

Social trends and the Simmel effect

Max Pellert

University of Konstanz

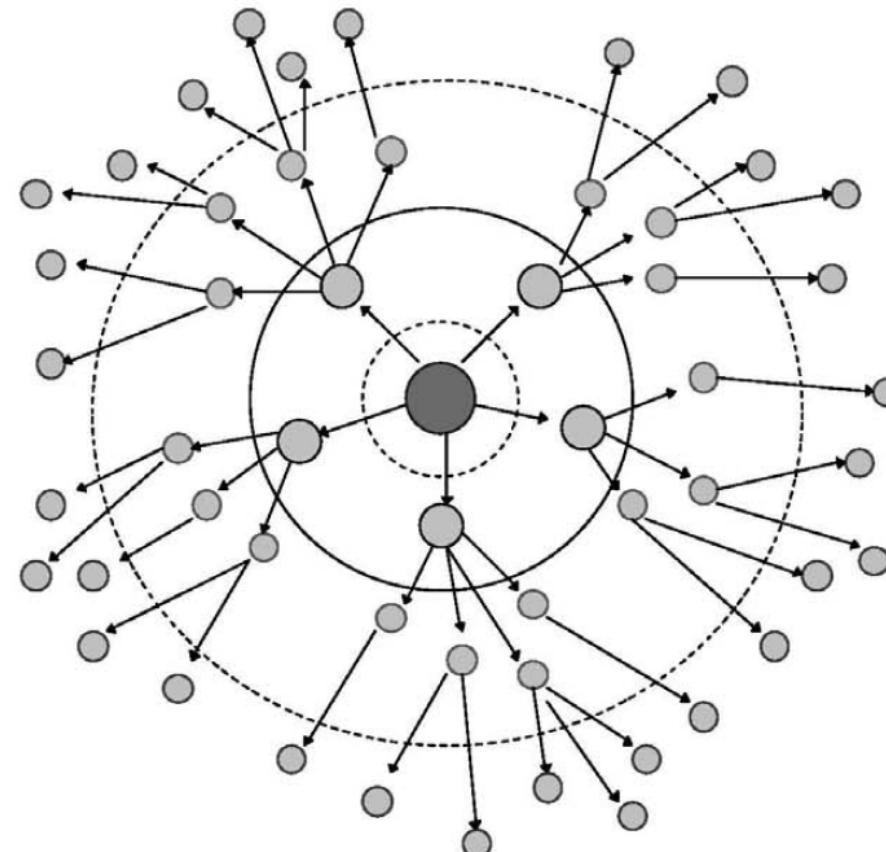
Social Media Data Analysis

Outline

1. The Simmel effect
2. Baby names: Big old trend data
3. Online trend dynamics
4. Bootstrapping

Trends and social impact

The question of trends: How does social impact aggregate in a society?

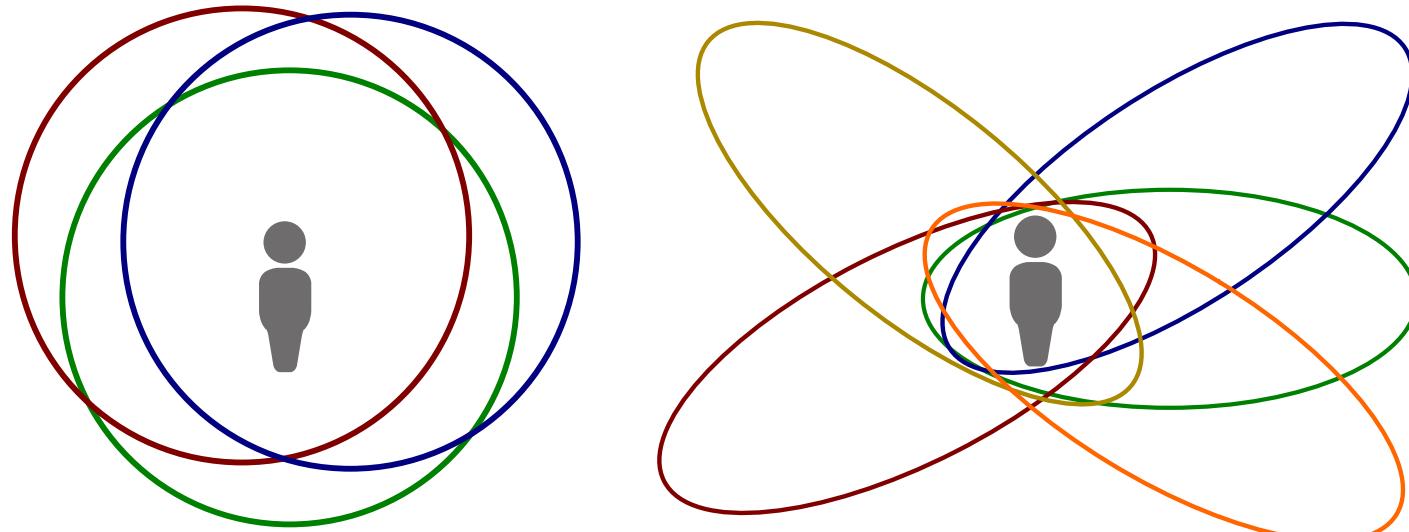


Hipsters and fashion



Oblivious to the paradox of their uniform individuality" (Dan Ashcroft, Nathan Barley)

Georg Simmel



- One of the fathers of Sociology
- Studied how social structures vary across societies
- Formulated theories on how premodern and modern societies differ

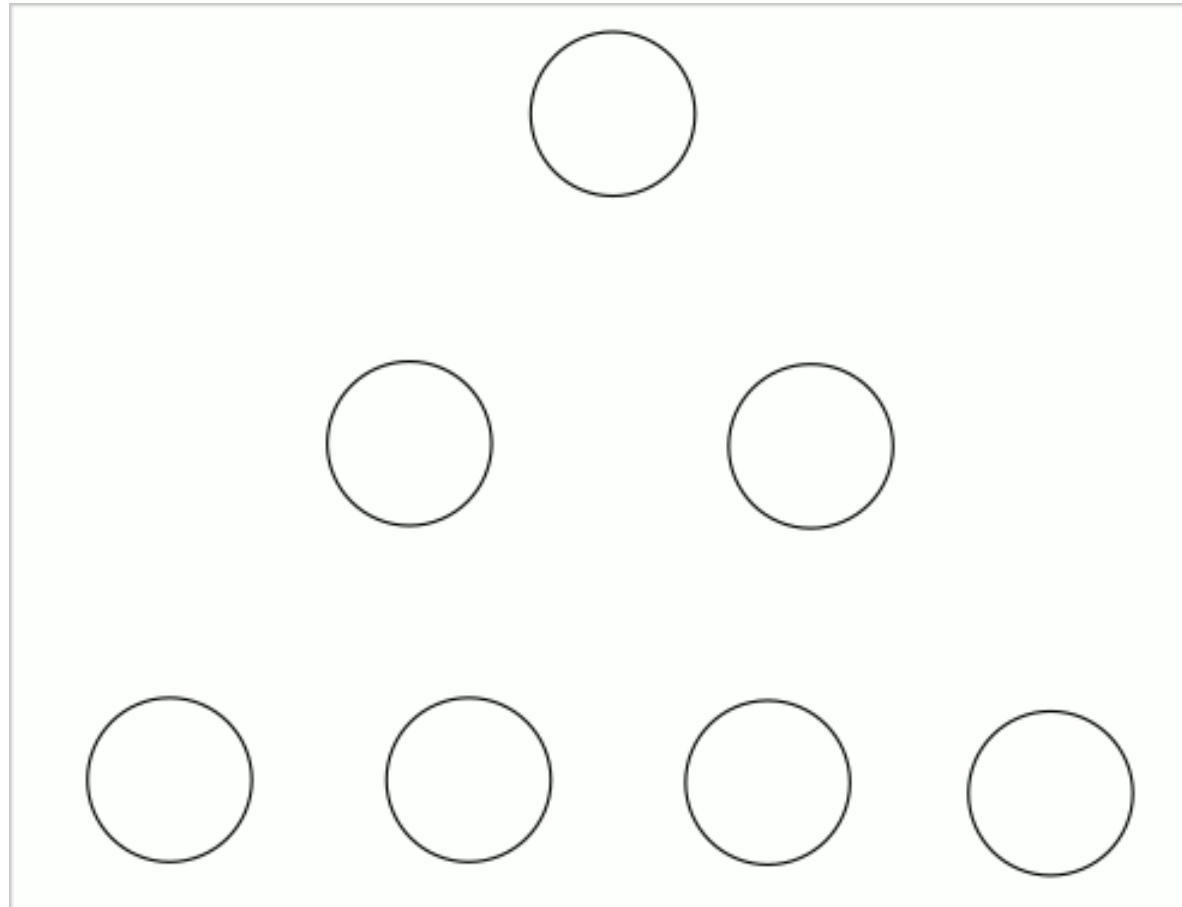
Theory of fashion: the Simmel Effect

Georg Simmel defined fashion as the *non-cumulative change in cultural features*, where cultural features are displayed as *status symbols*. Status symbols are externally displayed traits associated with high social class, e.g. surnames, clothing, sport, food, etc.

