Simulating Hot Topic Popularity with a Modified SIR Model

Shane Storks

Collaborators: Martial Kuntz, Alex Marinkovski

Faculty Advisor: Dr. Na Yu

sstorks@ltu.edu

Lawrence Technological University

Southfield, MI

MAA MathFest Student Paper Session

July 27, 2017

Introduction

The interest in a trending topic can be modeled similarly to the spread of a contagious disease.

Figure 1: The random spread of ideas over the Internet (T3Leads).

We observed this when modifying the classic susceptible-infected-recovered (SIR) model developed by Kermack and McKendrick (1929) and applying it to Google Trends data.

Definitions

Define the following terms:

- **Hot topic:** the rise and fall of the popularity of a topic on the Internet which may last days to months. A particular topic may become a *hot topic* more than once over time.
- **Google Trends index:** a normalized value between 0 and 100 representing the relative number of searches for a topic at a specific time compared to the rest of a time range.

Assumptions

To make the problem easier to describe, make the following assumptions:

- Multiple searches by a user for the same topic are negligible in Google Trends data.
- Consequently, the Google Trends index of a topic is proportional to the number of users searching Google for that topic.

Consider the population of all users who will ever search Google for a particular hot topic. Divide this into three subpopulations S, I, and R:

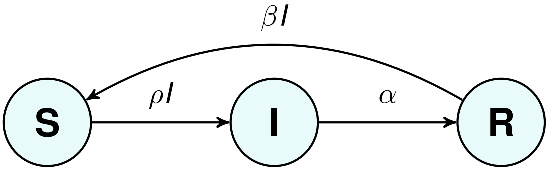

Figure 2: A flowchart for the dynamics of the susceptible (S), *interested* (I), and recovered (R) populations.

$S + I + R = 100$ in this closed population after rescaling to fit the range of the Google Trends index.

Model (cont.)

At time t for a specific hot topic,

$$\frac{dS}{dt} = -\rho SI + \beta IR \tag{1}$$

$$\frac{dI}{dt} = \rho SI - \alpha I \tag{2}$$

$$\frac{dR}{dt} = \alpha I - \beta IR \tag{3}$$

where

- $S(t) \propto$ users *still unaware of* the topic or *thinking about* the topic,
- $I(t) \propto$ users *searching for* the topic (i.e., interested in it), and
- $R(t) \propto$ users *no longer thinking about* the topic.

Model (cont.)

At time t for a specific hot topic,

$$\frac{dS}{dt} = -\rho SI + \beta IR \tag{1}$$

$$\frac{dI}{dt} = \rho SI - \alpha I \tag{2}$$

$$\frac{dR}{dt} = \alpha I - \beta IR \tag{3}$$

When users move between populations,

- $\rho I =$ the *gaining-interest* rate,
- $\alpha =$ the *losing-interest rate*, and
- $\beta I =$ the *remembrance rate*.

Due to the influence of word-of-mouth, mainstream media, and popular culture, virtually all gains in interest are proportional to the number of users already interested in a topic ($I(t)$).

Testing the Model

We tested three examples by using the Fmincon global search optimization solver in MATLAB with Google Trends data for each topic.

They show that the model can cover different kinds of past and present hot topics, including but not limited to

- historical events,
- Internet memes, and
- popular songs.

Past Example: Flint Water Crisis

In late 2015, it was revealed that the water supply in Flint, MI had been tainted with lead partially due to negligence by public officials. This humanitarian incident was a national hot topic in the following months.

Figure 3: Flint Water Plant water tower (Osorio).

We attempted to fit a curve to the erratic rise and fall in popularity of the topic.

