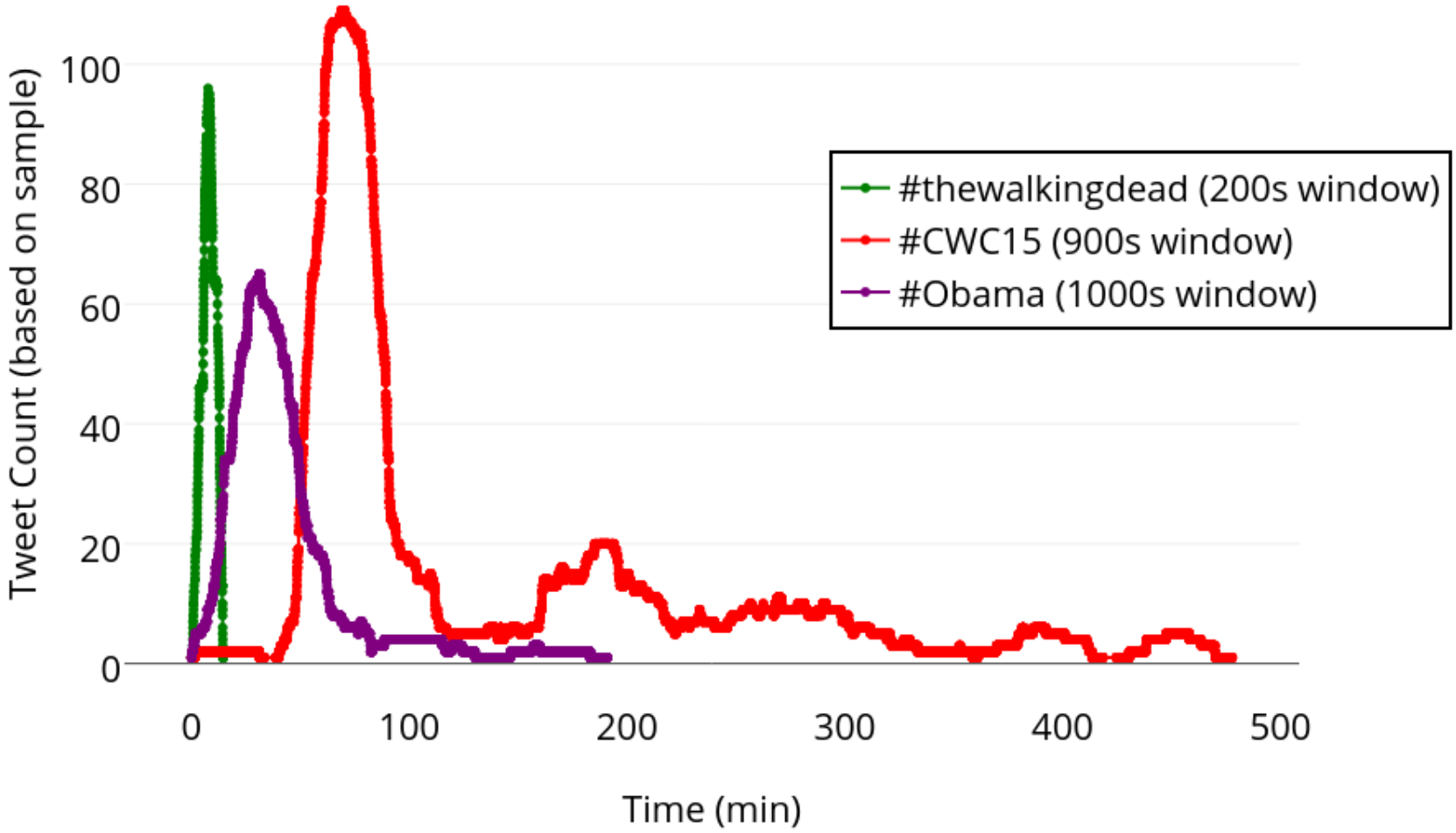
Mathematical Modeling of Trending Topics on Twitter

Jonathan Skaza

April 14, 2015

Comparison of Trending Topics



"Window" refers to the moving sum period (e.g., each point represents count in past 200s)

Project Objectives

- Quantify the diffusion of information on Twitter

Project Objectives

- Quantify the diffusion of information on Twitter
- Compare and contrast different trending topics

Project Objectives

- Quantify the diffusion of information on Twitter
- Compare and contrast different trending topics
- Compare spread of information to spread of infectious disease

Project Objectives

- Quantify the diffusion of information on Twitter
- Compare and contrast different trending topics
- Compare spread of information to spread of infectious disease
- Create a reproducible output product

Agenda

1. Twitter overview, facts, and figures

Agenda

1. Twitter overview, facts, and figures
2. Methodological strategy

Agenda

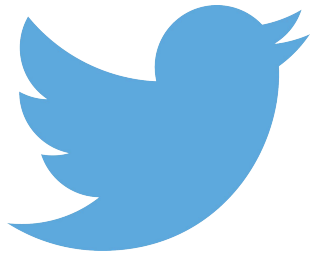
1. Twitter overview, facts, and figures
2. Methodological strategy
3. Previous studies

Agenda

1. Twitter overview, facts, and figures
2. Methodological strategy
3. Previous studies
4. Results and discussion

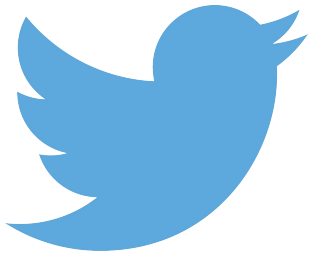
About Twitter

- Created in 2006



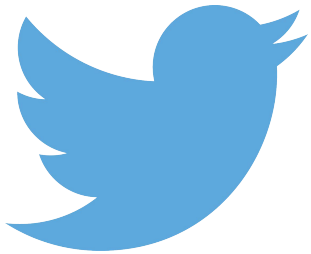
About Twitter

- Created in 2006
- Incorporated in 2007



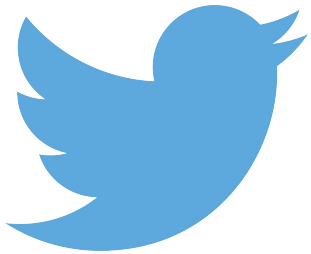
About Twitter

- Created in 2006
- Incorporated in 2007
- 288 million monthly active users – 500 million Tweets per day



About Twitter

- Created in 2006
- Incorporated in 2007
- 288 million monthly active users – 500 million Tweets per day
- Source: `about.twitter.com`



Anatomy of a Tweet

Jonathan Skaza
@SkazaSays

Follow

Fun fact: Since field was expanded to 64 teams, average sum of seeds in Final Four has been 11 (this year, it's 10) #MarchMadness

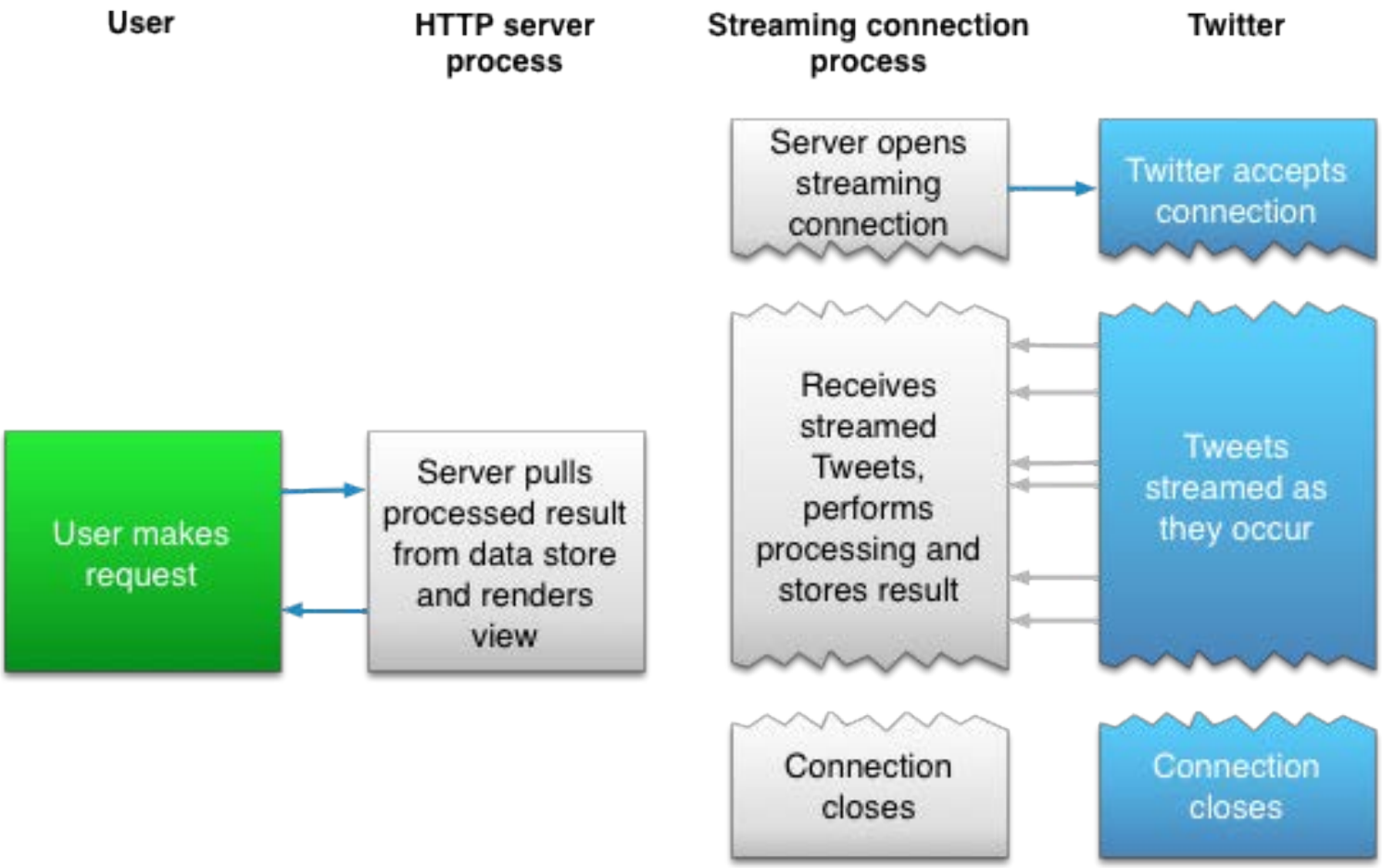
12:24 PM - 31 Mar 2015

1 RETWEET

Twitter Application Programming Interface (API)

- Two different flavors: REST and Streaming

Streaming API



Source: `dev.twitter.com`

Streaming API Request Parameters

delimited	locations
stall_warnings	count
filter_level	with
language	replies
follow	stringify_friend_id
track	

Source: Twitter Developers Documentation

stream.filter(track=['#'])