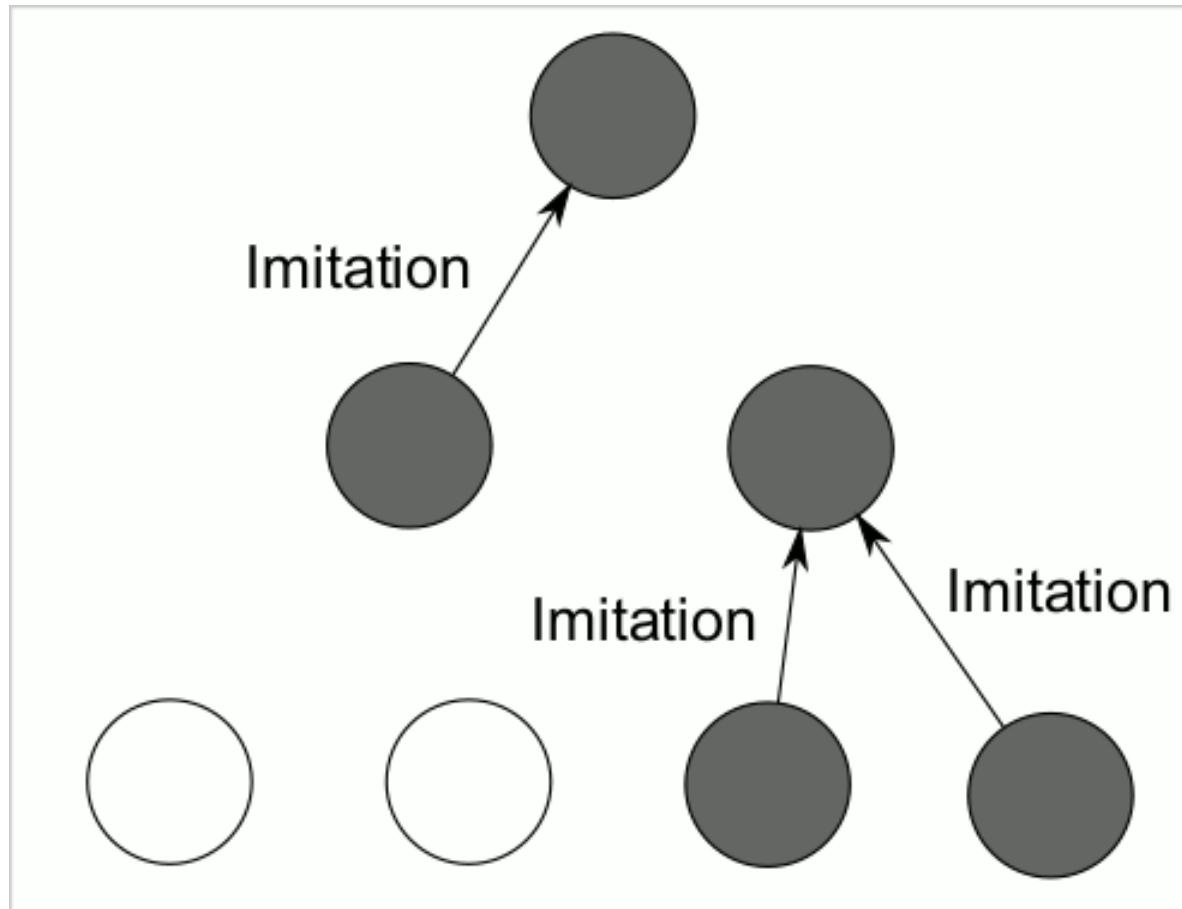
The Simmel effect: The persistence of social differences under the instability of status symbols

Simmel noticed that fashions come and go, but fashion is always present. When something becomes popular, it is bound to lose its popularity. Simmel introduced this theory in his 1904 article "Fashion", describing observations that are still relevant, such as how going against fashion is a way to acknowledge its relevance (the hipster paradox).

The mechanisms of Simmel's theory: Imitation

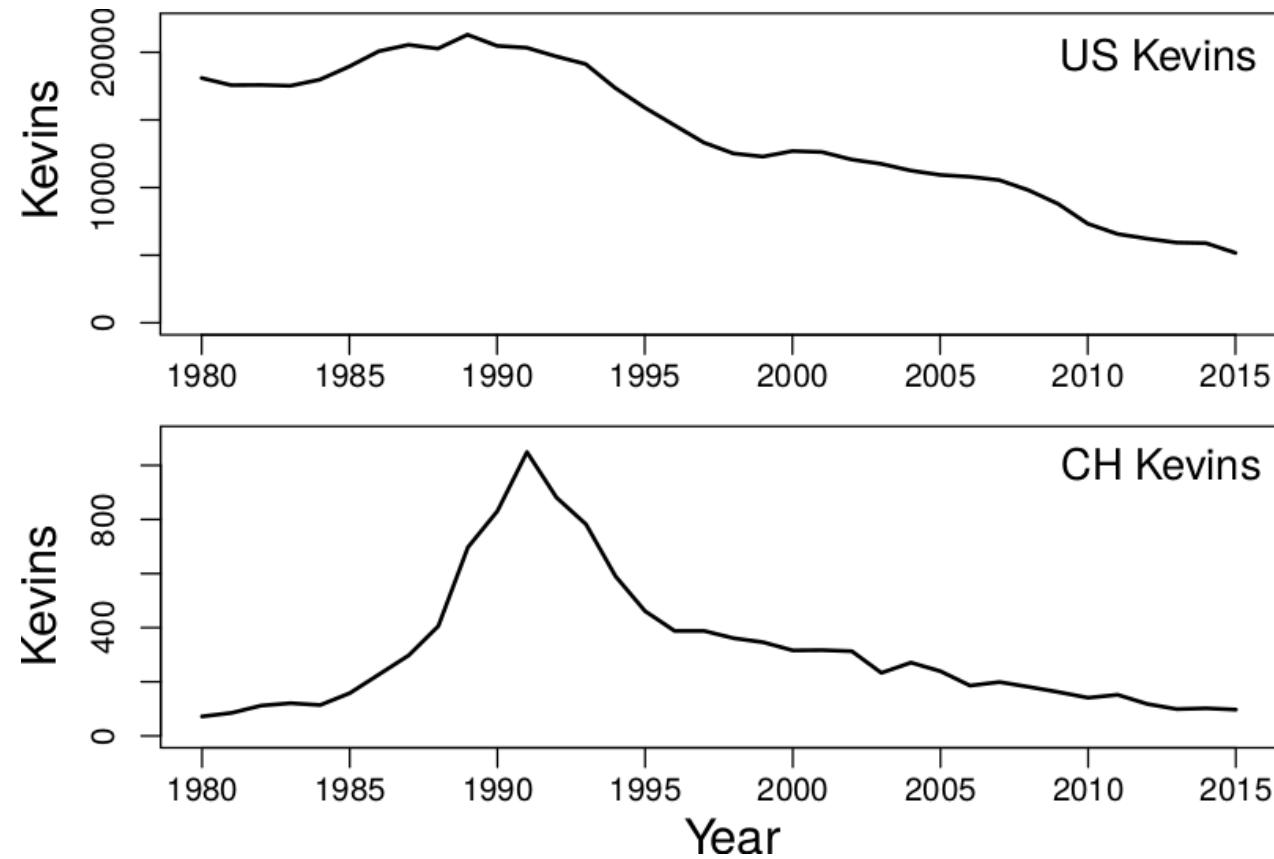


The mechanisms of Simmel's theory: Differentiation





The case of baby names



First names can be status symbols and carry subjective and social values.
Copying the name of your baby from someone else is an example of imitation.