Past Example: Flint Water Crisis

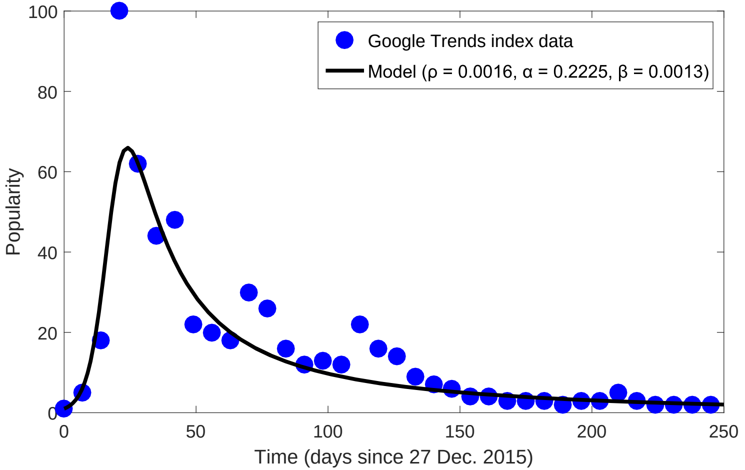

Figure 4: According to Google Trends, public interest in the topic has died down.

Recent Example: ”Catch Me Outside”

In January 2017, 13-year-old Danielle Bregoli appeared on *Dr. Phil* and became an Internet meme for her outrageous behavior and use of the phrase ”catch me outside.”

Figure 5: Danielle Bregoli (Collins).

In April 2017, we fitted a curve to the data that predicted the end of the topic’s decline. Updating the data in July 2017, we observed that the curve predicted the data well.

Recent Example: ”Catch Me Outside”

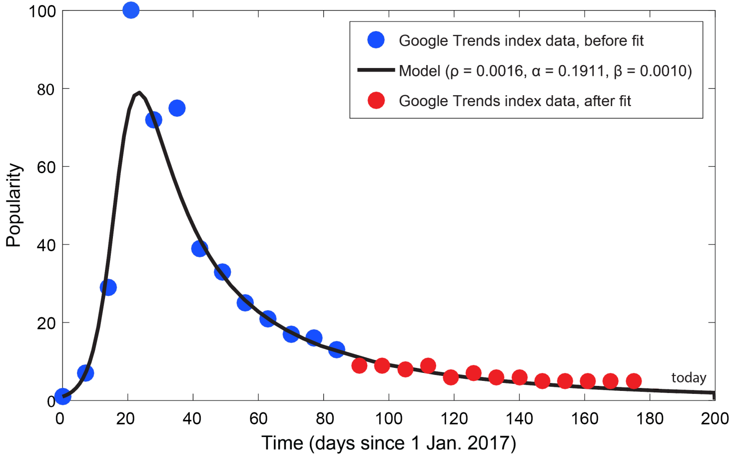

Figure 6: Calculated parameter values show that interest in Danielle Bregoli grew at the same rate as the Flint water crisis.

Current Example: ”Bad Things”

”Bad Things” was a popular song during the time of the project, by artists Machine Gun Kelly and Camila Cabello.

Figure 7: ”Bad Things” cover artwork (Redfearn).

In April 2017, we fitted a curve to incomplete data to predict that the song would no longer be popular by August 2017. In July 2017, we collected more data and observed that this prediction was holding true.

Current Example: "Bad Things"

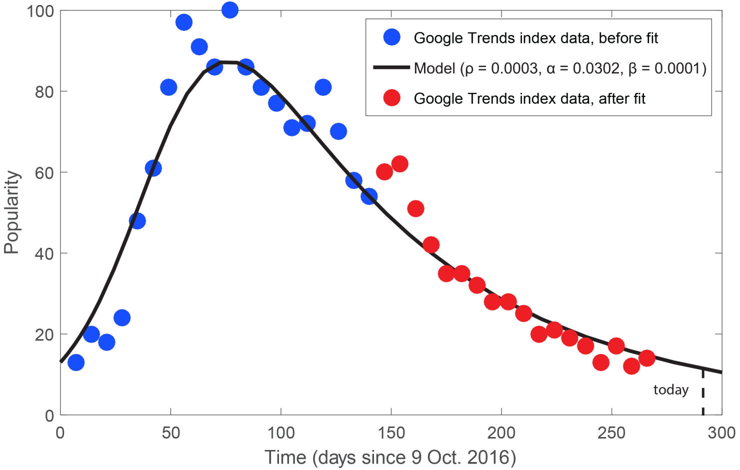

Figure 8: Additional data gathered in July 2017 supports the predicted decline of the song's popularity over the summer of 2017.