Data Collection

Raw Tweet

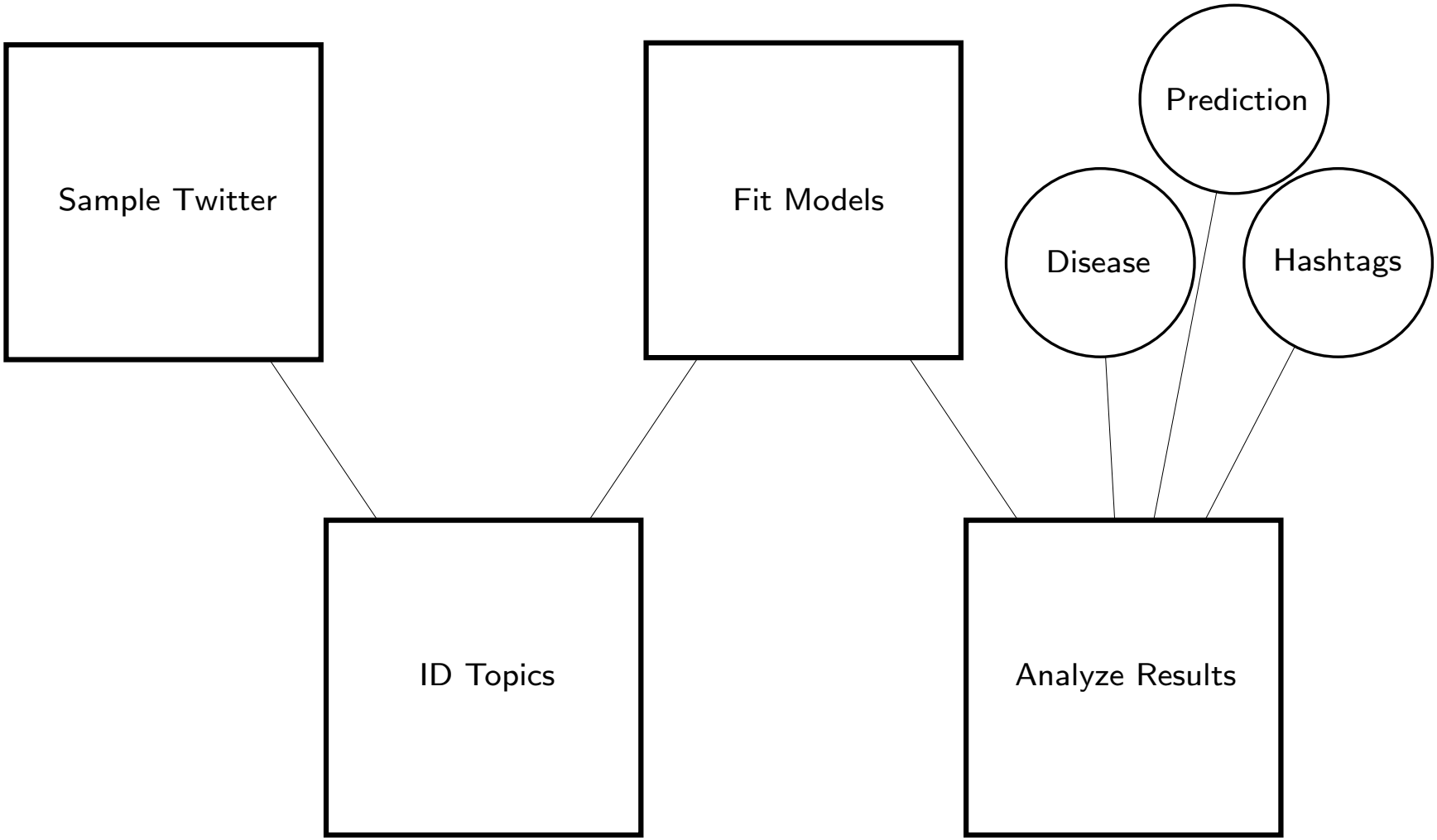

```
{"created_at":"Fri Mar 27 18:16:52 +0000 2015","id":581520276292280320,"id_str":"581520276292280320","text":
"Loving the #NCAA #MarchMadness? Find out fun facts like which states listened most, overall listening hours
and more! http:\/\/t.co\/DWfDTDnDg8","source":"\u003cahref=\"http:\/\/twitter.com\" rel=\"nofollow\"\u003e
Twitter Web Client\u003c\/a\u003e","truncated":false,"in_reply_to_status_id":null,"in_reply_to_status_id_str"
:null,"in_reply_to_user_id":null,"in_reply_to_user_id_str":null,"in_reply_to_screen_name":null,"user":{"id":
1694596596,"id_str":"1694596596","name":"Westwood One","screen_name":"WestwoodOne","location":"In your
speakers","url":"http:\/\/www.westwoodone.com","description":"Westwood One offers audio products and content
to reach listeners whenever, wherever they are. #powerofsound","protected":false,"verified":true,"followers_
count":1123,"friends_count":337,"listed_count":24,"favourites_count":1923,"statuses_count":2113,"created_at"
:"Fri Aug 23 19:28:54 +0000 2013","utc_offset":-10800,"time_zone":"Atlantic Time (Canada)","geo_enabled":
false,"lang":"en","contributors_enabled":false,"is_translator":false,"profile_background_color":"FAFAFA",
"profile_background_image_url":"http:\/\/pbs.twimg.com\/profile_background_images\/378800000066715369\/34
9a5b97fca21c477dd28089d909936b.png","profile_background_image_url_https":"https:\/\/pbs.twimg.com\/profi
le_background_images\/378800000066715369\/349a5b97fca21c477dd28089d909936b.png","profile_background_tile"
:false,"profile_link_color":"0A0A0A","profile_sidebar_border_color":"FFFFFF","profile_sidebar_fill_color"
:"DDEEF6","profile_text_color":"333333","profile_use_background_image":true,"profile_image_url":"http:\/\
/pbs.twimg.com\/profile_images\/489073660854935553\/a2WsGpB-_normal.jpeg","profile_image_url_https":"https:\
/\/pbs.twimg.com\/profile_images\/489073660854935553\/a2WsGpB-_normal.jpeg","profile_banner_url":"https:\/\
/pbs.twimg.com\/profile_banners\/1694596596\/1422292326","default_profile":false,"default_profile_image":
false,"following":null,"follow_request_sent":null,"notifications":null},"geo":null,"coordinates":null,"plac
e":null,"contributors":null,"retweet_count":0,"favorite_count":0,"entities":{"hashtags":[{"text":"NCAA","i
ndices":[11,16]},{"text":"MarchMadness","indices":[17,30]}],"trends":[],"urls[{"url":"http:\/\/t.co\/DWfDTD
nDg8","expanded_url":"http:\/\/bit.ly\/1CiR90h","display_url":"bit.ly\/1CiR90h","indices":[118,140]}],"user
_mentions":[],"symbols":[]},"favorited":false,"retweeted":false,"possibly_sensitive":false,"filter_level":
"low","lang":"en","timestamp_ms":"1427480212649"}
```

Data Collection

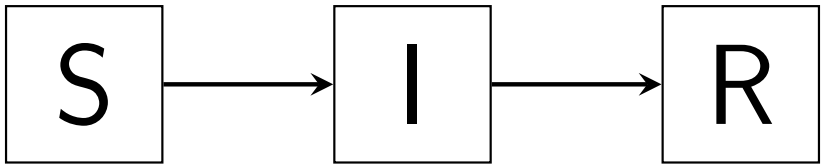
Processed Tweet

```
Fri Mar 27 18:16:52 +0000 2015,['NCAA', 'MarchMadness']
```

Methodology



SIR Model

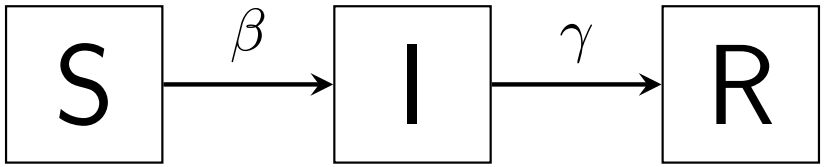


Developed by Kermack and McKendrick (1927)

Disease: Proximal to infected individual $\rightarrow$ Catch disease $\rightarrow$ Recover from disease

Meme: Twitter user $\rightarrow$ Tweet about topic $\rightarrow$ Move on in life

SIR Model



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

Use Markov Chain Monte Carlo (MCMC) simulation techniques to estimate β, γ, initial S, and initial I (Coelho, Codeco, and Gomes, 2011)

Previous Work

- Infectious Disease

Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)

Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)

Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)
- Other Applications

Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)
- Other Applications
  - Feynman diagrams (Bettencourt et al., 2006)

Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)
- Other Applications
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)

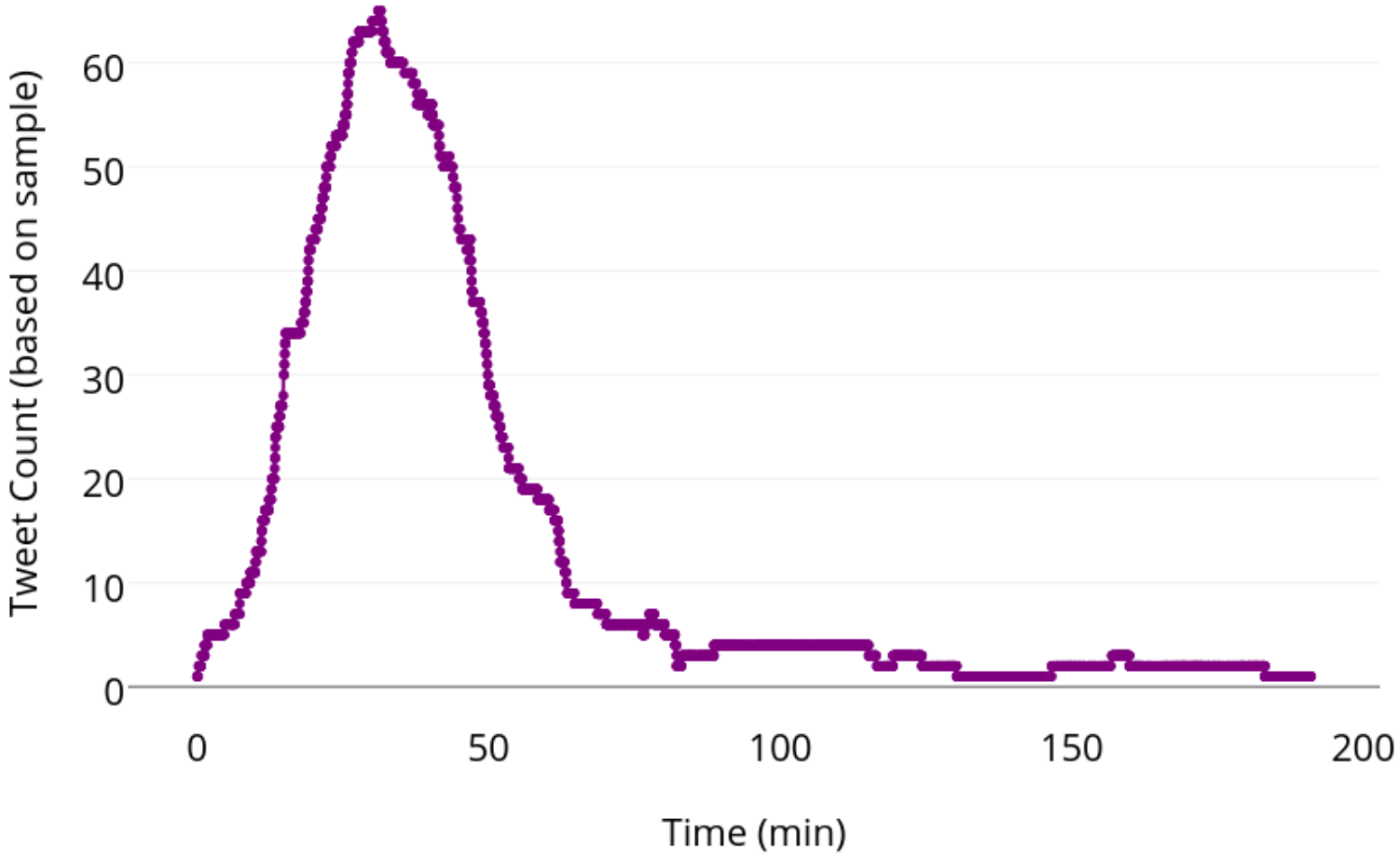
Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)
- Other Applications
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
  - Spread of rumors in social networks (Zhao et al., 2012)