Baby names: Big old trend data

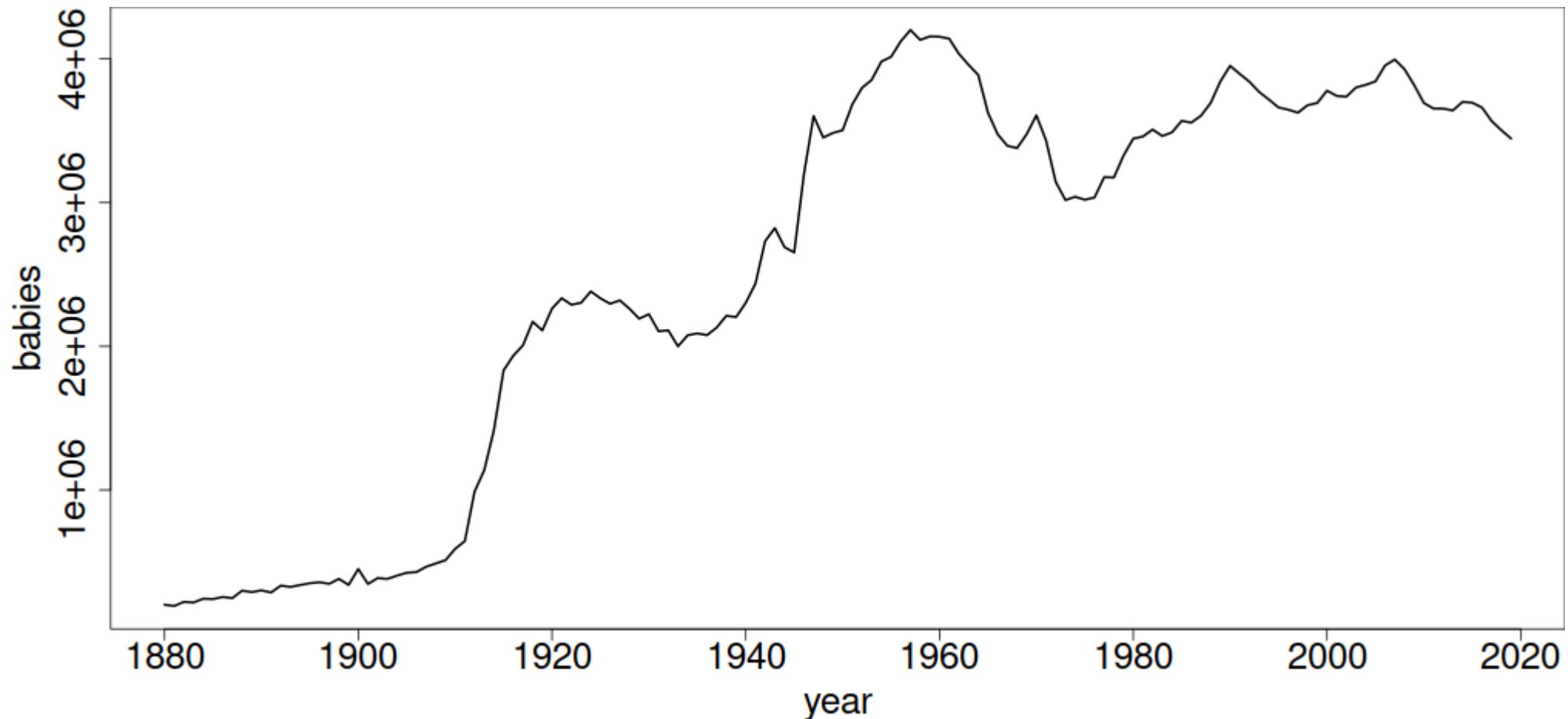
1. The Simmel effect

2. *Baby names: Big old trend data*

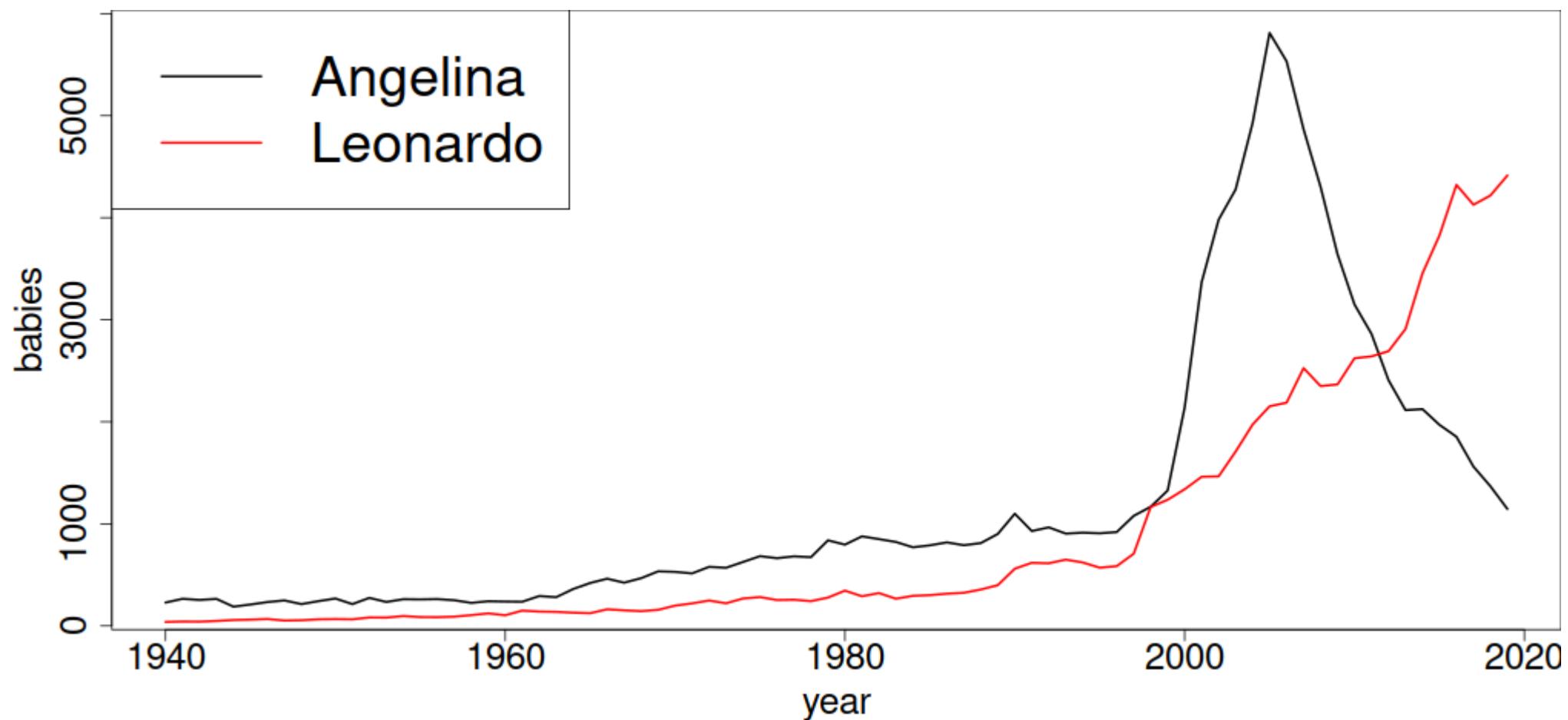
3. Online trend dynamics

4. Bootstrapping

US SSA baby name data

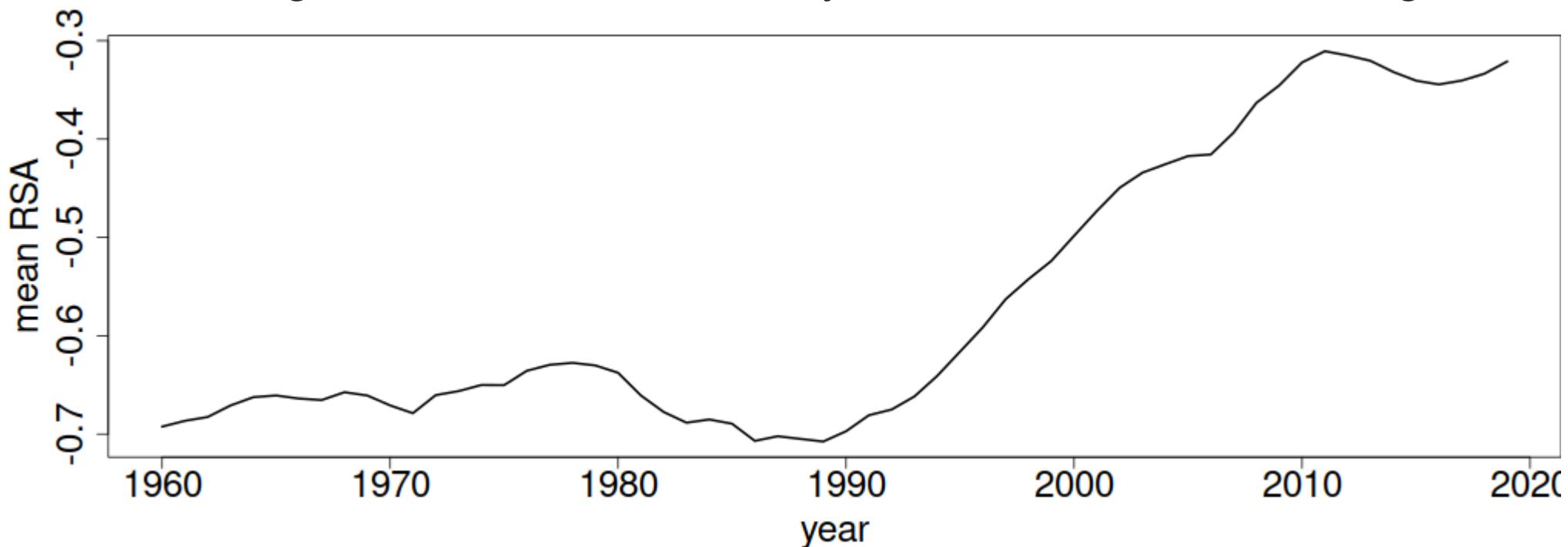


Baby name trend examples

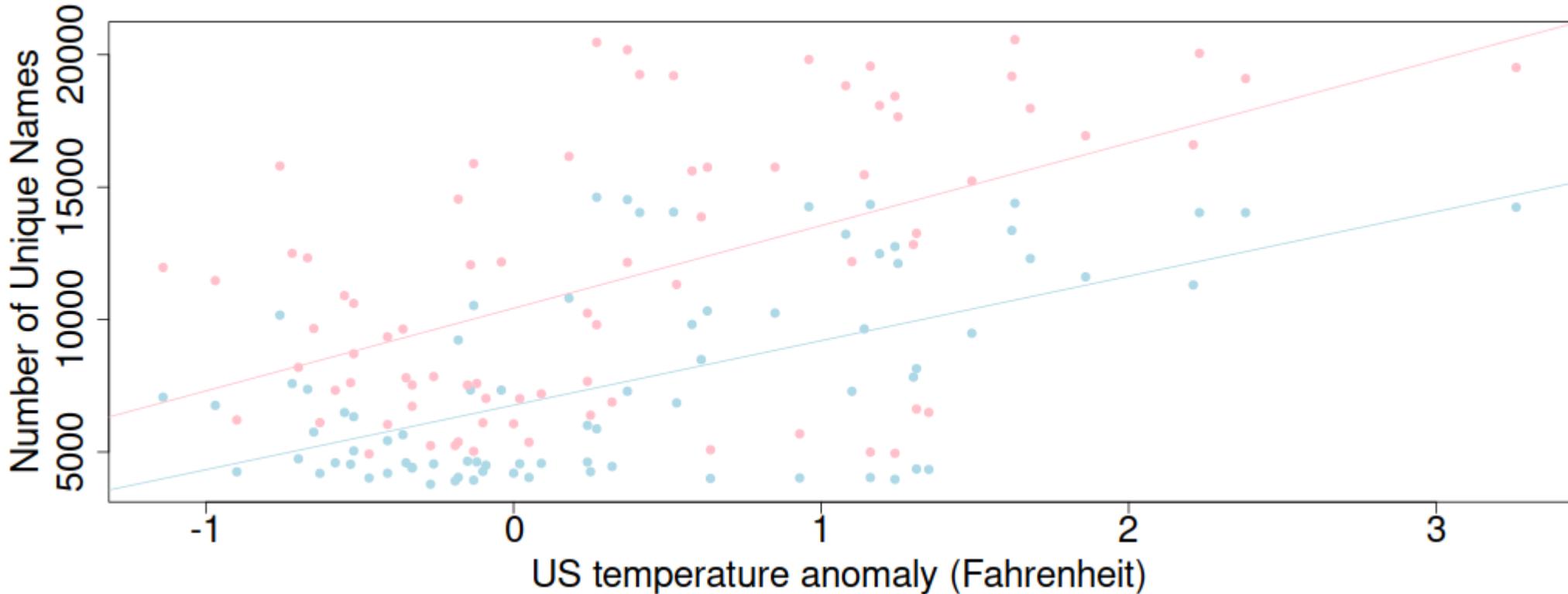


The QWERTY effect in baby names

The QWERTY effect is a hypothesis in Psychology that postulates that words that are written with more right-hand letters of the keyboard are, on average, more positive than words that are written with more left-hand letters of the keyboard. The fraction of right-hand letters in US baby names has been increasing:



Wacky baby name research



A parody paper titled "We are entering an unprecedented age in baby name flux" reported; "baby name diversity also seems to have risen with the increasing annual temperature of the US (i.e., climate change)".

The limits of baby name predictability

Baby names are a popular example to illustrate scientific topics. The book *Freakonomics* explains the imitation part of the Simmel effect and explains how people imitate their richer neighbors when naming their babies. The book goes as far as making a prediction of what will be the top US baby names in 2015, based on a data analysis exercise that is never explained in detail in the article. Here is the prediction:

MOST POPULAR GIRLS' NAMES OF 2015?

Annika	Eleanora	Isabel	Maya
Ansley	Ella	Kate	Philippa
