Your Two Weeks of Fame and your Grandmother's

James Cook [0]   Atish Das Sarma [1]

Alex Fabrikant [2]   Andrew Tomkins [2]

[0]UC Berkeley

[1]eBay Research Labs

[2]Google Research

WWW 2012

CNN is widely credited with initiating the acceleration of the modern news cycle with the fall 2006 debut of its spin-off channel CNN:24, which provides a breaking news story, an update on that story, and a news recap all within 24 seconds.   - The Onion

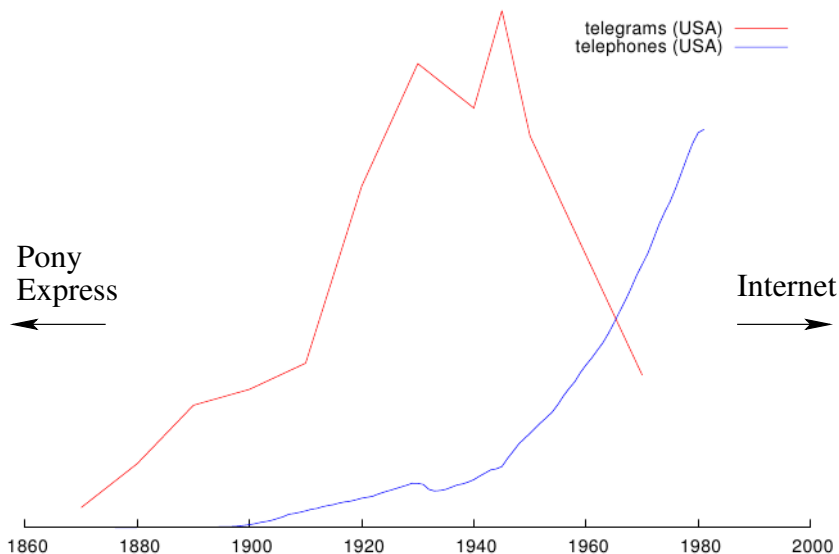## "In the future everyone will be world-famous for 15 minutes."

- Andy Warhol

- ▶ Can we measure changes in the public's attention span?
- ▶ Today, we can measure public behavior to the level of an individual using data sets like Twitter.
- ▶ What about before the Internet and personal digital records?
- ▶ Let's use news articles as a proxy for what the public is thinking about.
- ▶ Take-away: our intuitions are wrong. The typical person has always been famous for the same length of time, and the most famous are staying in the news for longer than ever before.

Outline

- Working with the news archive
- Measuring public attention
- Results

It's getting easier to communicate.

telegrams (USA)
telephones (USA)

Pony
Express
←

Internet
→

1860 1880 1900 1920 1940 1960 1980 2000

U.S. Census via http://eh.net/encyclopedia/article/nonnenmacher.industry.telegraphic.us

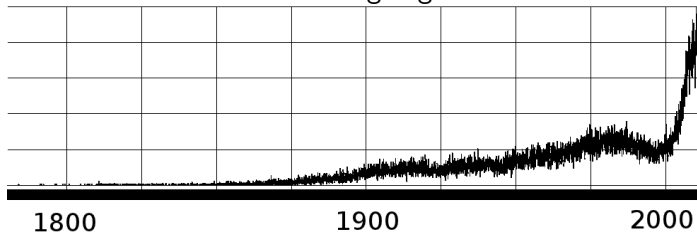FCC stats via http://www.galbithink.org/telcos/early-telephone-data.htm.

Google's News Archive

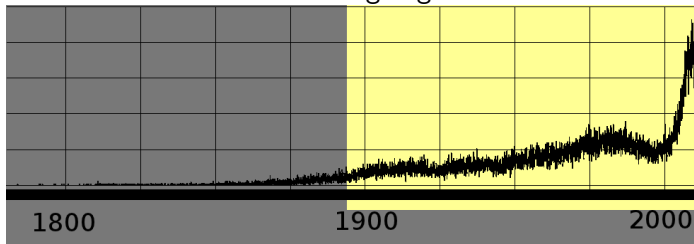- Can we measure changes in the public's attention span?