Previous Work

- Infectious Disease
  - Measles (McGilchrist et al., 1996; Grais et al., 2006; Tuckwell and Williams, 2007; Kuniya, 2006)
  - Influenza (Tuckwell and Williams, 2007; Li, Li, and Ghosh, 2009; Hooten, Anderson, and Waller, 2010; Coelho, Codeco, and Gomes, 2011)
- Other Applications
  - Feynman diagrams (Bettencourt et al., 2006)
  - News and rumors on Twitter (Jin et al., 2013)
  - Spread of rumors in social networks (Zhao et al., 2012)
  - Zombie apocalypse (Witkowski and Blais, 2013)

Application of Methodology to #Obama



Specify Prior Probability Distributions

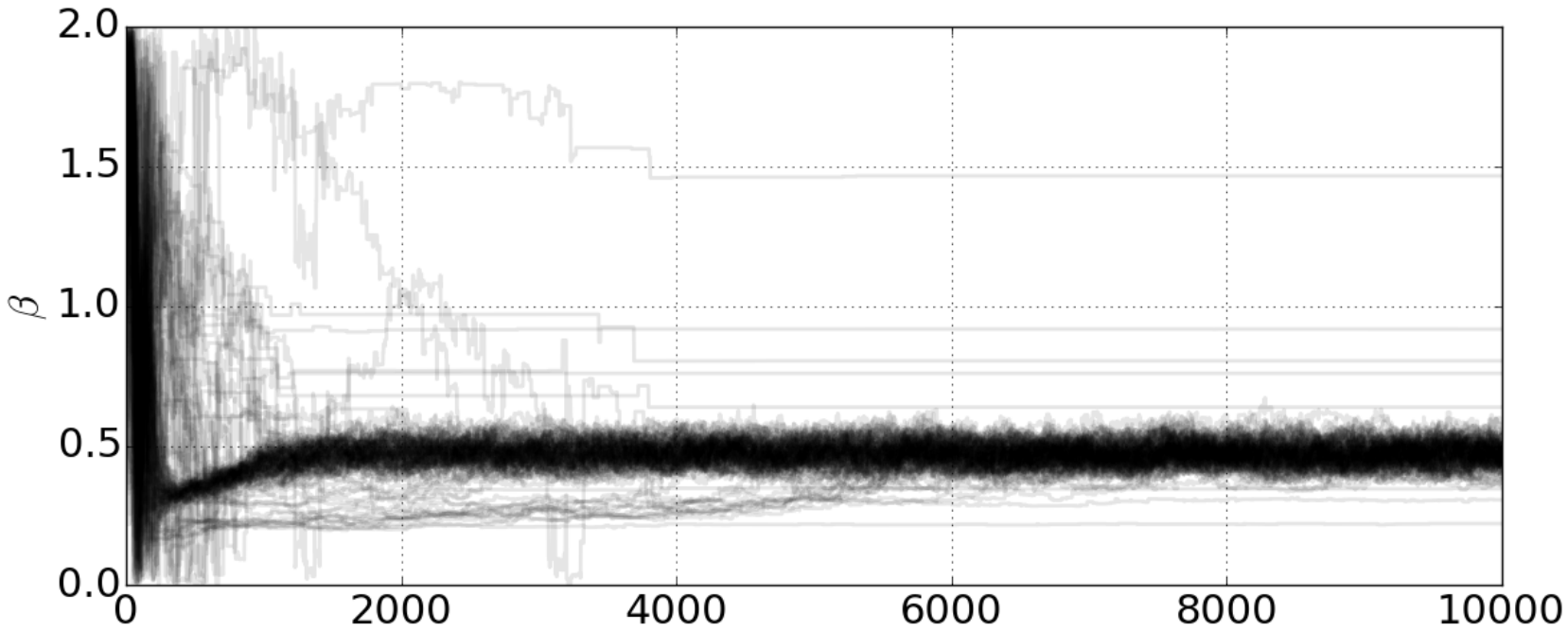
Prior Probability Distributions

- $\beta \sim U(0, 2)$
- $\gamma \sim U(0, 2)$
- $S_0 \sim U(30, 5000)$
- $I_0 \sim U(0, 10)$

Example of MCMC Parameter Estimation

Code Snippet

```
model = MCMCModel(sim, beta = Uniform(0,2))
model.run_mcmc(10000)
```