Ava	Emma	Lara	Phoebe
Avery	Fiona	Linden	Quinn
Aviva	Flannery	Maeve	Sophie
Clementine	Grace	Marie-Claire	Waverly

MOST POPULAR BOYS' NAMES OF 2015?

Aidan	Bennett	Johan	Reagan
Aldo	Carter	Keyon	Sander
Anderson	Cooper	Liam	Sumner
Ansel	Finnegan	Maximilian	Will
Asher	Harper	McGregor	
Beckett	Jackson	Oliver	

Top 10 names in 2015 and 2004

topFemale2015	topFemale2004	topMale2015	topMale2004
Abigail	Abigail	Alexander	Andrew
Ava	Ashley	Benjamin	Christopher
Charlotte	Elizabeth	Ethan	Daniel
Emily	Emily	Jacob	Ethan
Emma	Emma	James	Jacob
Harper	Hannah	Liam	Joseph
Isabella	Isabella	Mason	Joshua
Mia	Madison	Michael	Matthew
Olivia	Olivia	Noah	Michael
Sophia	Samantha	William	William

Predicting is hard

There is not much overlap between the prediction and the results for 2015. Just using the 2004 list, you would have made a better prediction.

What you see is that predicting which names in particular will be the most popular is a very difficult task. The Simmel effect describes forces that create observable patterns, but that does not mean that the model is predictive to tell us which of all names will become popular ten years from now, even if we had data of the social status of parents.

This is the difference between explanatory and predictive power of a model. A model can explain phenomena without being useful to make predictions, as in this case, but can also be predictive without giving explanations, like in the case of deep learning or other black-box approaches.