Further Improvements: Stochastic Model

The fitted curves do not represent erratic regions of the data as well, as seen in the Flint water crisis example.

One way this can be mitigated is by adding a stochastic process to the losing-interest rate α:

$$\frac{dS}{dt} = -\rho SI + \beta IR \quad (1)$$

$$\frac{dI}{dt} = \rho SI - (\alpha + \sigma\xi)I \quad (2)$$

$$R = 100 - S - I \quad (3)$$

Where

- $\xi =$ low-pass filtered white noise with mean $= 0$ and std $= 1$, and
- $\sigma =$ noise intensity.

All other parameters are the same as in the deterministic model.

Stochastic Simulation: Flint Water Crisis

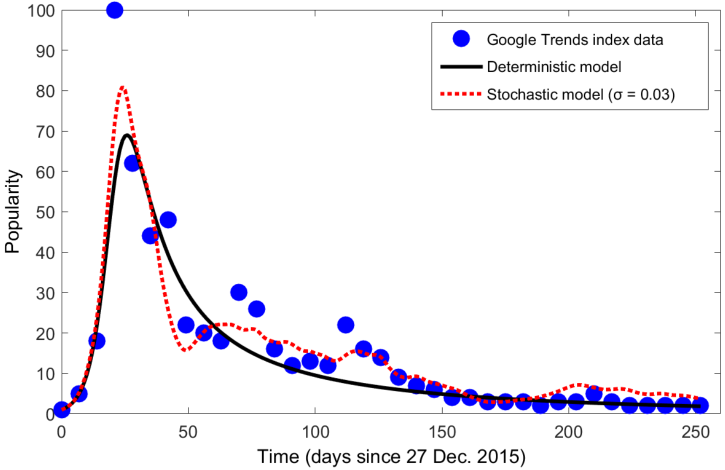

Figure 9: The stochastic model better simulates the fluctuation of the data.

Conclusion

Although the propagation of hot topics over the Internet is a random process, our numerical analysis demonstrates that it can be simulated and predicted by our simple, deterministic model quite well.

Though it can be improved to cover more rare types of interactions and trends, the simple version accurately represents the majority of situations when modeling the popularity of hot topics online.

References

1. Collins, D. (2017). Figure 5: Danielle Bregoli. Retrieved from https://www.thesun.co.uk/news/3063240/cash-me-ousside-girl-danielle-bregoli-dr-phil-show-argument/

2. Kermack, W.O. & A.G. McKendrick. (1927). Contributions to the mathematical theory of epidemics-I. *Proceedings of the Royal Society of London, 115(772)*, 700–721. Retrieved from http://www.math.utah.edu/ bkohler/Journalclub/kermack1927.pdf

3. Osorio, C. (2016). Figure 3: Flint Water Plant water tower. Retrieved from http://www.npr.org/sections/thetwo-way/2016/03/23/471585633/independent-investigators-state-officials-mostly-to-blame-for-flint-water-crisis

4. Redfearn, D. (2016). Figure 7: ”Bad Things” cover artwork. Retrieved from http://www.billboard.com/articles/news/7540998/machine-gun-kelly-camila-cabello-bad-things-single-art

5. Storks, S., Kuntz, M., Marinkovski, A., and Yu, N. (2017, April). Simulating hot topic popularity with the SIR model. Poster presented at Lawrence Technological University’s seventh annual Research Day, Southfield, MI.

6. T3Leads. (2014). Figure 1: The random spread of ideas over the Internet. Retrieved from http://www.billboard.com/articles/news/7540998/machine-gun-kelly-camila-cabello-bad-things-single-art