Investigating the Observability of Complex Contagion in Empirical Social Networks

Clay Fink[1], Aurora Schmidt[1], Vladimir Barash[2], John Kelly[2], Christopher Cameron[3], Michael Macy[3]

[1]Johns Hopkins University Applied Physics Laboratory

[2]Graphika, Inc.

[3]Cornell University

NSF-IBSS, San Diego, CA, August 1-2, 2016

This work was funded by the Minerva Initiative through the United States Air Force Office of Scientific Research (AFOSR) under grant FA9550-15-1-0036.

Complex contagions and social movements

Threshold-based, or *complex*, models of social contagion may partly explain the initiation of mass mobilizations and social movements

Prior work

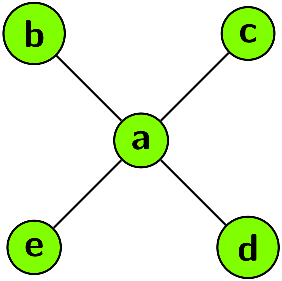
- Threshold models of collective behavior and theoretical predictions

  (Granovetter 1978, 1973); (Centola, Macy 2007); (Barash, Cameron, Macy 2012)

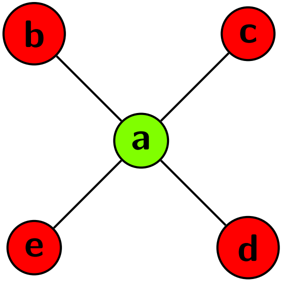
- Observational Studies: focus on empirical adoption thresholds

  Coleman, et al. (1966); Valente (1996): empirical studies of social reinforcement for medical practices and diffusion of innovations; Romero, et al. (2011), Fink, et al. (2016): spread of hashtags on Twitter; State and Adamic (2015): adoption of Equal-Sign profile pictures on Facebook

Overestimation of adoption thresholds



At time t none of a's neighbors have adopted

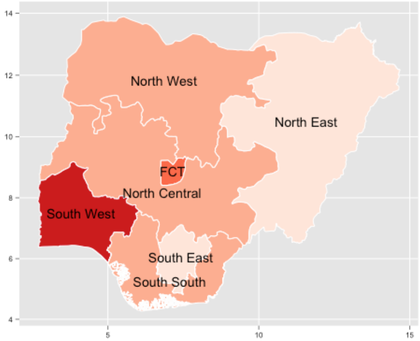


By time $t + dt$ all neighbors have adopted. If a now adopts, what was their actual adoption threshold?

This work

- We formulate *comparable* probabilistic models of simple and complex contagion to generate predictions of Twitter hashtag diffusion events
- Using the follow network of 53K Nigerian 2014 users and true adoption curves of 20 popular Nigerian hashtags, we
  1) Perform efficient search for unknown infection parameters under both simple and complex probabilistic models
  2) Compare the explanatory power of each model with optimized parameters against the true cumulative adoption curves
  3) Show, under asynchronous simulation and an empirical follow network, that when probabilistic infection parameters are not known, simple and complex models are not distinguishable by distributions of observed adoption thresholds

Dataset

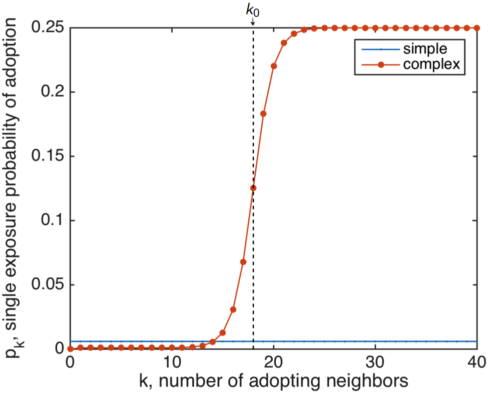


- Used the public API to query for tweets from around 45 Nigerian cities
- Queried timelines of users returned by geo queries
- Looked at hashtag cascades that began in 2014
- Retrieved follow graph for 53K active 2014 Nigerian Twitter users using the public API

Simulating asynchronous user check-in schedules

- Define a *check-in time* as the end of any 15-minute period during which user has tweeted
- Calculate each user's hourly check-in rate, λ
- For simplicity, we assume stationarity and a common check-in rate across users
- Used an average λ of 0.38 based on the calculated check-in rates of 39K of the studied users
- Resampled check-in times for all users every 10 simulations

Probabilistic models of simple and complex contagion



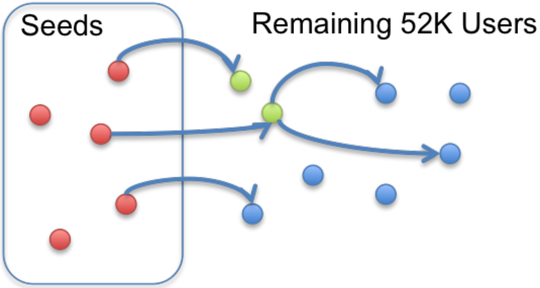
Simple contagion model:

$$p_k = p$$

Complex contagion model:

$$p_k = \epsilon_{\mathrm{lo}} + \frac{\epsilon_{\mathrm{hi}} - \epsilon_{\mathrm{lo}}}{1 + e^{-g(k-k_0)}}$$

