Estimating Contagion Rates in Kickstarter Twitter Cliques

VIRGINIA TECH™

DAVON WOODARD/MD SALMAN AHMED CS 6604 FALL 2017

Outline

1. Introduction and Initial Motivation
2. Data Collection
3. Assumptions
4. Simple Contagion
    1. Model
    2. Results
5. Complex Contagion
    1. Model
    2. Results
6. Challenges
7. Conclusion

Introduction

- Crowdfunding has emerged as a popular community-based, micro-financing model for entrepreneurs, artists, and activists alike to bring their respective dreams into fruition.
- Successful campaigns, those which meet their financial goals, bring with them not only the financial utility for the creator, but also social utility for the backers.

Motivation

Initial: Can the prediction power of Twitter data be extended by complementing static data with the model of social media exposure curves (stickiness and persistence) presented by Romero[1], et. al coupled with the use of censored data presented by Li, et. al[2]?

Secondary: Given a set of assumptions, can the rate of spread of Kickstarter campaigns in a Twitter network be estimated using simple contagion and complex contagion models?

Data Collection

- **Crowdfunding dataset**
  - From Chandan K. Reddy's Team

- **Twitter dataset**
  - Using Twitter public API & GetOldTweets-python[1])
  - Modifying the APIs of GetOldTweets-python to meet our needs

1. https://github.com/Jefferson-Henrique/GetOldTweets-python

Dataset Characteristics

- Crowdfunding dataset[1] – 18k total records
  - Each record corresponds a project
  - Contains project id, name, URL, duration, goal amount, pledged amount, …
- Twitter dataset[2] – 162k total records
  - Each record corresponds to a tweet
  - Contains the text, user, date of tweet, tweet link, retweet, etc.

1 & 2. You can get the two datasets on http://people.cs.vt.edu/ahmedms/cs6604.html

Relevancy [Crowdfunding dataset x Twitter dataset]

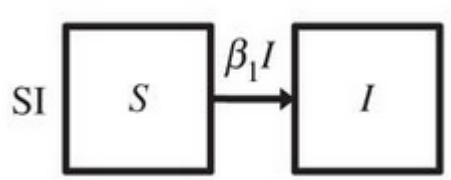
- Out of 18k projects, 10k projects have tweets
- Out of the 10k, 4k projects have enough tweets to take part in the model develop

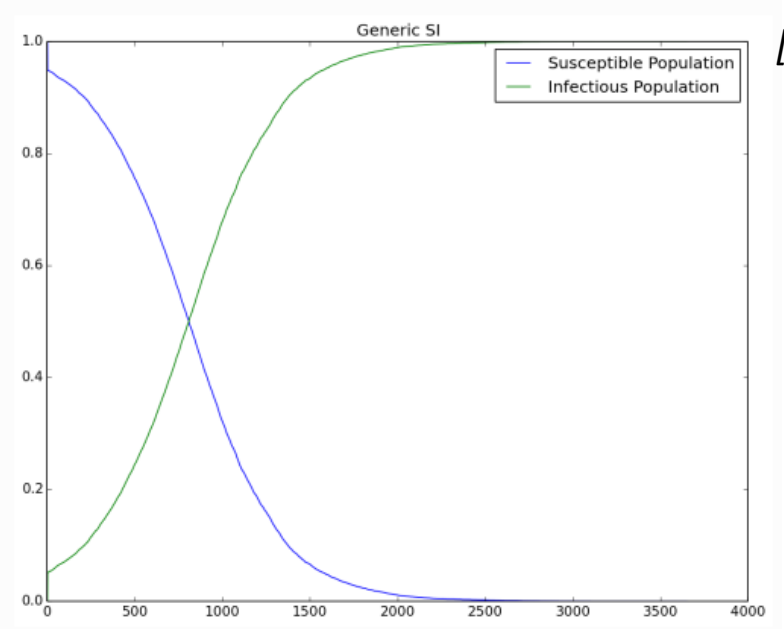
Assumptions

Our crawling program retrieves only the information about tweets.