Google's News Archive

▶ Can we measure changes in the public's attention span?

▶ Over 60 million news articles going back to the 18th century.

Google's News Archive
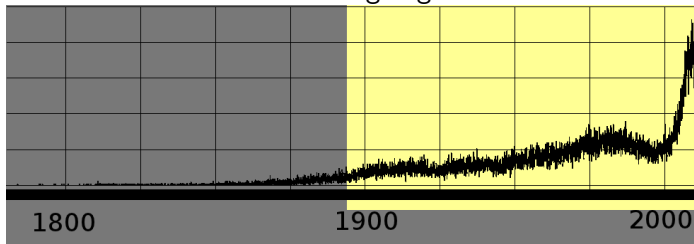
- Can we measure changes in the public's attention span?
- Over 60 million news articles going back to the 18th century.

- Substantial daily volume from 1895 to 2011. (Before that, low media volume and literacy rates start to fall off.)

Google's News Archive

- Can we measure changes in the public's attention span?
- Over 60 million news articles going back to the 18th century.

- Substantial daily volume from 1895 to 2011. (Before that, low media volume and literacy rates start to fall off.)
- Let's measure how long things stay in the news.

Measuring Public Attention

The categories of news have changed.

- 1909 Youngstown Vindicator:

  DAUGHTERS AT TEA TODAY

  In Preparation For Next Week's Contest Over Presidency.

  Washington, April 17.—One of the chief events preceding the congress of the Daughters of the American Revolution, which opens Monday, is the tea to be given this afternoon by Mrs. James S. Sherman, wife of the vice president.

- 2009 Telegraph:

  Couple's wedding entrance dance becomes YouTube viral hit

  Kevin Heinz and Jill Peterson, the wedding couple filmed dancing down the aisle, have become an internet sensation after the video received six million YouTube hits.

Still, News articles have always been about people.

Measuring Public Attention

The categories of news have changed.

- 1909 Youngstown Vindicator:

  DAUGHTERS AT TEA TODAY

  In Preparation For Next Week's Contest Over Presidency.

  Washington, April 17.—One of the chief events preceding the congress of the Daughters of the American Revolution, which opens Monday, is the tea to be given this afternoon by Mrs. James S. Sherman, wife of the vice president.

- 2009 Telegraph:

  Couple's wedding entrance dance becomes YouTube viral hit

  Kevin Heinz and Jill Peterson, the wedding couple filmed dancing down the aisle, have become an internet sensation after the video received six million YouTube hits.

Still, News articles have always been about people.

Measuring Public Attention

Measure how long personal names stay in the news.
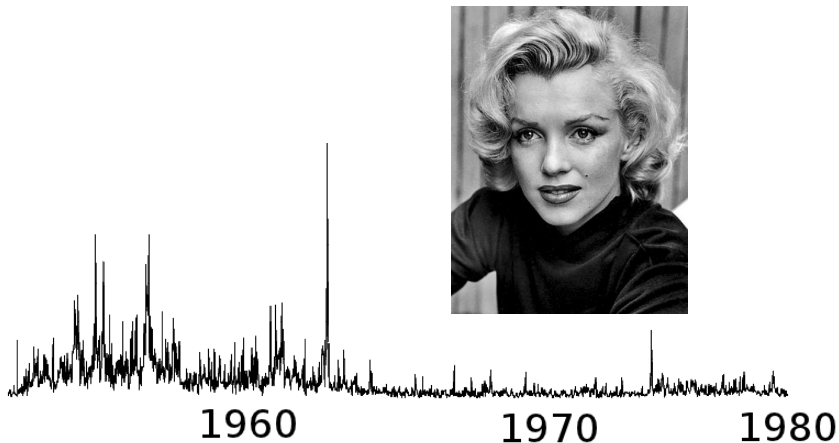
Timeline for Marilyn Monroe
photo: Life Magazine

Working with the News Archive

Is this a news article?

