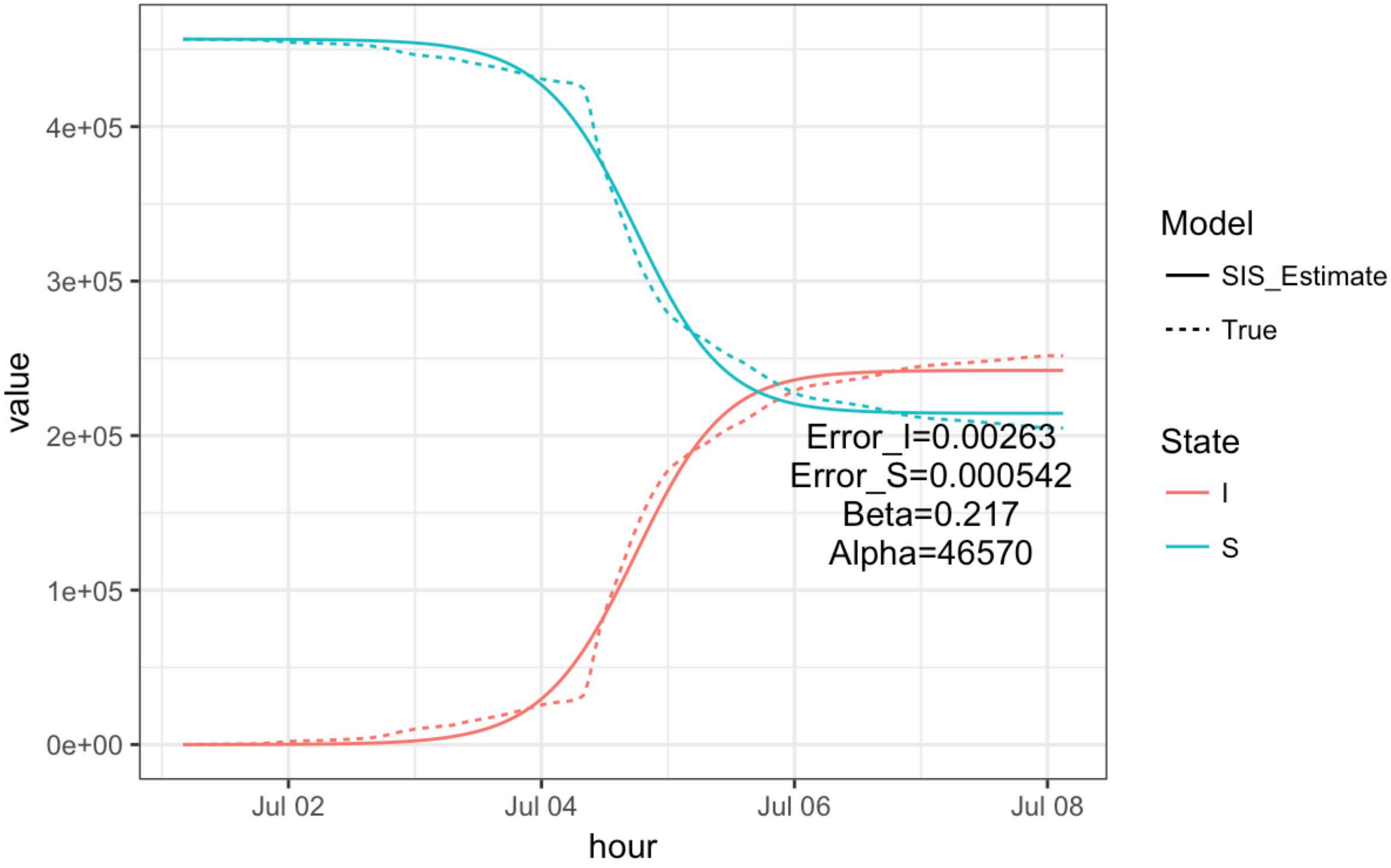


SIS Model: Friendship Network

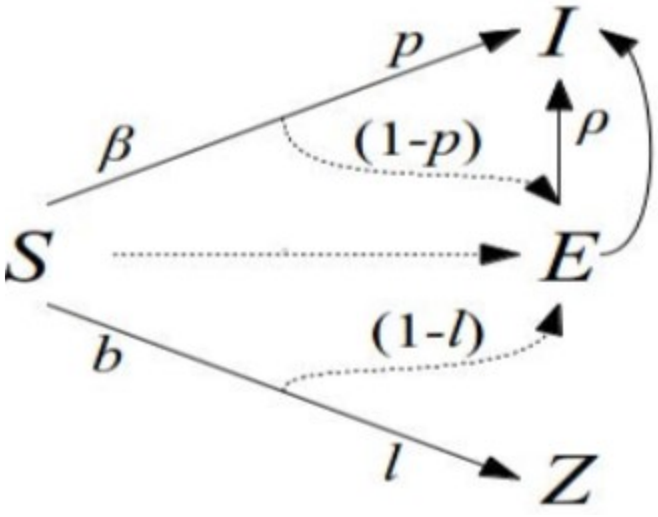


$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \alpha I$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \alpha I$$