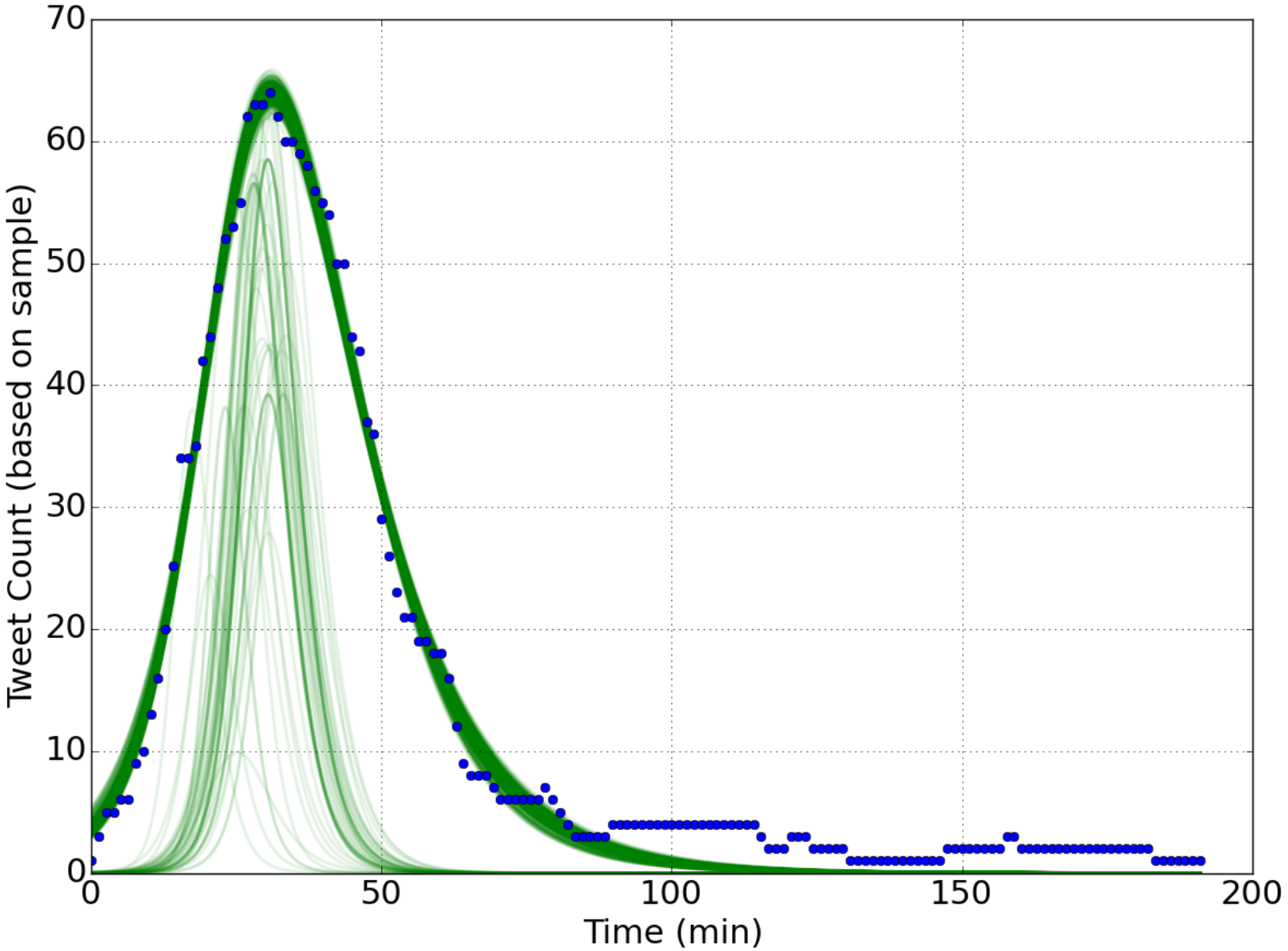
Run #Obama Simulation

Simulate 500 times, drawing from posterior probability distributions

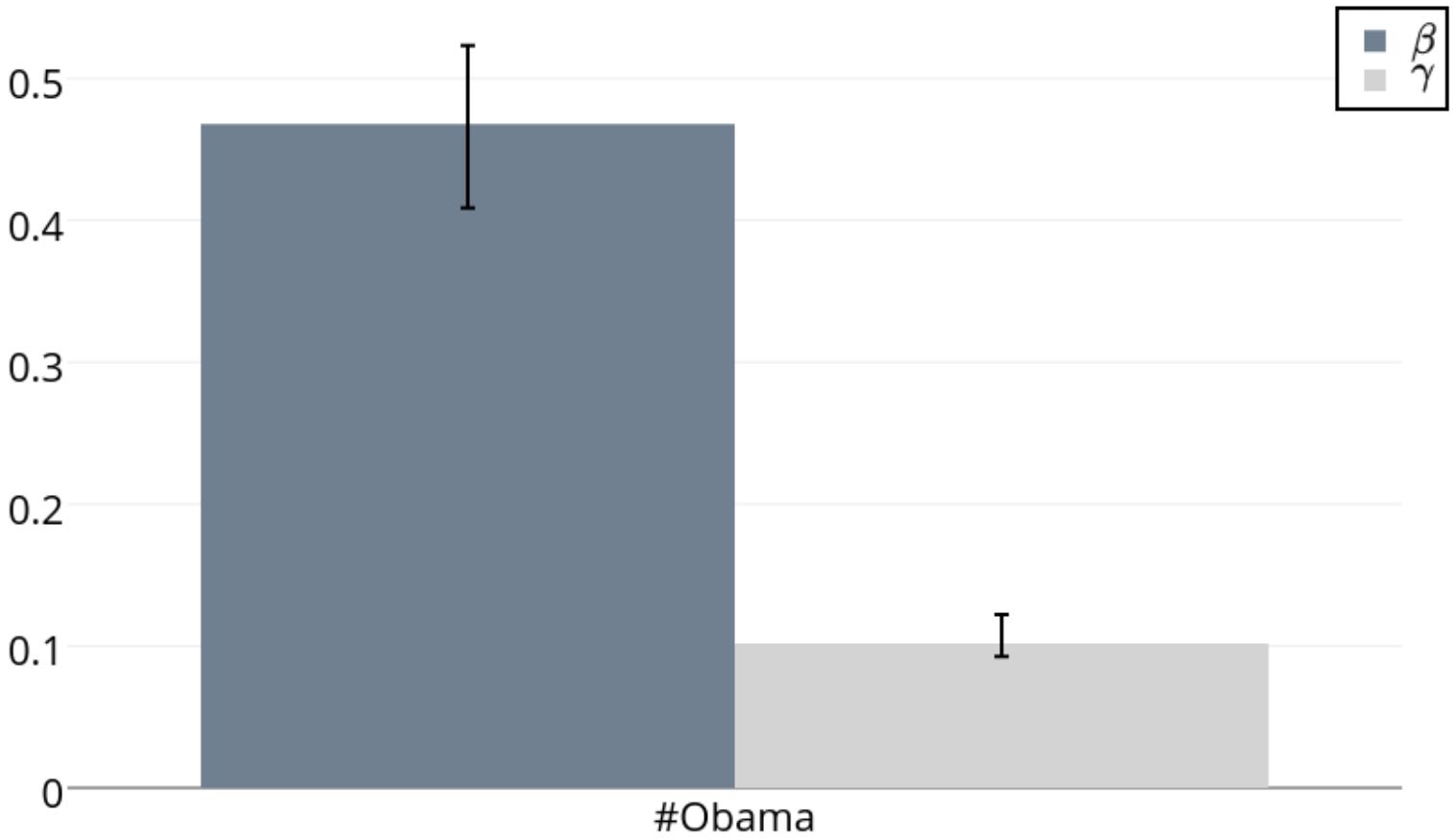
Code Snippet

```
for i in range(500):
    model.draw()
    sim.run(0,191)
    plot(sim.t,sim.I,'g-',alpha = .1)
```