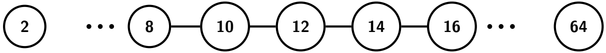
Simulations

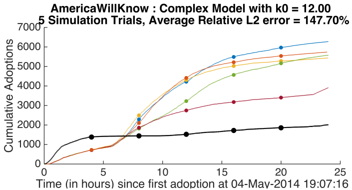


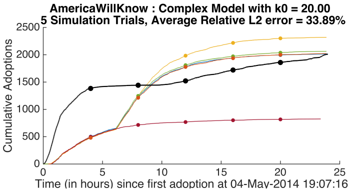
- Restrict analysis to first 24 hours after the first day of significant usage of a tag
- We do not model the adoptions of instigators nor adoptions due to external influence. We use the true schedule of these "seeds" and predict adoption cascades of remaining non-seed users of the full 52K-sized network, modeling asynchronous random time-line checking

Parameter fitting

Range of 32 k_0 values to search



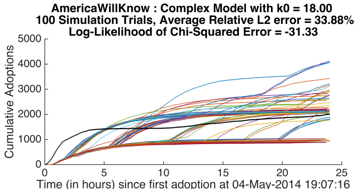




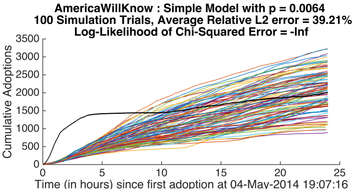
1) Use binary search over unknown parameter values. After 5 simulations, if majority of interpolated adoption predictions are above the true adoption curve, we try larger k_0; if below, test smaller k_0

2) Use lower and upper bounds found in stage 1) for a finer-grained search. Select parameter with the lowest average ℓ_2 prediction error

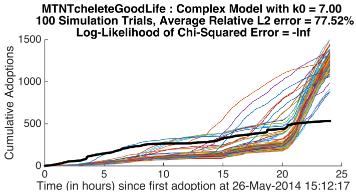
Simulation results (based on 100 simulations per tag)



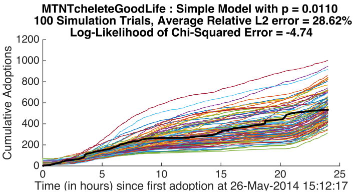
(a) Complex ($k_0 = 18$)



(b) Simple ($p = 0.0064$)

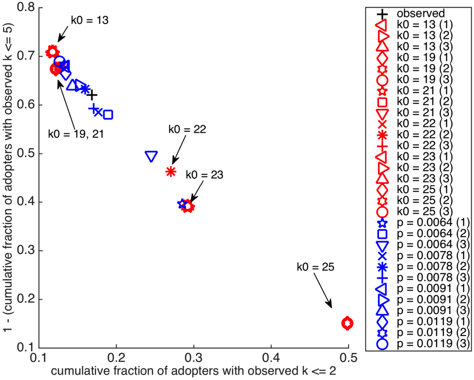


(c) Complex ($k_0 = 7.00$)



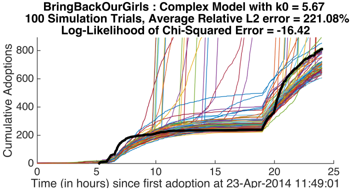
(d) Simple ($p = 0.0110$)

Using observed thresholds (k) to separate complex from simple contagion (#AmericaWillKnow)

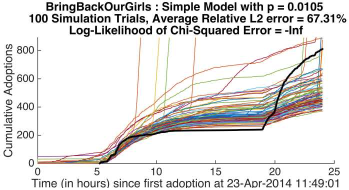


- For three simulations of each model, plotted the percentage of adopting users with $k \leq 2$ vs. those with $k > 5$
- If k were a reliable measure, we would expect to see a separation between the simulations of complex and simple models and their parameters; we do not
- We also see that k consistently overestimates the level of social reinforcement indicated by the model parameters

Conclusions and future work



(e) Complex ($k_0 = 7.00$)



(f) Simple ($p = 0.0110$)

- Probabilistic simulation methods show promise at identifying phenomena that is better explained by social reinforcement
- Most tags are explained by the simple model (as expected), some tags yield large variations in adoption curves (e.g. BringBackOurGirls), and some tags fit neither model most likely due to external spread
- Plan to reduce effects of "superemitters" to address early takeoff predictions and more accurately model user check-in rates for non-stationary and heterogeneous behavior

Further reading

Fink, C., Schmidt, A., Barash, V., Cameron, C., & Macy, M. (2016a). Complex contagions and the diffusion of popular Twitter hashtags in Nigeria. Social Network Analysis and Mining, 6(1), 1-19.

Fink, C., Schmidt, A. C., Barash, V., Kelly, J., Cameron, C., & Macy, M. (2016b). Investigating the Observability of Complex Contagion in Empirical Social Networks. In Tenth International AAAI Conference on Web and Social Media (pp. 121-130). Menlo Park, CA: AAAI

Thank You!

Questions?

Contact: clayton.fink@jhuapl.com