- Twitter user network as a clique
- Total nodes of the Twitter user network is twice the unique Twitter users

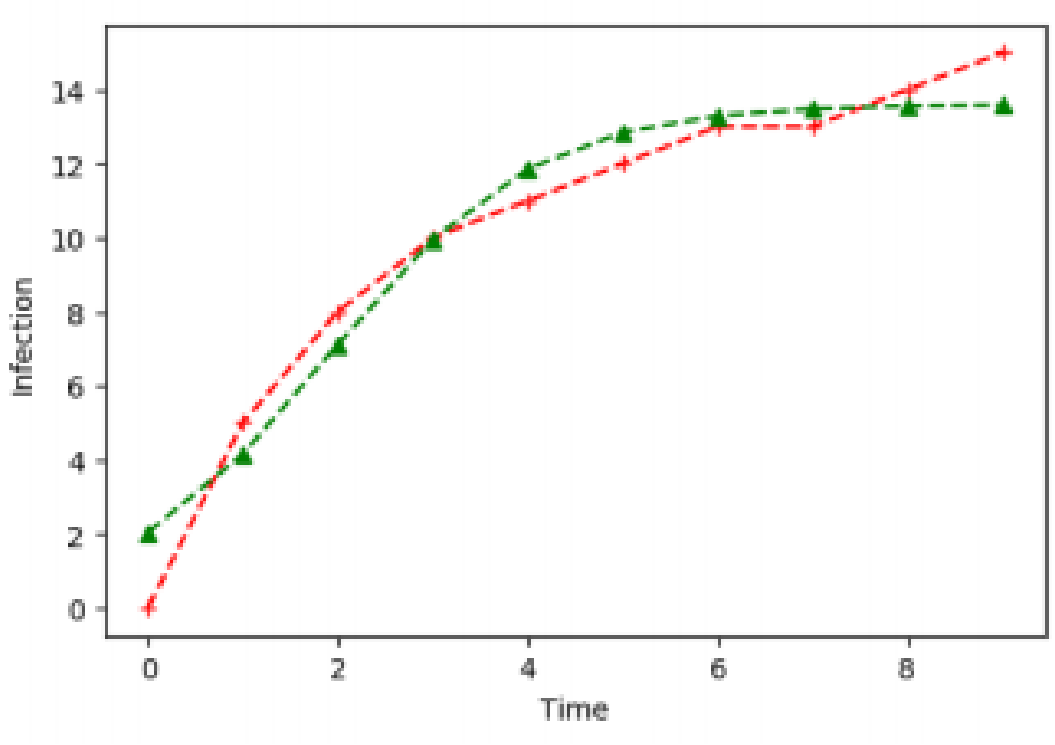
Model – Simple Contagion





[4]

Model – Simple Contagion

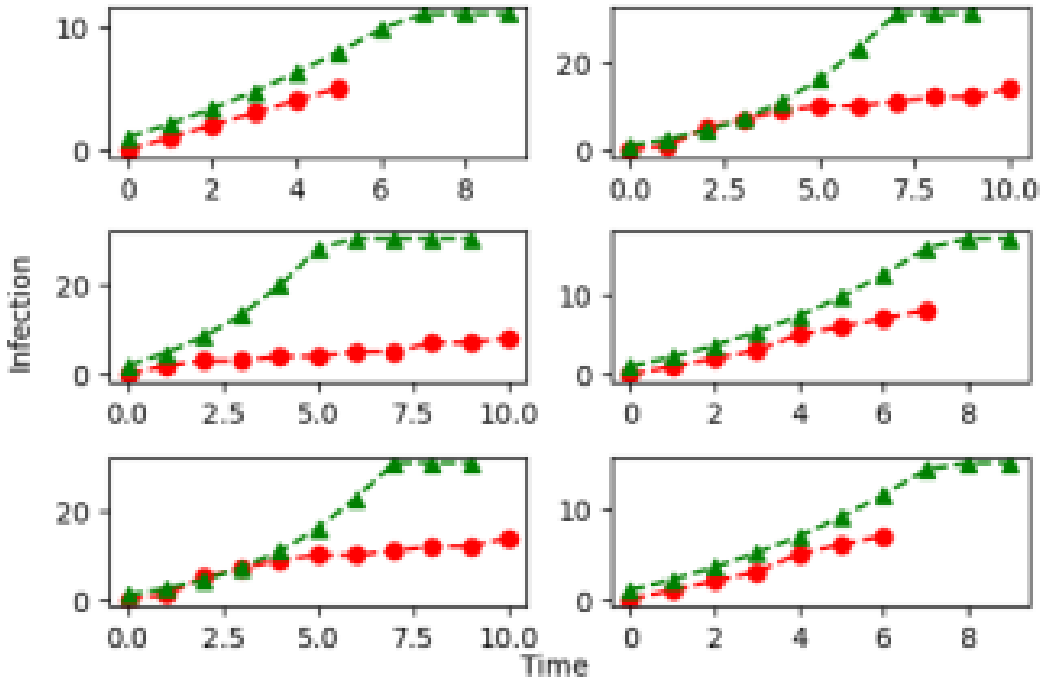


$$\frac{dS}{dt} = -\beta SI$$
$$\frac{dI}{dt} = \beta SI - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