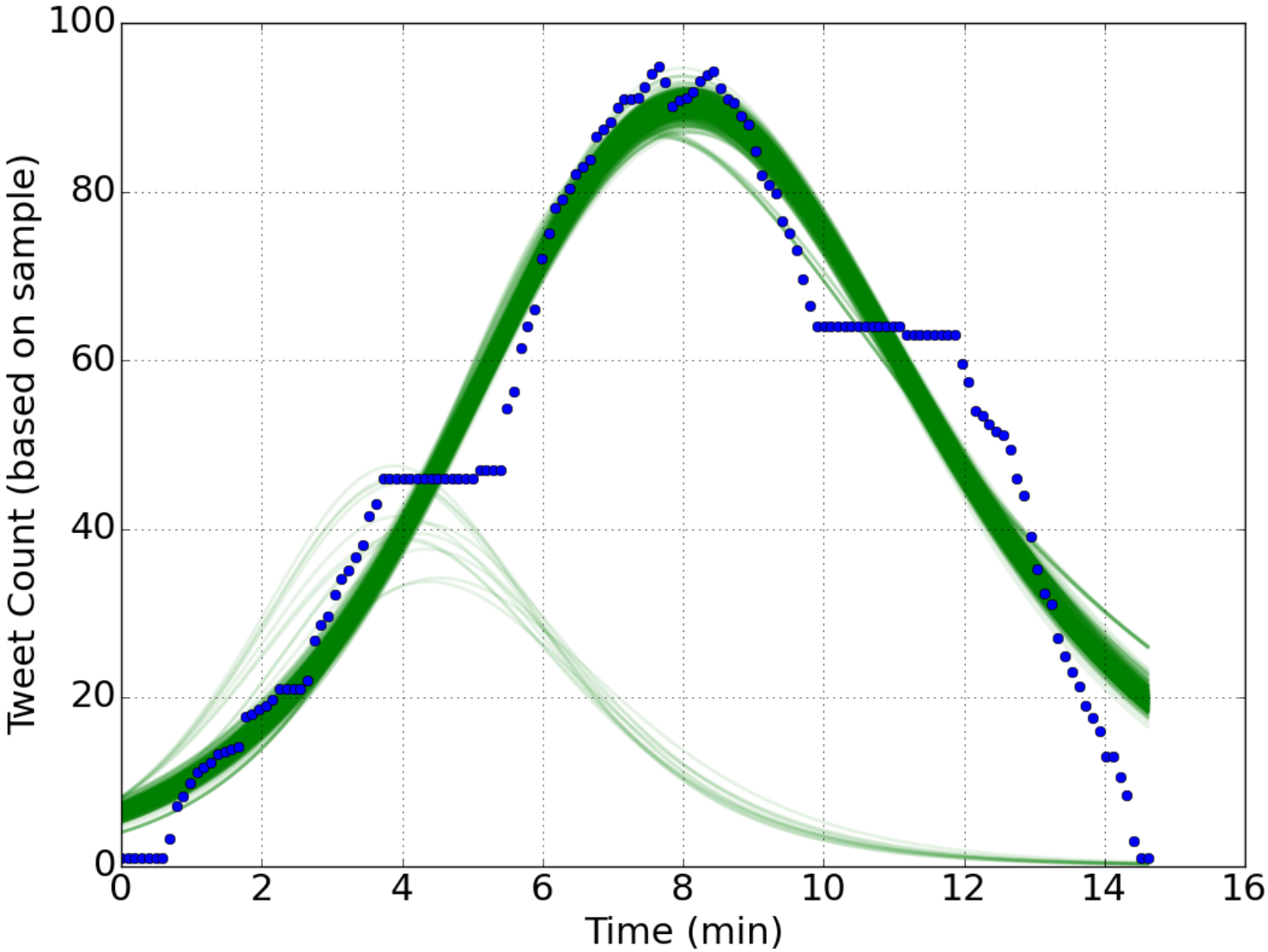
Simulation Results for #Obama



Best Parameter Estimates & Credible Intervals

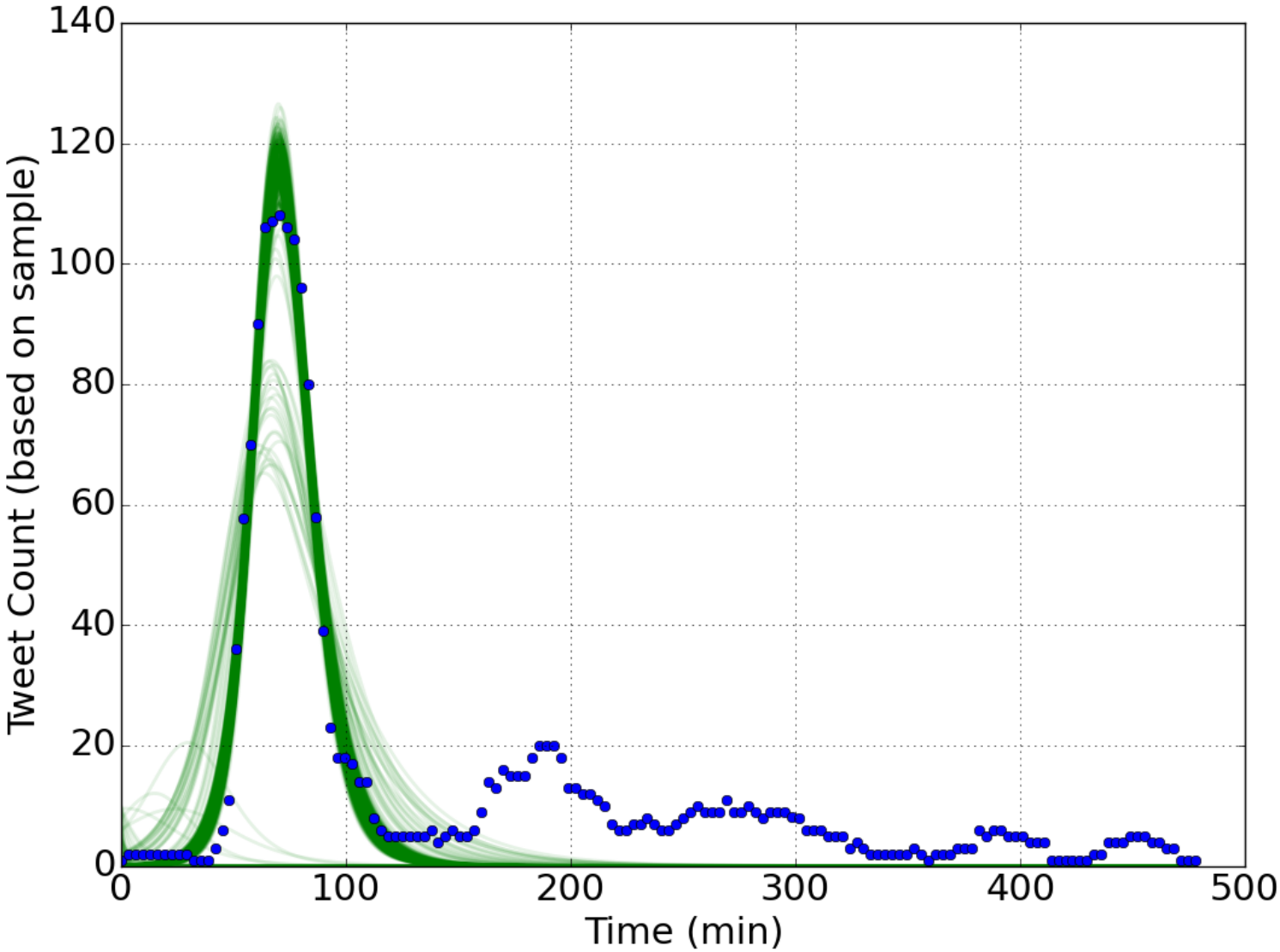


#thewalkingdead Simulation

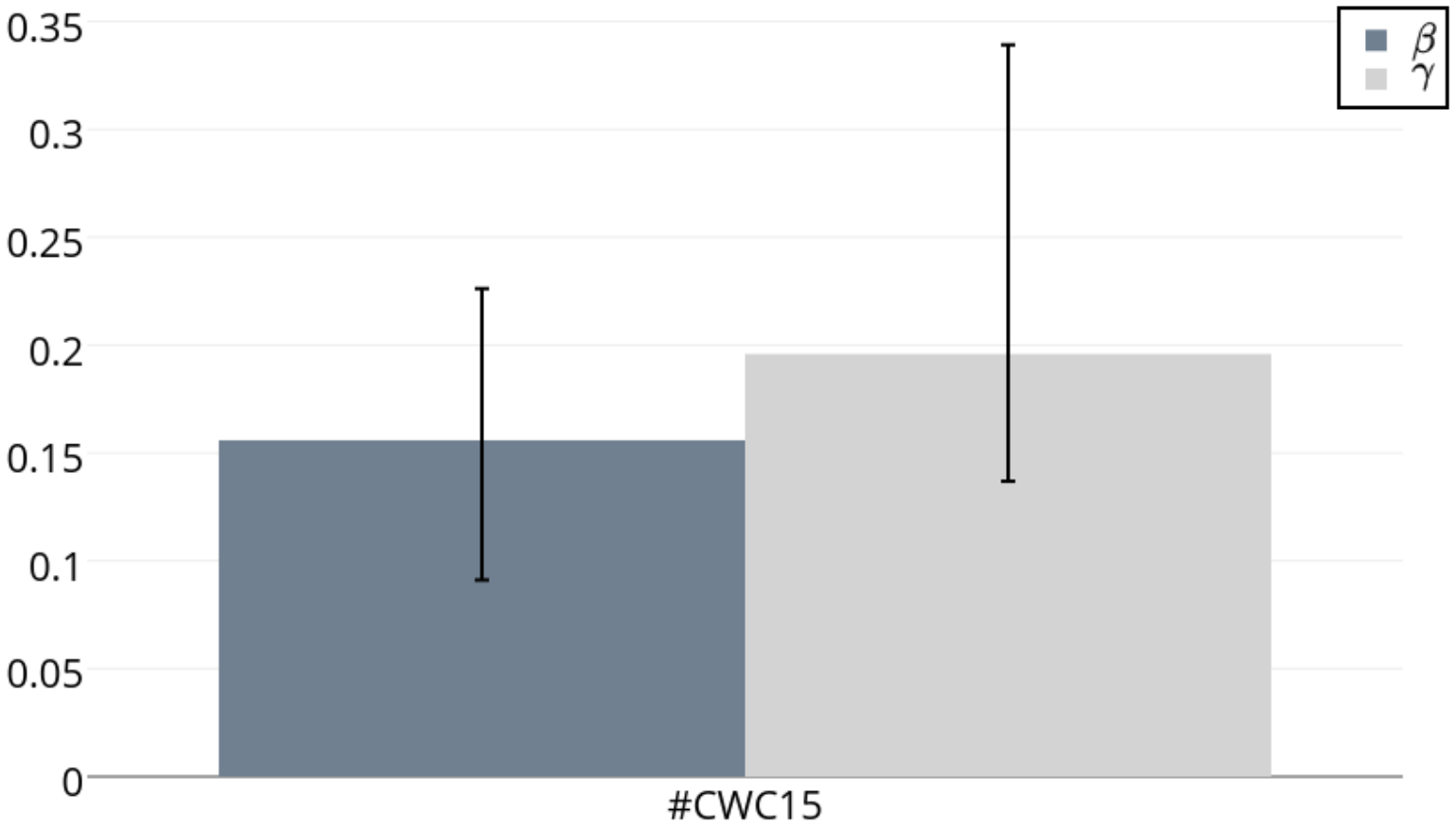