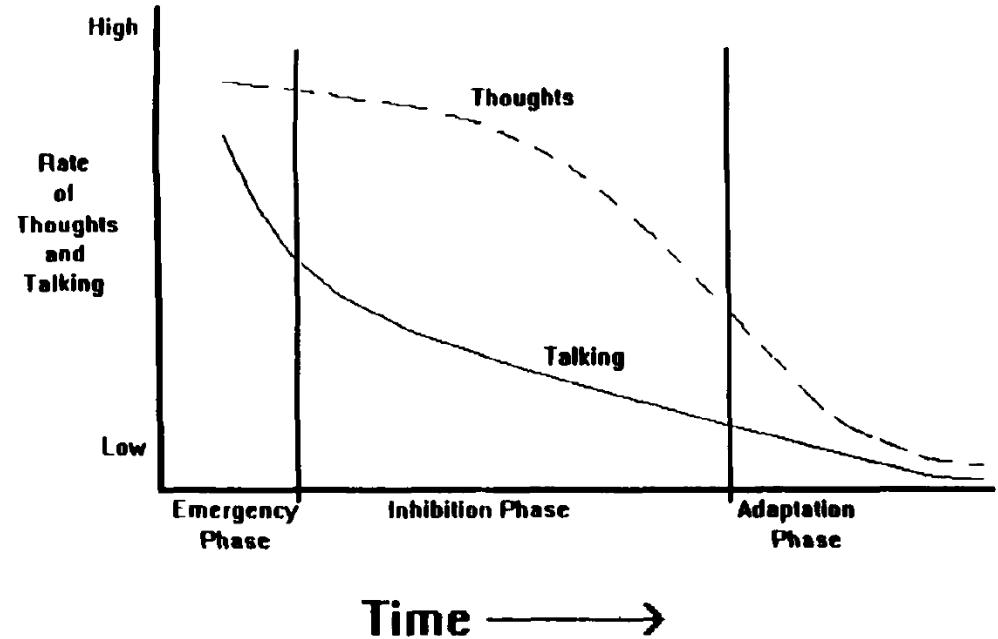
| **Take home message:** understanding does not imply predictive power

Online trend dynamics

1. The Simmel effect
2. Baby names: Big old trend data
3. *Online trend dynamics*
4. Bootstrapping

Social stage model of collective coping

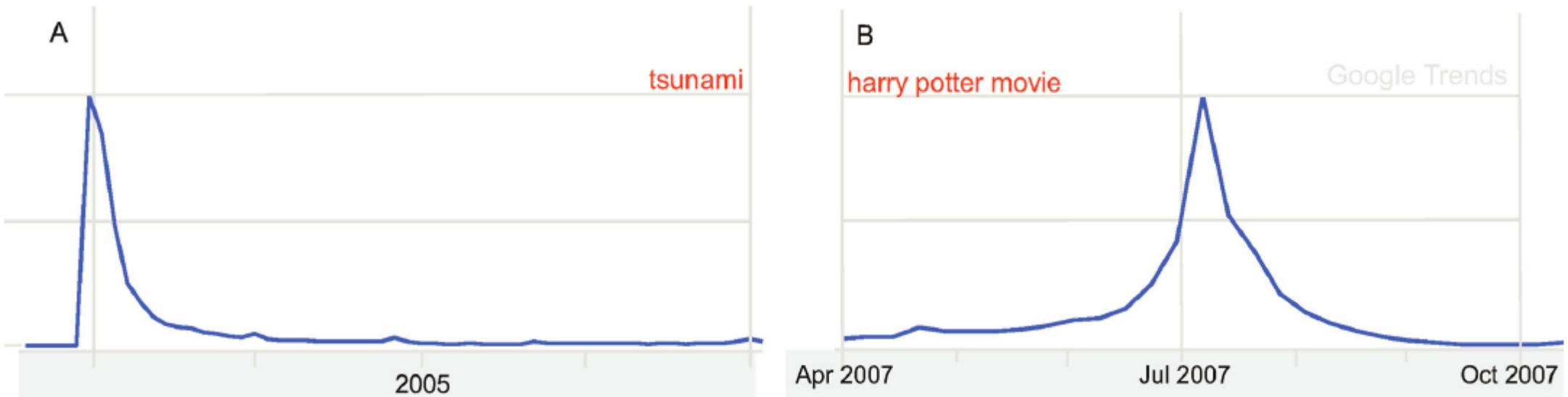
- People react to an external event (e.g emergencies)
- Trend of talking and thinking about the event
- **Emergency phase:** High levels of thinking and talking
- **Inhibition/satiation phase:** Fast decrease in talking, slower in thinking
- **Adaptation phase:** Both talking and thinking go back to baseline



A Social Stage Model of Collective Coping: The Loma Prieta Earthquake and The Persian Gulf War. J. Pennebaker and K. Harber. Journal of Social Issues (1993)

Social trends in online platforms

Google search trends can capture data about these large-scale social trends:



Exogenously triggered search volume after a tsunami (left) and endogenously driven search for the Harry Potter movie (right).

Exogenous versus endogenous bursts



- **Exogenous burst:** Spike created by an **external event**
 - Unexpected terrorist attack creates a very fast increase, the event is *external to the community*
- **Endogenous burst:** Peak driven by **social influence in the community**
 - Anticipation for a movie and word of mouth **within a community** creates a slower increase
- Exogenously triggered search volume after a terrorist attach (left) and endogenously driven search for the a Star Wars movie (right).

Subcritical and critical dynamics

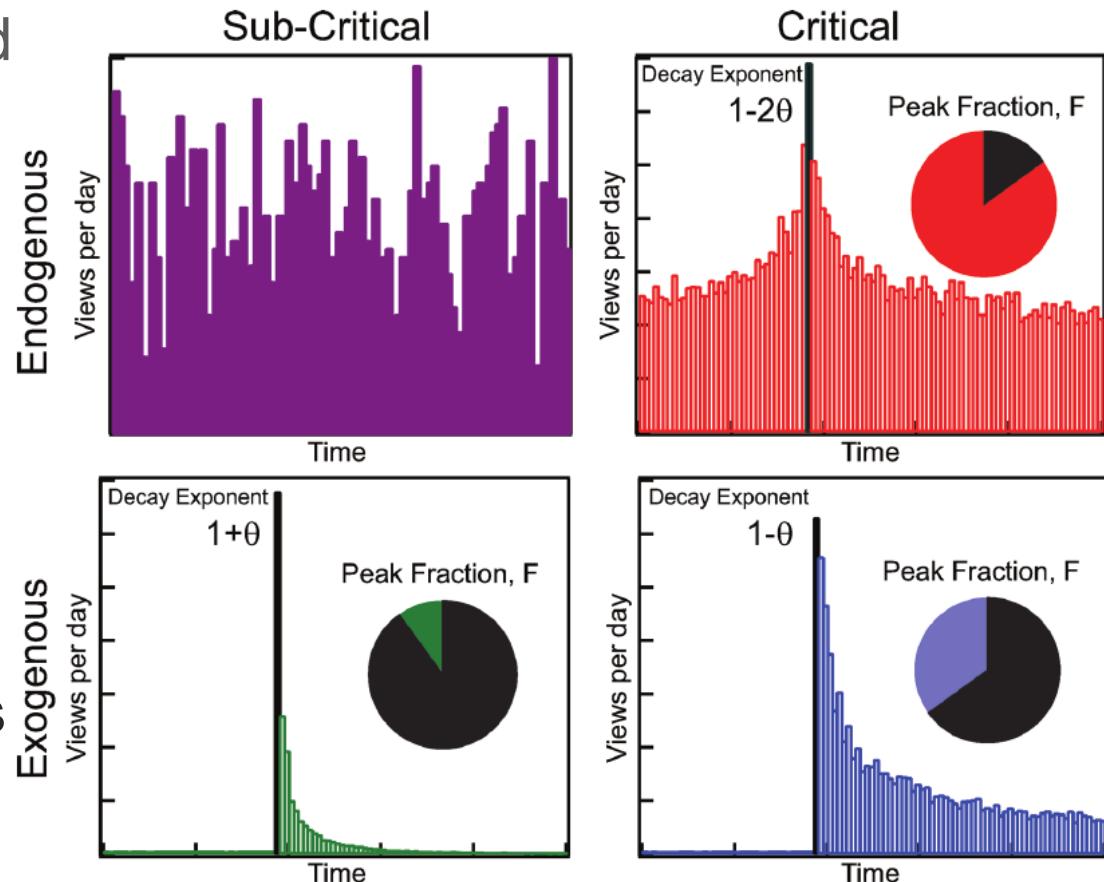


- Types of social trends after the peak of a collective aggregate of a community
- Examples with Google trends volume
- **Subcritical dynamics:** Fast relaxation due to **social influence weaker than novelty decay**
 - Interest in the US about the attacks was limited, the peak relaxed fast
- **Critical dynamics:** Slow relaxation due to **social influence stronger than novelty decay**
 - Hype about Star Wars movie kept people talking and searching about it