$B_{(successful)} = .0092$

$B_{(failed)} = .0120$

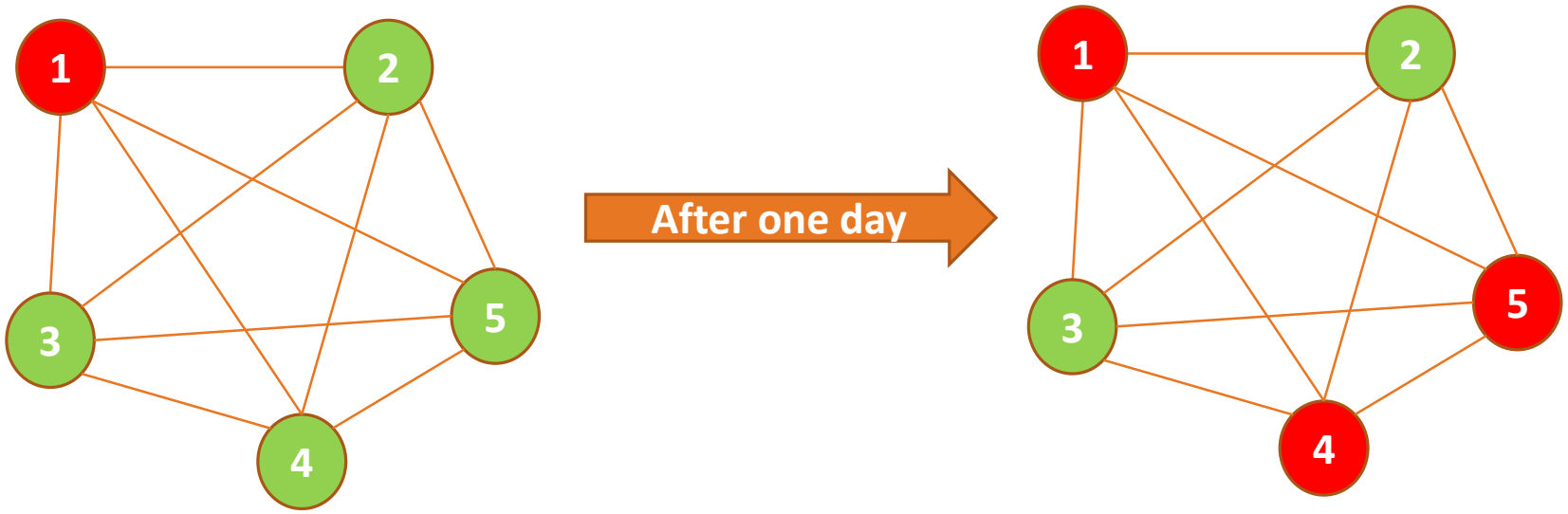
Results– Simple Contagion



$$I(t) = \frac{NI_0 e^{\beta N t}}{N + I_0[e^{\beta N t} - 1]}$$

- Variance:
  - Assumptions
  - Network Structure

Model - Complex Contagion [Concept of Exposure Curve]