#thewalkingdead Parameter Estimation

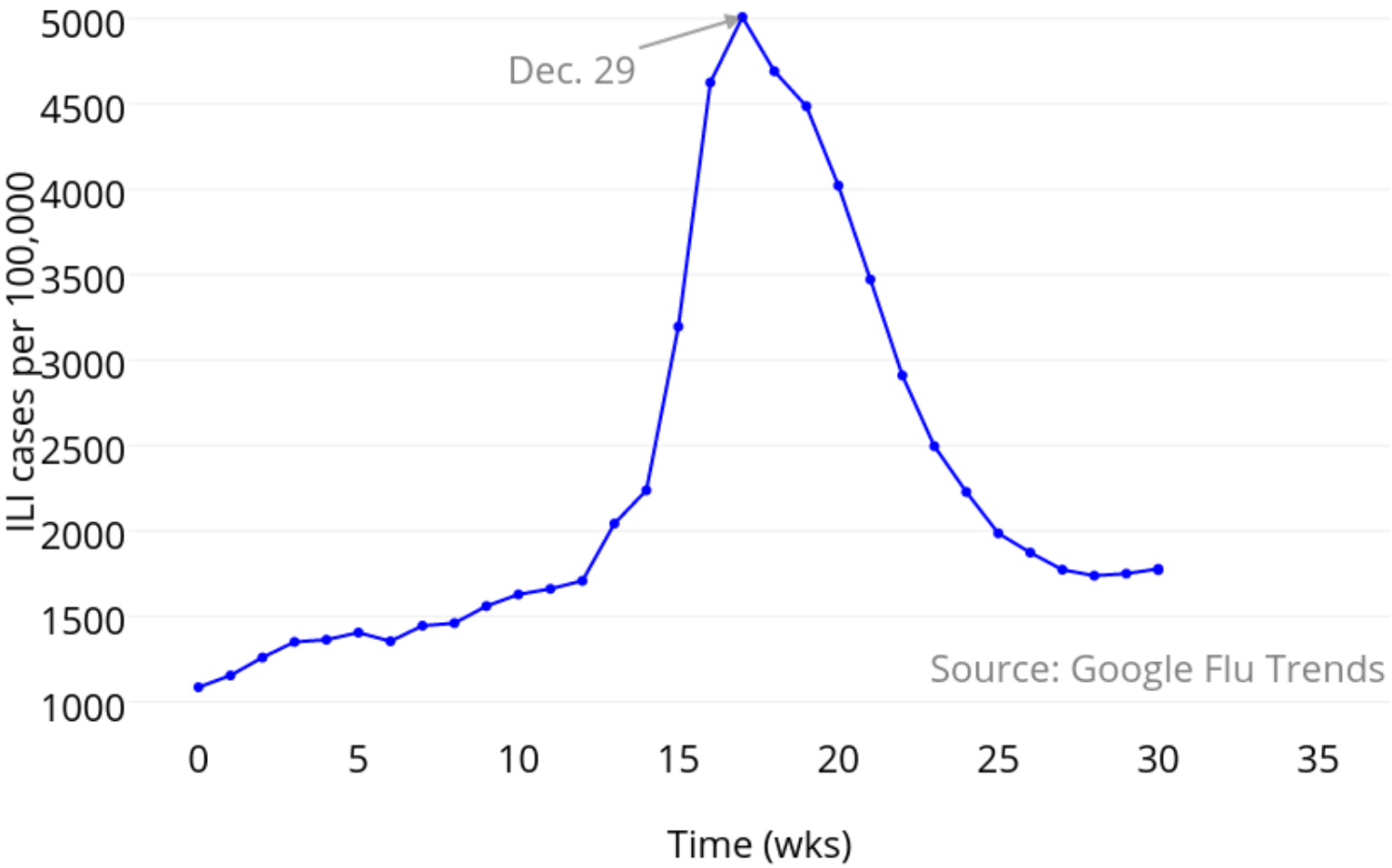
#CWC15 Simulation



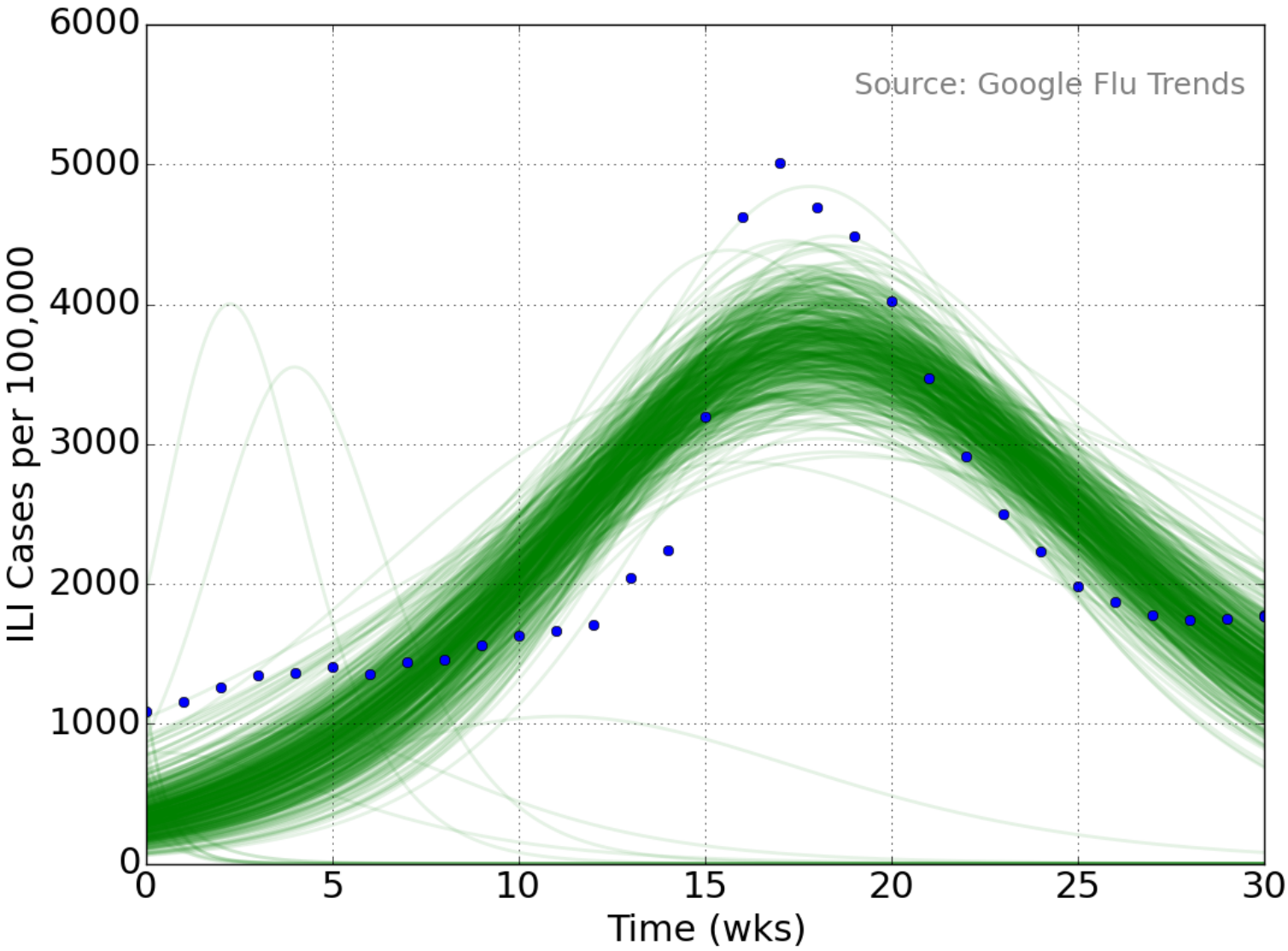
#CWC15 Parameter Estimation



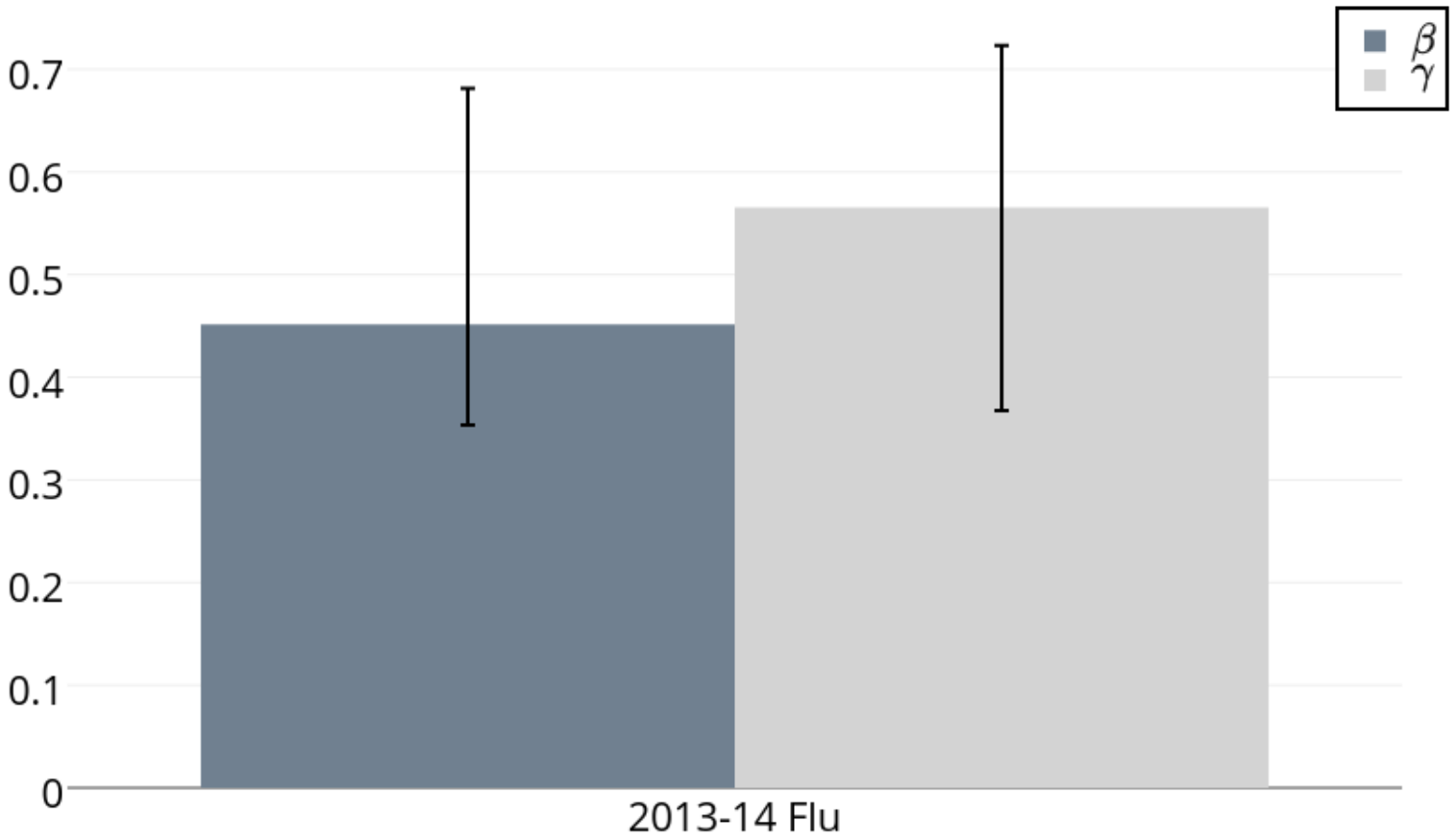
2013-14 U.S. Flu Season (September 1st - April 6th)