SEIZ Model: Friendship Network

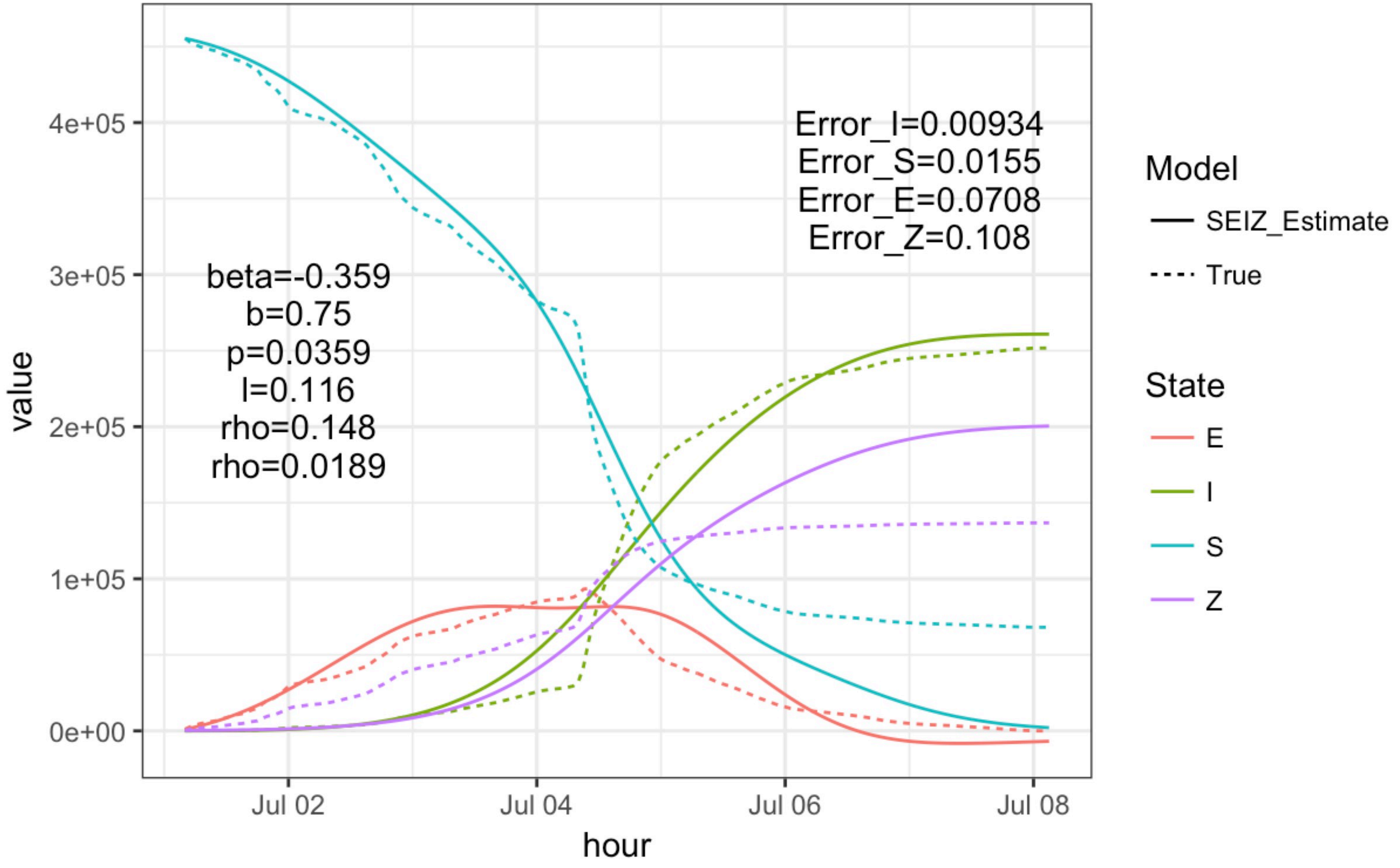


$$\frac{d[S]}{dt} = -\beta S \frac{I}{N} - bS\frac{Z}{N}$$

$$\frac{d[E]}{dt} = (1-p)\beta S\frac{I}{N} + (1-l)bS\frac{Z}{N} - \rho E\frac{I}{N} - \epsilon E$$

$$\frac{d[I]}{dt} = p\beta S\frac{I}{N} + \rho E\frac{I}{N} + \epsilon E$$

$$\frac{d[Z]}{dt} = lbS\frac{Z}{N}$$



$$\frac{\epsilon}{\rho} \approx 0.127 \qquad R_{SI} = \frac{(1-p)\beta + (1-l)b}{\rho + \epsilon} \approx 1.900$$

Conclusion

- The peak and decay in information spread largely rely on the attendance and dropout of a few large influential users. As these influential users dropout, they would leave a large quantity of "node-to-node" patterns, which turns out to be broad local discussion to the public.
- The information spread on Twitter can be accurately captured by compartmental models in epidemiology. The speed of information spreading is relatively high (contact rate is relatively high). Meanwhile the fading speed of information spreading is relatively high
- The SEIZ model are more interpretable compared to traditional compartmental models.

Reference

[1] Luís M.A. Bettencourt, Ariel Cintrón-Arias, David I. Kaiser, Carlos Castillo-Chávez. The power of a good idea: quantitative modeling of the spread of ideas from epidemiological models, 2006.

[2] M. De Domenico, A. Lima, P. Mougel, M. Musolesi. The Anatomy of a Scientific Rumor, 2013.

[3] Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, Naren Ramakrishnan. Epidemiological modeling of news and rumors on twitter, 2013.

[4] M. E. Newman. pread of epidemic disease on networks, 2002.

[5] Laijun Zhao, Hongxin Cui, Xiaoyan Qiu, Xiaoli Wang and Jiajia Wang. Sir rumorspreading model in the new media age, 2013.