The endo-exo model

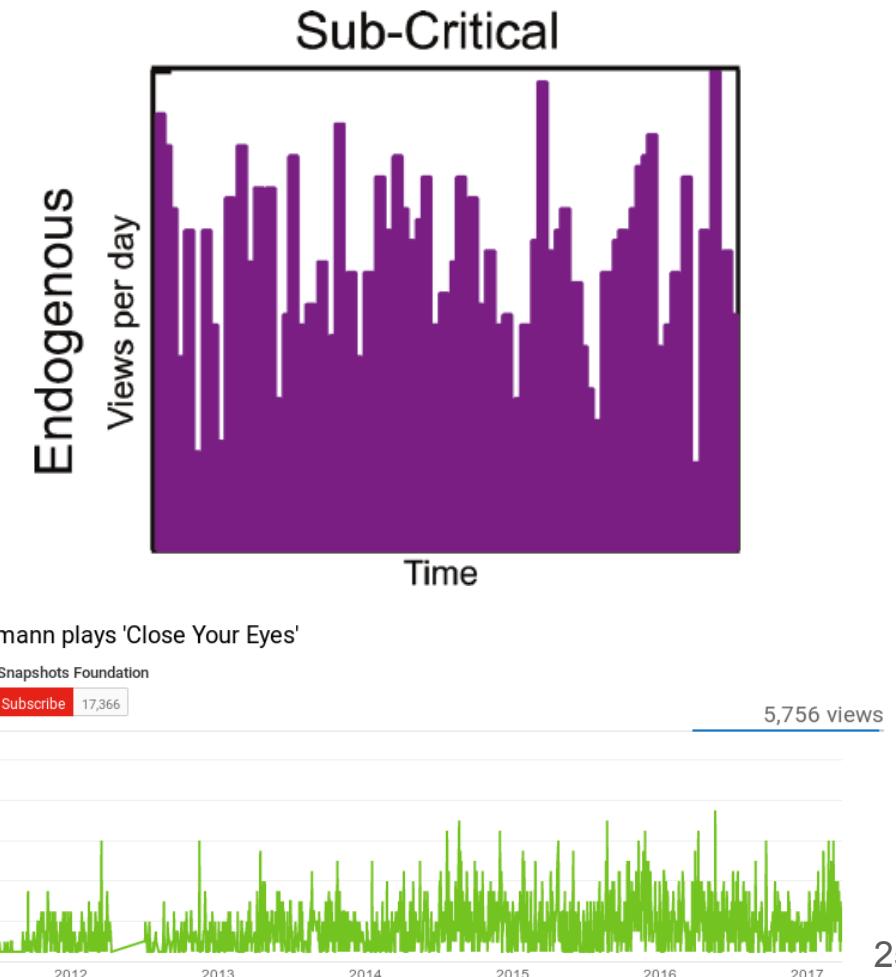
The endo-exo model of Riley Crane and Didier Sornette captures these types of dynamics. In this model, a trend can have two properties:

- It can have an **exogenous trigger** when a central event influences lots of people at the same time, as in the tsunami example.
- It can be **critical** when the social interaction between individuals leads to further responses and it is stronger than the rate of losing interest.



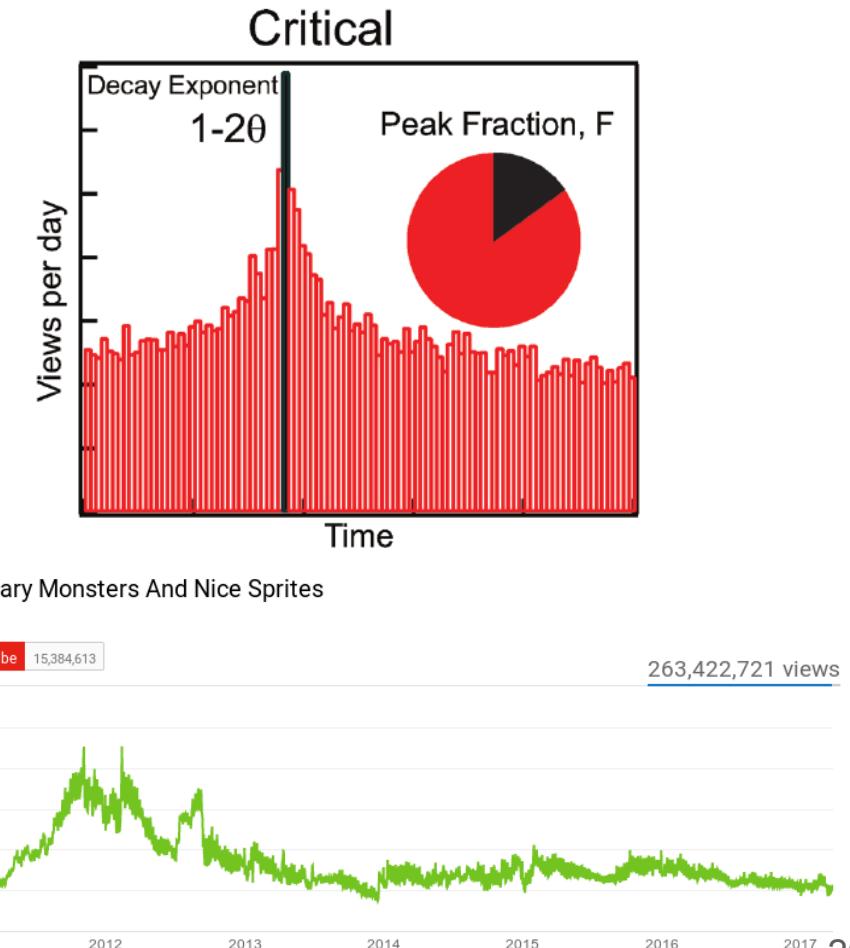
Endogenous subcritical

- **Endogenous:** absence of strong spike
- **Subcritical:** social influence absent or weaker than novelty decay
- **Trend:** noisy signal of incoherent activity
 - No identifiable peak
 - Example: most of content in social media



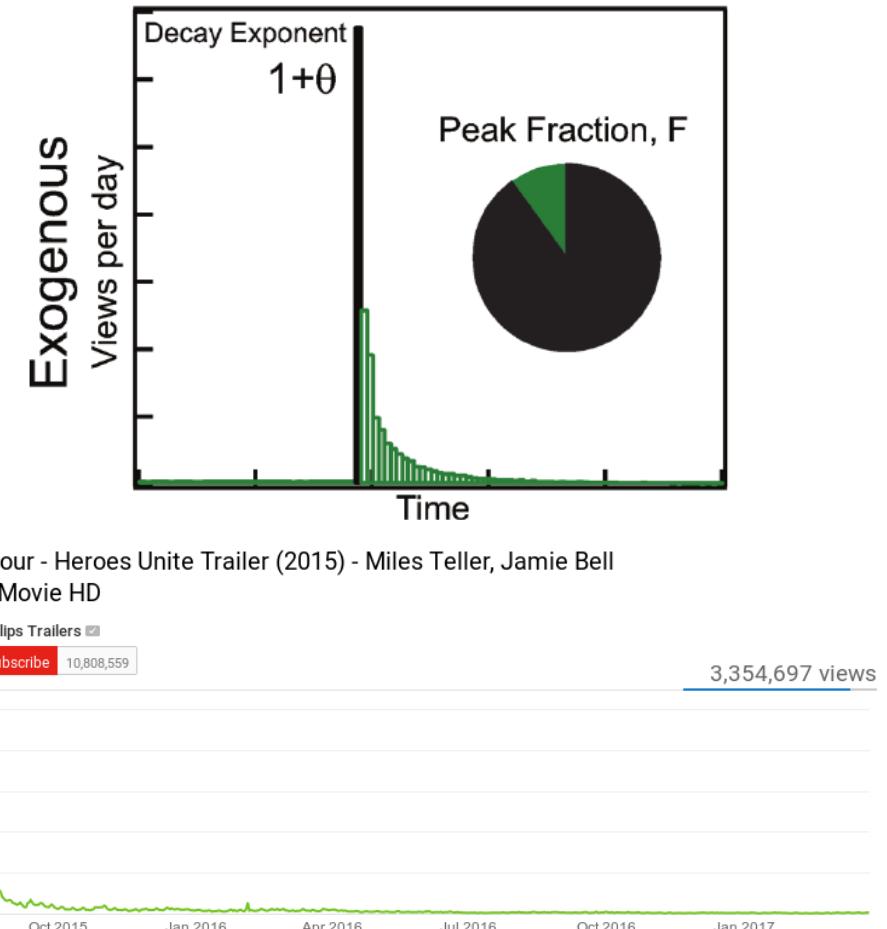
Endogenous critical

- **Endogenous:** peak present without spikes
- **Critical:** social influence stronger than novelty decay
- **Trend:** large wave with slow increase and decrease
 - Less than 20% total volume at the peak
 - Example: content that becomes popular in social media