Day 1

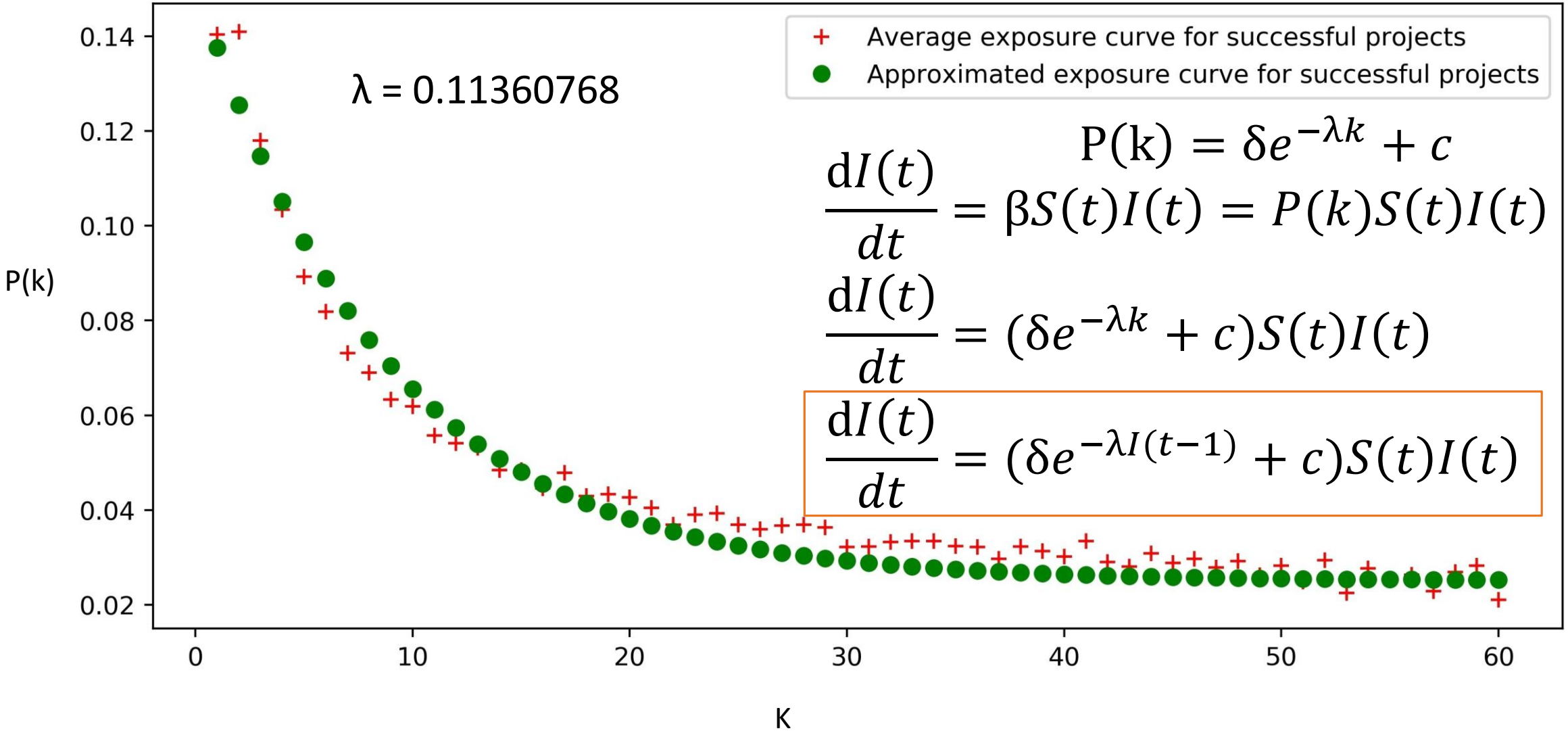
Day 2

$E_k = \{2, 3, 4, 5\}$

$P(k) = 2/4$
Where $k = 1$

$I_k = \{4, 5\}$

Exposure Curve



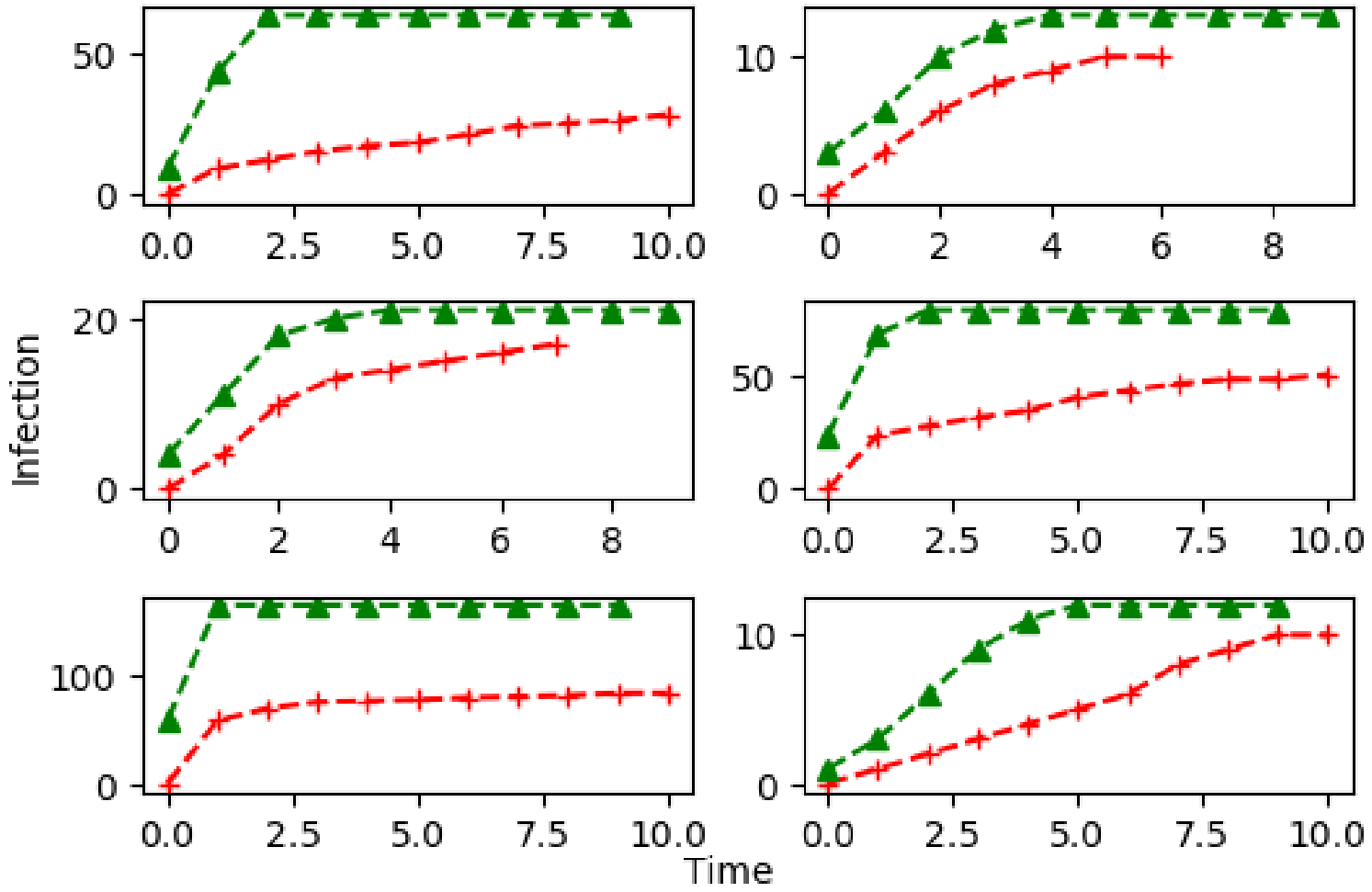
$$P(k) = \delta e^{-\lambda k} + c$$

$$\frac{dI(t)}{dt} = \beta S(t) I(t) = P(k) S(t) I(t)$$

$$\frac{dI(t)}{dt} = (\delta e^{-\lambda k} + c) S(t) I(t)$$

$$\frac{dI(t)}{dt} = (\delta e^{-\lambda I(t-1)} + c) S(t) I(t)$$

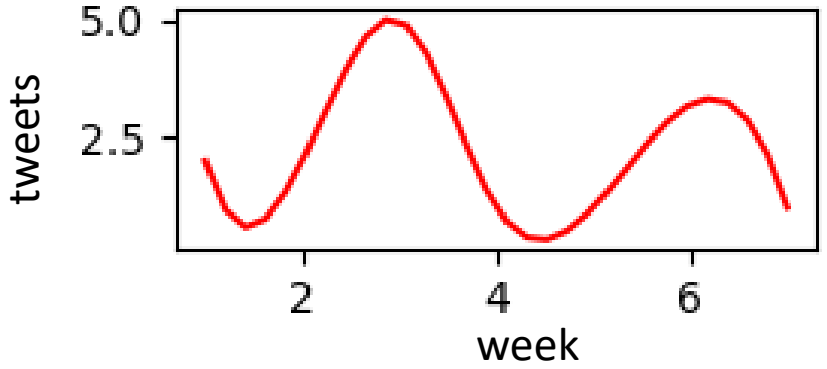
Results - Complex Contagion [Actual vs. Prediction]



- A big approximation error
- Inadequate twitter data
- Poor assumptions

Challenges

- Heading wrong direction until the last moment
- Thinking the output as the input to our models
- Inadequate associated Twitter data
- Poor assumptions



Conclusion

- Assumptions are critical
- Model should fit the dataset available
- With additional Network Information additional research on information diffusion of Kickstarter campaigns help guide marketing efforts

References

1. Romero, et. al . “Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter”

2. Yan Li, Vineeth Rakesh, and Chandan K. Reddy. 2016. Project Success Prediction in Crowdfunding Environments.

3. Prakash, Aditya. “Epidemics: Probabilistic Models”, Lecture, 9/25/17, VT

4. https://institutefordiseasemodeling.github.io/Documentation/general/generic-tutorial7SI.html