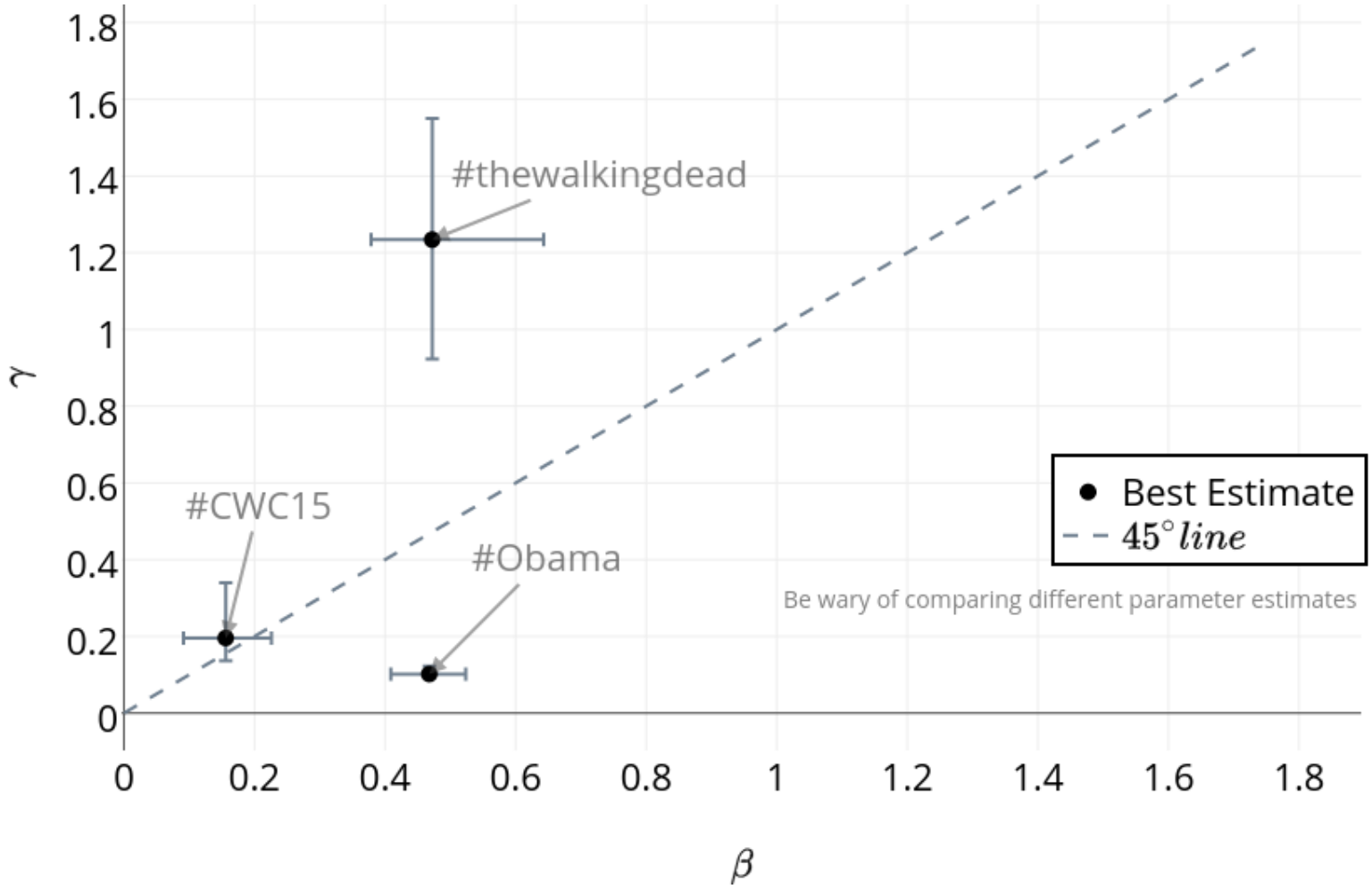
Flu Simulation



Flu Parameter Estimation

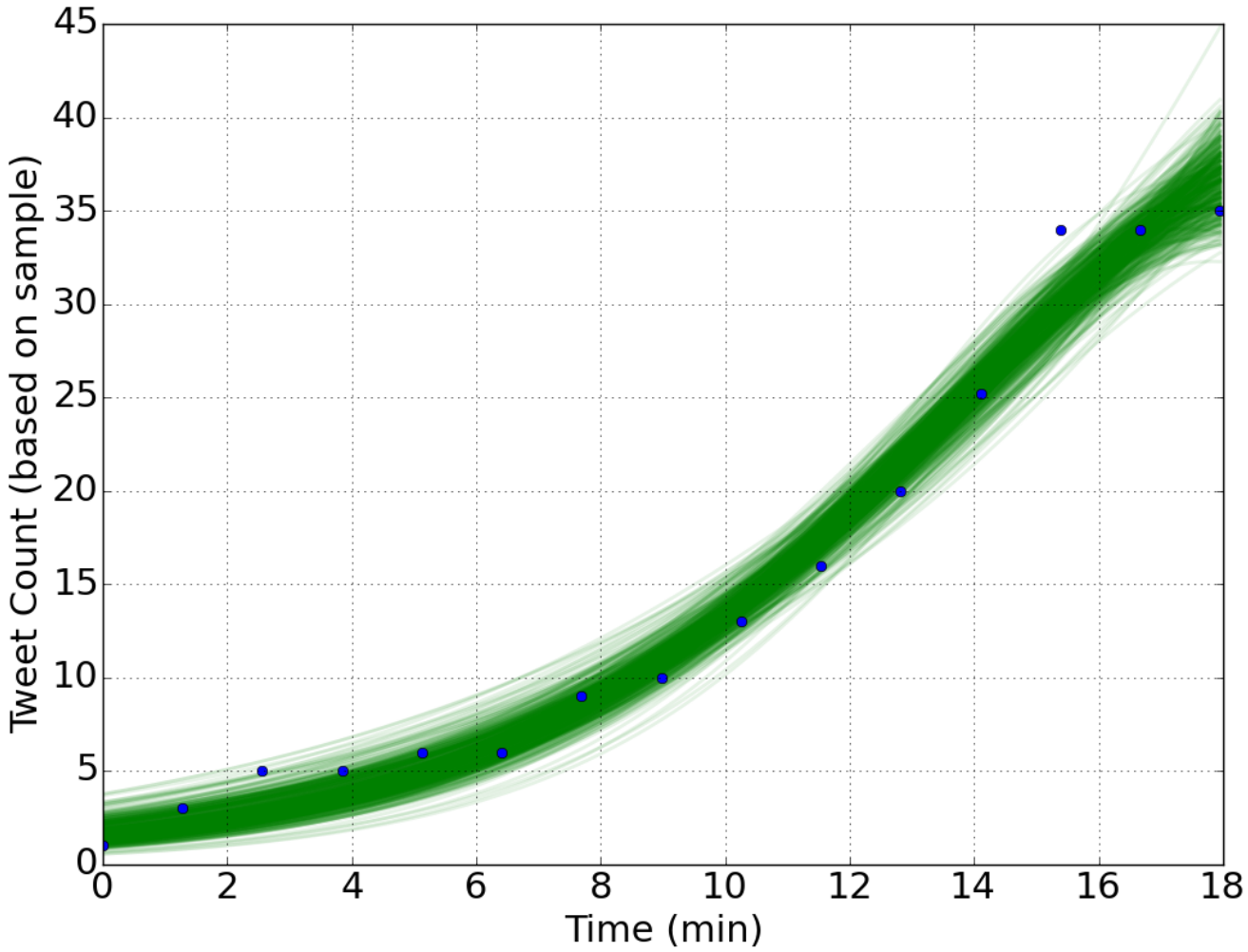


Comparison of Model Parameters



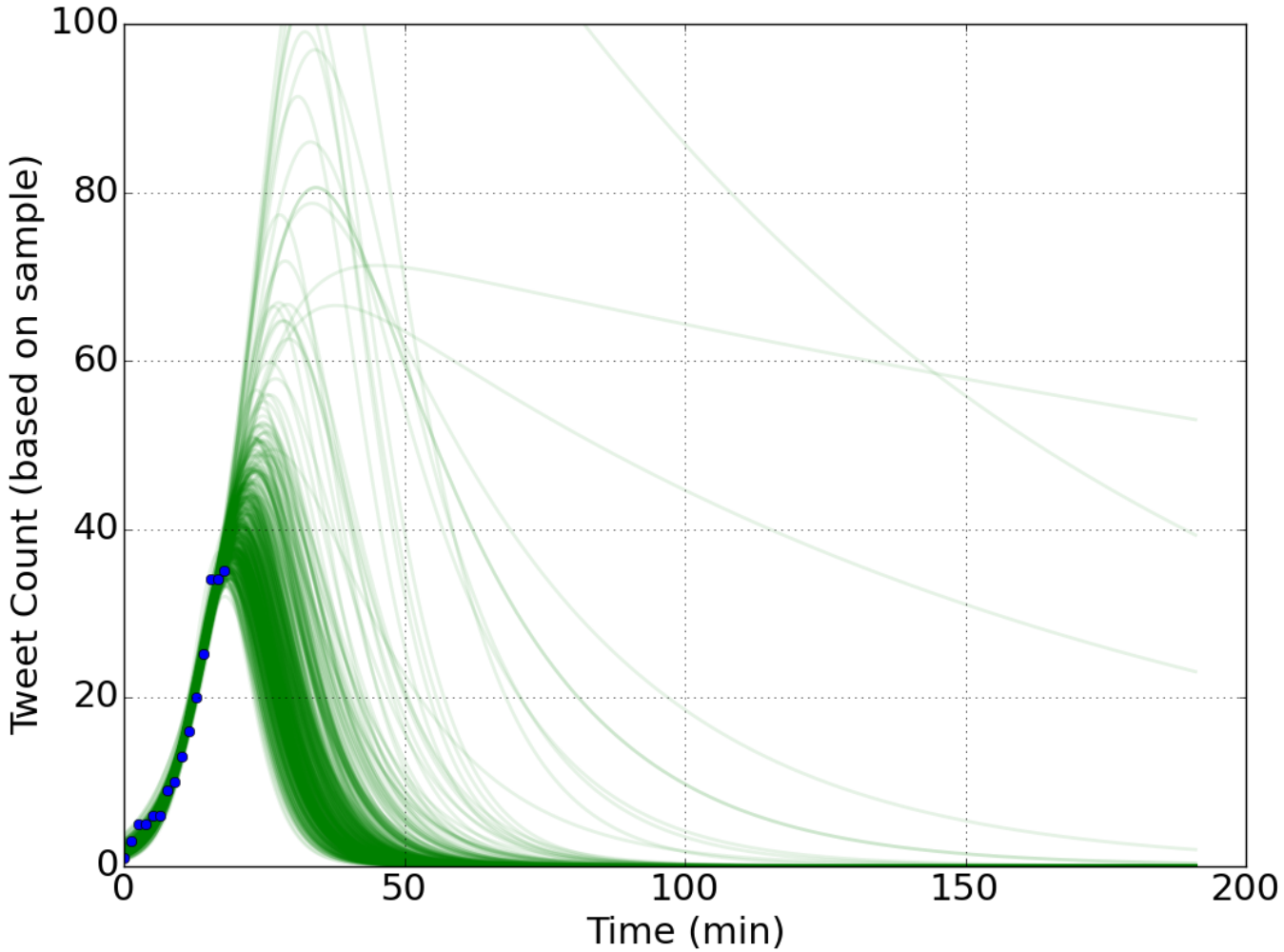
Prediction Using #Obama

Fit model to training set



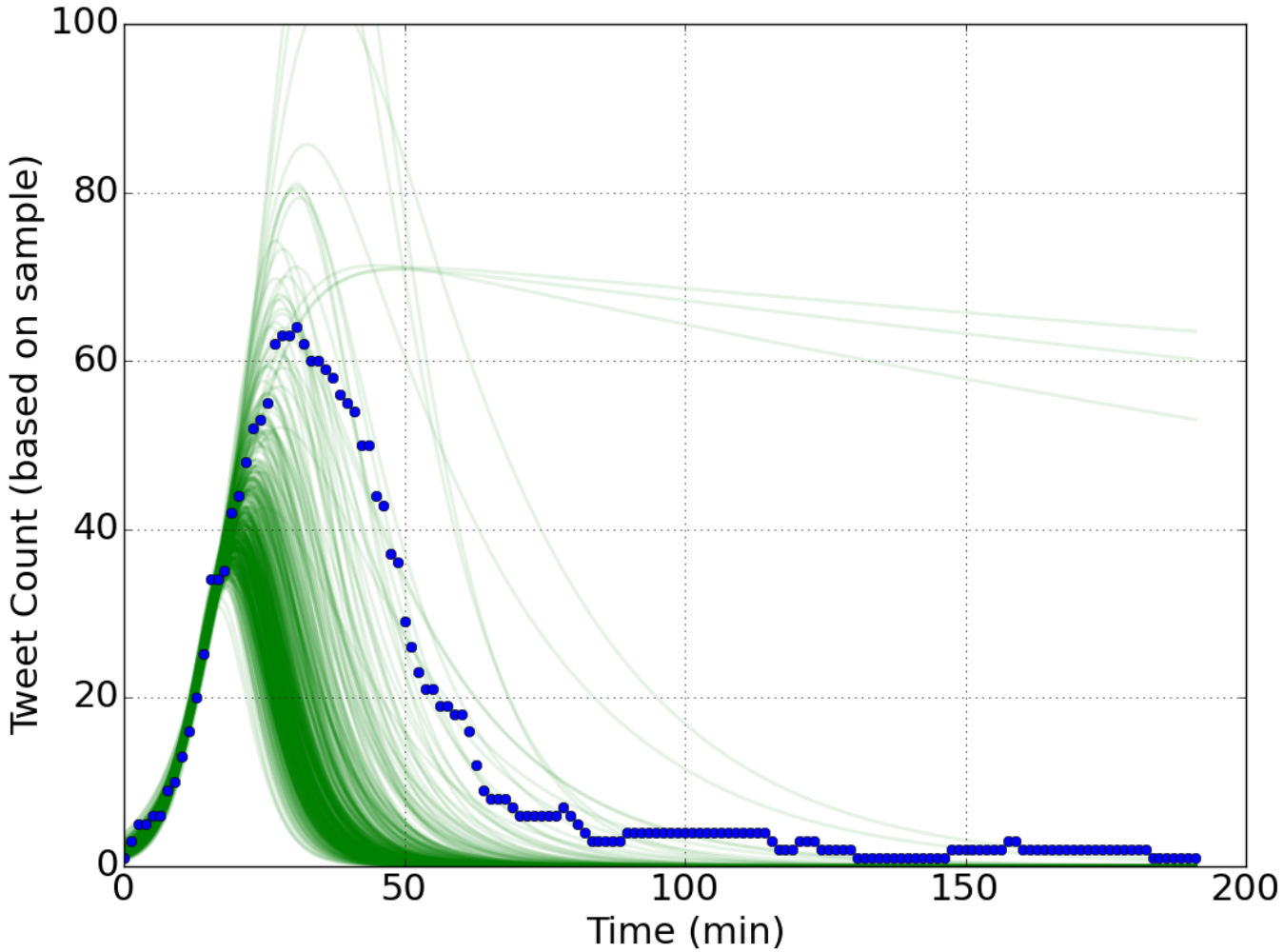
Prediction Results for #Obama

Run simulation over longer timescale



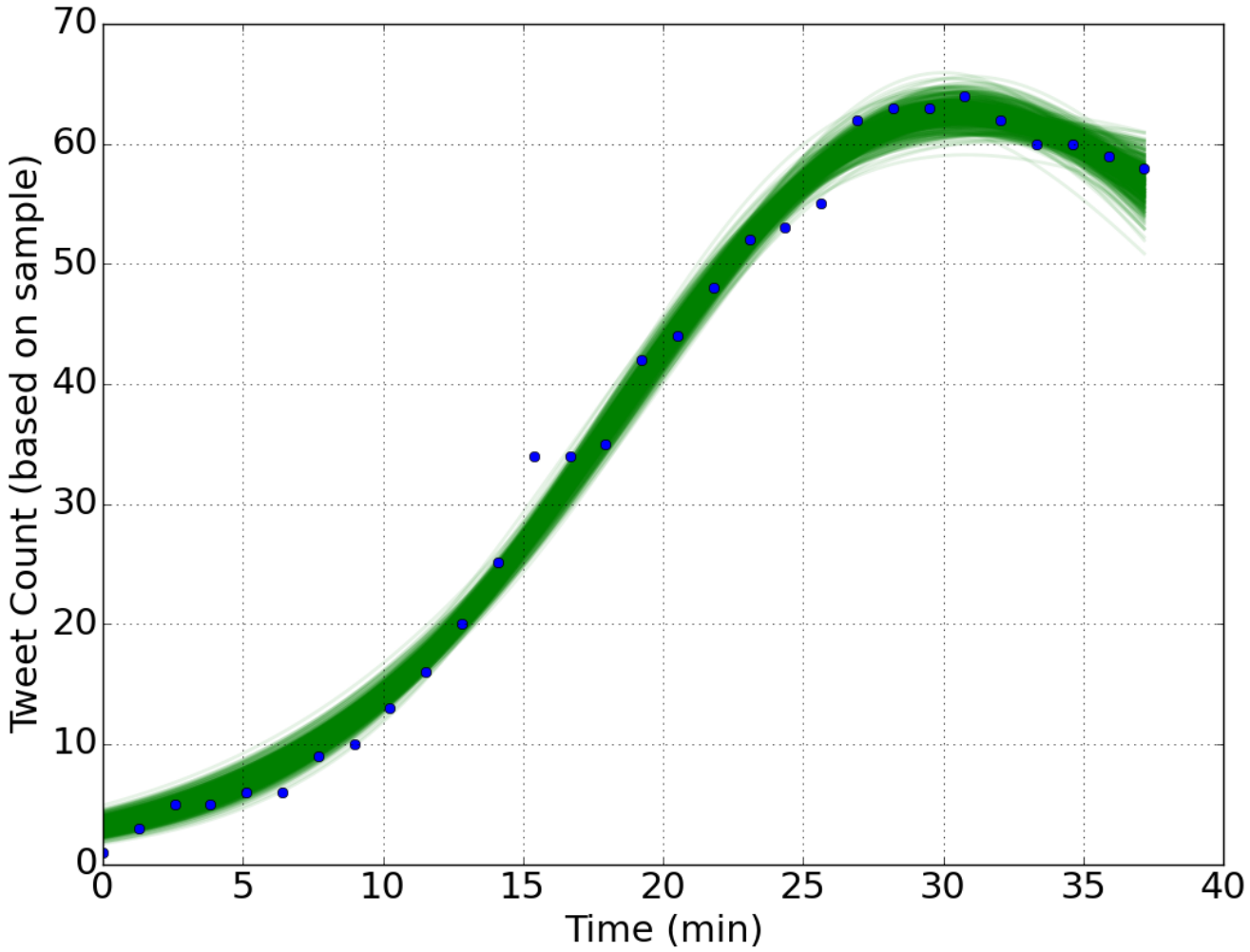
Prediction Results for #Obama

Comparison to actual (i.e., validation) data

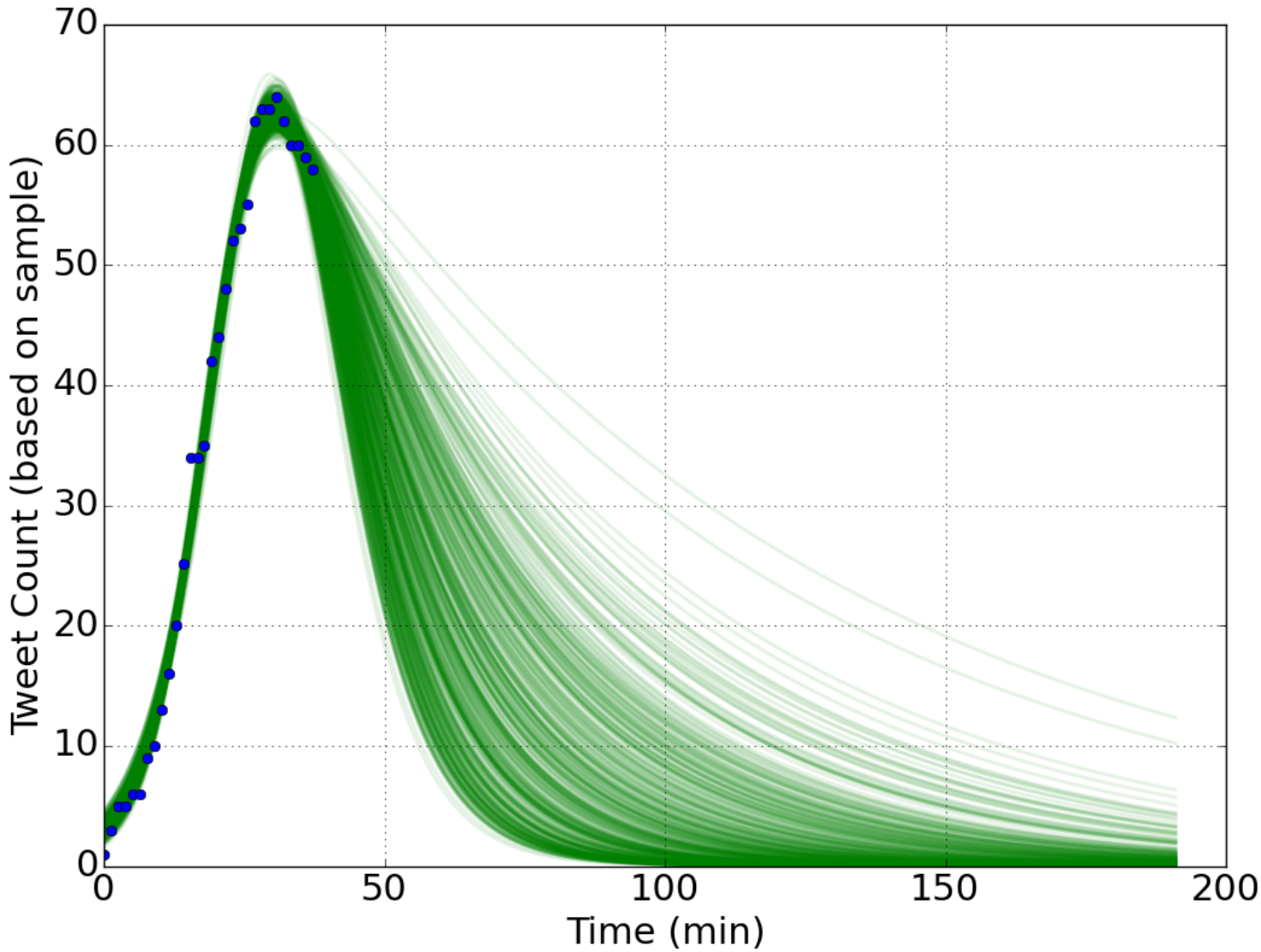


Prediction with a Larger Training Set

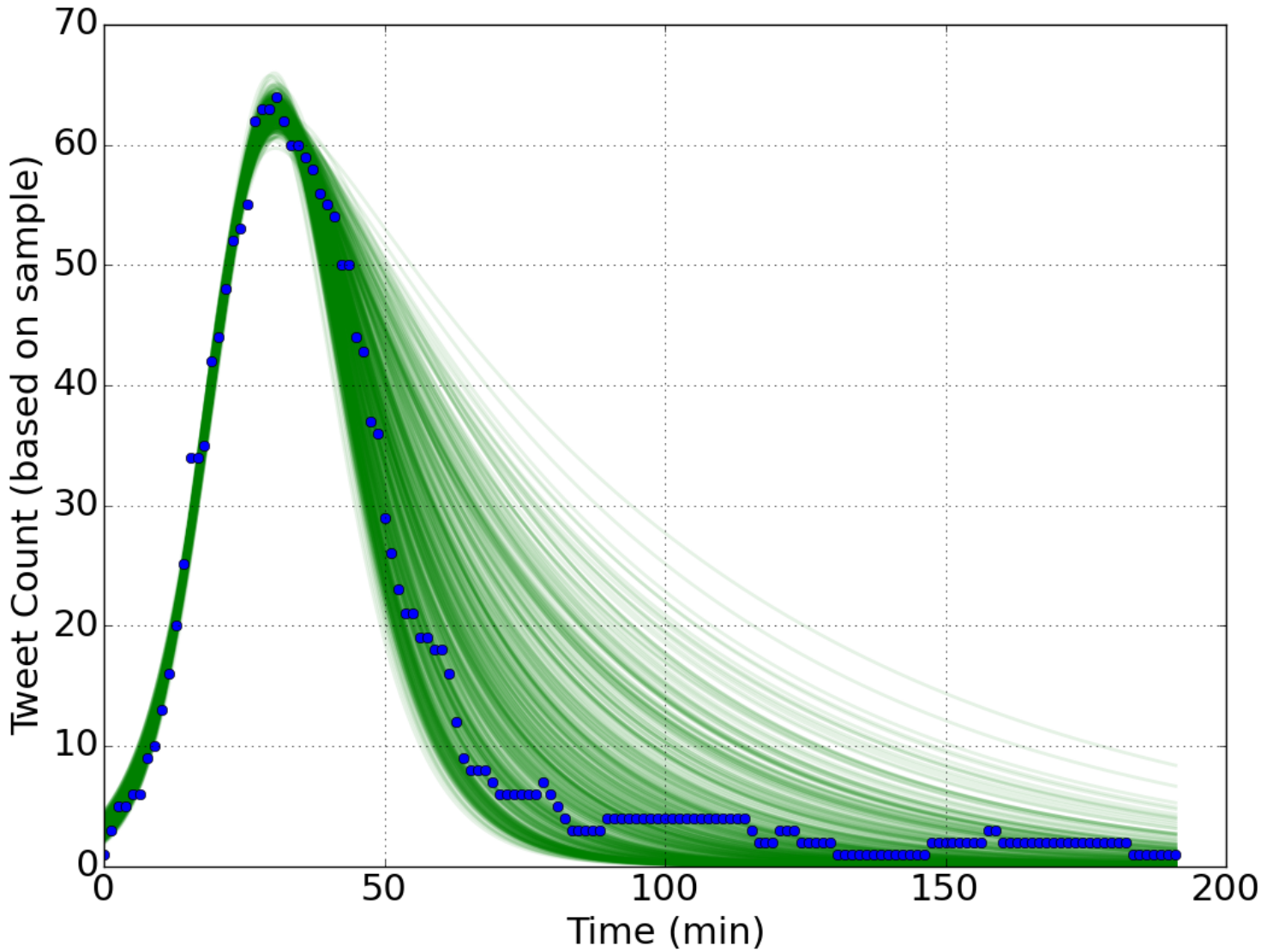
Notice that the training set now captures the peak



Prediction Results for #Obama



Prediction Results for #Obama



Potential Future Work

- Dynamical model tweaks

Potential Future Work

- Dynamical model tweaks
- Create better identification (i.e, hashtag selection) tools

Potential Future Work

- Dynamical model tweaks
- Create better identification (i.e, hashtag selection) tools
- Optimize “window” selection

Potential Future Work

- Dynamical model tweaks
- Create better identification (i.e, hashtag selection) tools
- Optimize “window” selection
- More prediction applications

Potential Future Work

- Dynamical model tweaks
- Create better identification (i.e, hashtag selection) tools
- Optimize “window” selection
- More prediction applications
- Develop interactive display

Potential Future Work

- Dynamical model tweaks
- Create better identification (i.e, hashtag selection) tools
- Optimize “window” selection
- More prediction applications
- Develop interactive display
- Compare to stochastic modeling strategy

Acknowledgments

Advisor: Brian Blais, PhD

Editorial Reviewer: John Quinn, PhD

Honors Program Coordinator: James Segovis, PhD

Questions?

Contact Info

Email: jonathan.skaza@gmail.com
Twitter: @SkazaSays