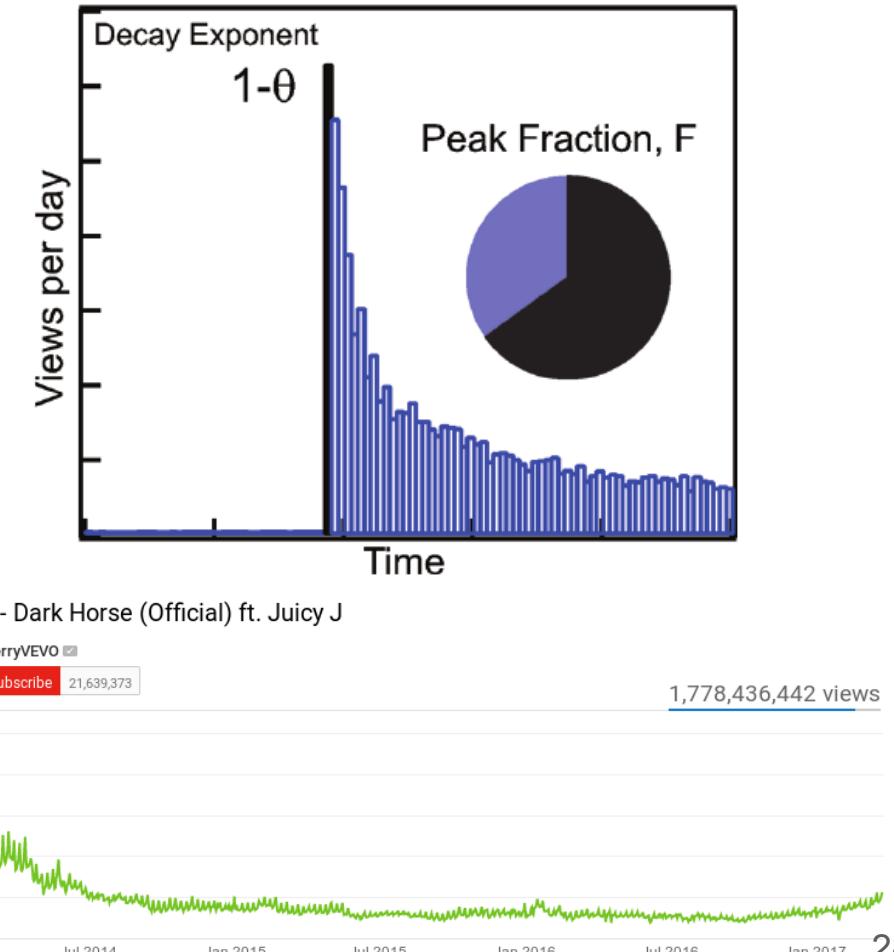
Exogenous subcritical

- **Exogenous:** peak as spike produced by an external event
- **Subcritical:** Social influence weaker than novelty decay
- **Trend:** sudden growth and decrease
 - More than 80% total volume at the peak
 - Example: junk content featured in mass media



Exogenous critical

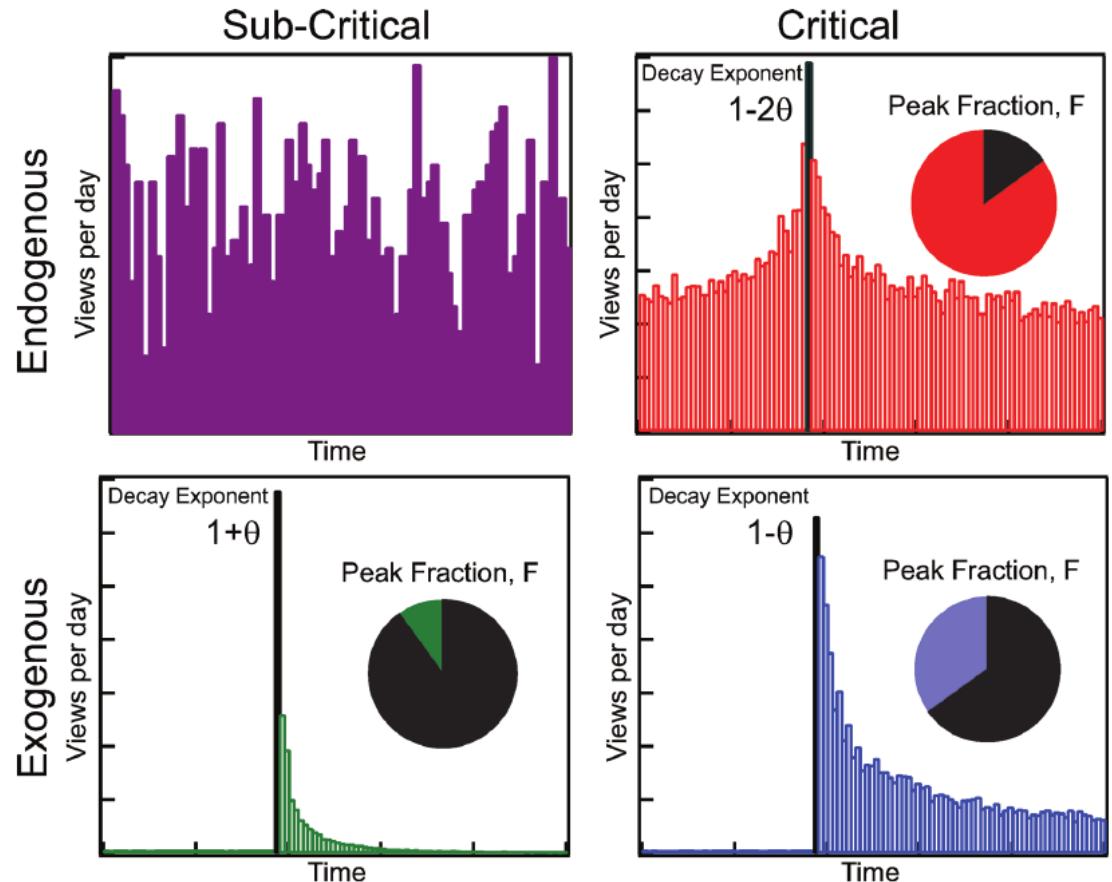
- **Exogenous:** Spike generated by external factor
- **Critical:** Social influence stronger than novelty decay
- **Trend:** Spike growth with slow decay
 - Between 20% and 80% total volume at the peak
 - Example: Central channels creating successful popular content



Peaks in the endo-exo model: summary

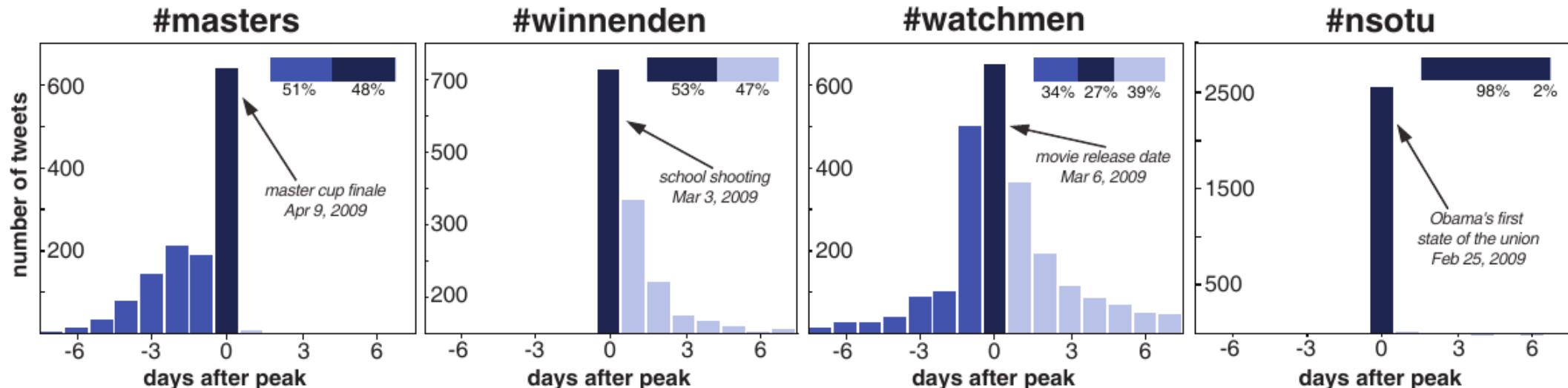
These two properties are not exclusive, leading to four types of responses:

1. **Endogenous sub-critical:** no clear peak, absence of trend.
2. **Endogenous critical:** "viral" peak driven by word of mouth.
3. **Exogenous sub-critical:** sharp peak but fast decay due to lack of strong social interaction.
4. **Exogenous critical:** sharp peak but slow decay due to strong interaction after shock.



Trends on Twitter

This model has been applied to classify Twitter hashtag trends by Lehmann et al. The figure shows activity volumes related to three of the classes in the model.

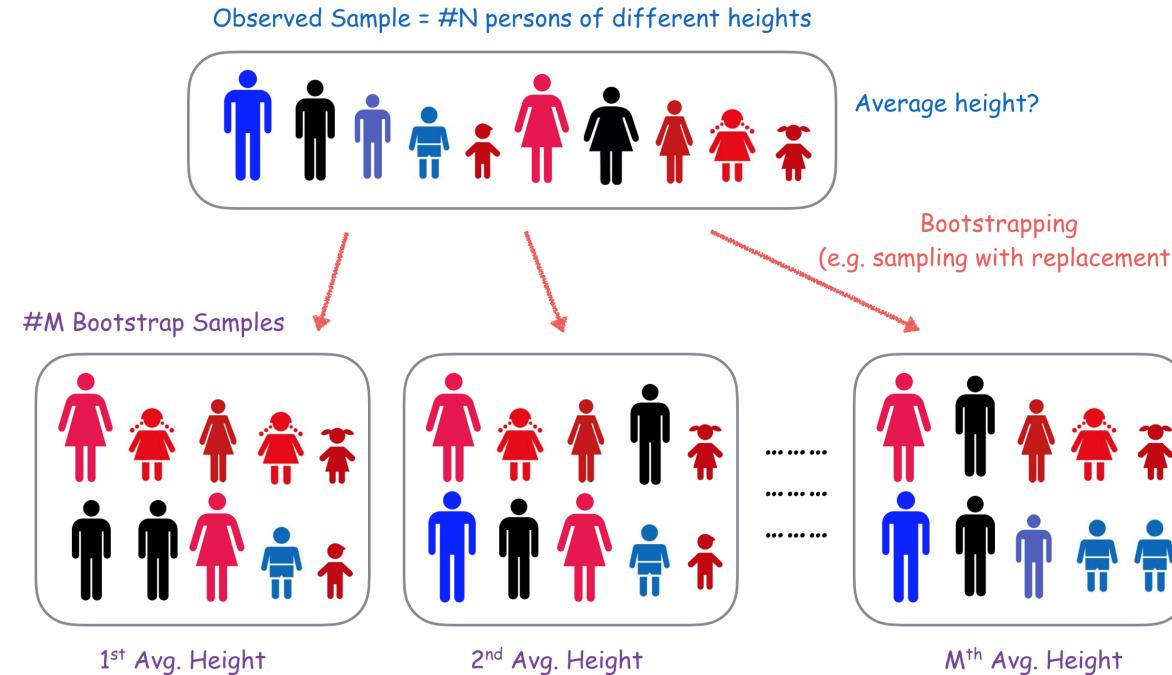


Gathering this kind of volume data is best done by using the Twitter API. You will learn how to do this from Python in the course exercises.

Bootstrapping

1. The Simmel effect
2. Baby names: Big old trend data
3. Online trend dynamics
4. *Bootstrapping*

Assessing uncertainty via bootstrapping



Bootstrap simulates the sampling of our original data.

To bootstrap, we resample from our sample by generating new samples of the same size as our original sample **with replacement**

Example: the mean of a variable

First, we load the data and measure the mean over our sample. We will use the height Data from National Health Interview Survey (NHIS) 2007. The first few heights in the dataset look like this (inches):

74 70 61 68 66 98

With the mean height over the sample being 69.5782654 inches.

To generate a bootstrap sample we should resample once with replacement and compute the mean over the resulting sample. An example of bootstrap mean is 69.7490073 inches.

Running the resampling

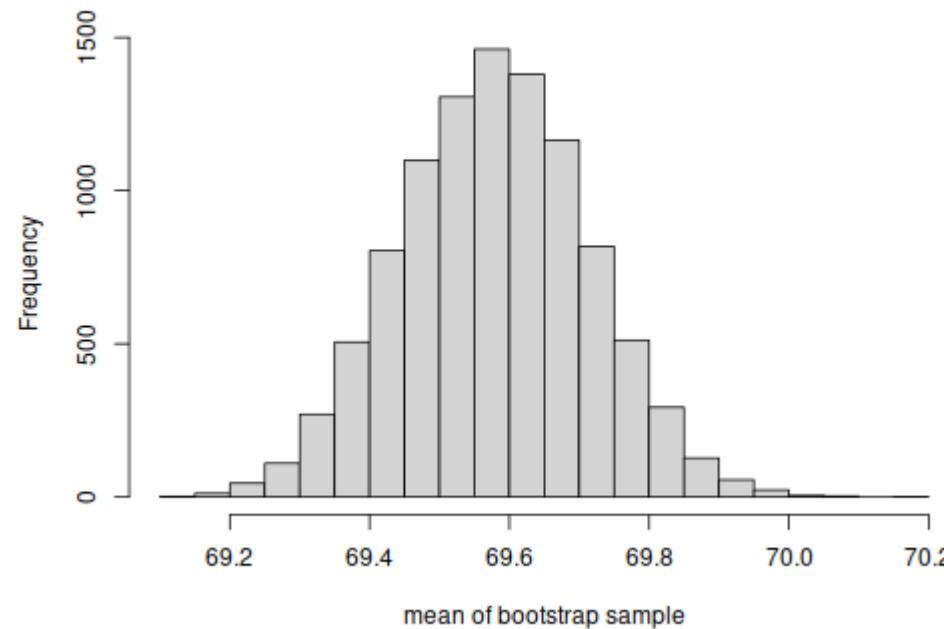
You will see that this mean value is slightly different as the first one. We have simulated what would be another sample from the total population, assuming that the first sample was representative of the total population.

Now we can repeat the bootstrap sample and measurement a lot of times, here we do it 10000 with a loop, saving the results in a vector. The vector will look like this:

```
69.67712 69.51181 69.6558 69.67377 69.68422 69.39791
```

Over the results we can now see the median, which is the point that separates 50% of the results on one side and 50% on the other. In this case the median is 69.5782654 inches.

Uncertainty in the measure



- Histogram of means measured over each bootstrap sample
- 95% of the examples fell between 69.3184848 and 69.8428527, this is called the 95% confidence interval.

Summary

- The Simmel effect
 - Studying fashions as changes in status symbols
 - Model explains them as generated by imitation and differentiation
- Baby names statistics
 - An example of trend analysis of status symbols
 - Old registry data can also be big data
- Online trend dynamics and the endo-exo model
 - Trend patterns as a combination of exogenous shocks and word of mouth dynamics
- Bootstrapping
 - Assessing uncertainty in the measurement of a statistic
 - Idea: resampling with replacement many times