The Milwaukee Sentinal - Apr 9, 1921:

TOO WEAK TO WORK

Lydia E. Pinkham's Vegetable Compound Restored Mrs. Quinly's Health. Now She Does Her Housework

Shelbyville, Mo. — "I was only able to do light housework because for months my periods were excessive. I had seen your medicine extensively advertised and thought I would give it a fair trial. I took about eight boxes of Lydia E. Pinkham's Vegetable

Working with the News Archive

- A variety of things appeared as items in the corpus.
  - news articles
  - things like articles: photo captions, groups of articles accidentally identified as one
  - non-news: advertisements, sports scores, recipes
- Fortunately, the distribution hasn't changed much:

	full corpus sample	1900–1925 sample
news articles	31	28
news-like items	3	2
non-news items	16	20

- Solution:
  - Include all three classes of item in the study.
  - Count each individual occurrence of a name, so article boundaries don't matter.

Working with the News Archive

- How can we measure how long people stay in the news?
- Idea: take the first and last dates the name appears in the news.
- One of many bugs: lots of names are famous for exactly 20 years from 1960s to 1980s. (Why?)

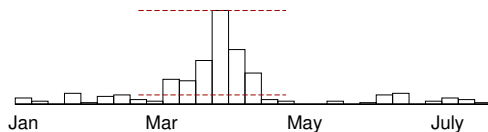# Working with the News Archive

- Some OCRd dates are off by several years.
- People can share the same name, and the same person can appear in the news more than once.

1850　　　　　　　1925　　　　　　2000
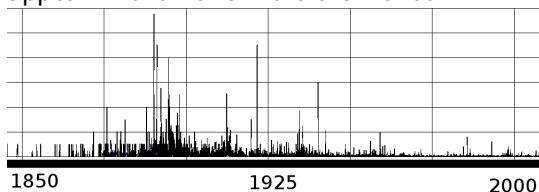
- - Solution: look at contiguous periods of attention, not global properties.
- Many more articles in 2010 than 1910.
  - Solution: sample the same number of articles in each month.

A Name's Period of Fame

- ▶ Plan: look at occurrences of names in Google's news archive to study fame durations now and in the past.
- ▶ We used two heuristics to identify periods of fame.
  1. *Spike method*
     - ▶ The spike around a news story: extends from week with most mentions to 10% threshhold.

  2. *Continuity method*
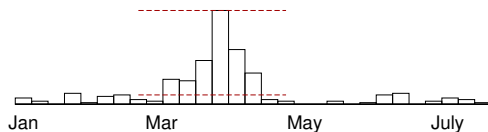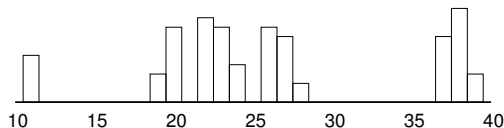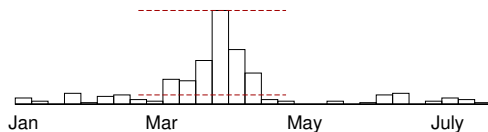   - ▶ Continuous public interest: longest stretch without a 7-day gap.

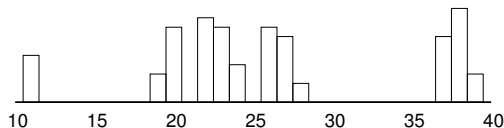

- ▶ We chose one period per name.

A Name's Period of Fame

- ▶ Plan: look at occurrences of names in Google's news archive to study fame durations now and in the past.
- ▶ We used two heuristics to identify periods of fame.
  1. *Spike method*
     - ▶ The spike around a news story: extends from week with most mentions to 10% threshhold.
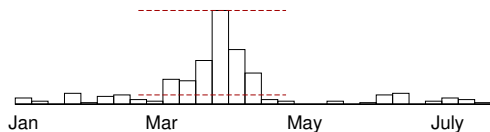

  2. *Continuity method*
     - ▶ Continuous public interest: longest stretch without a 7-day gap.
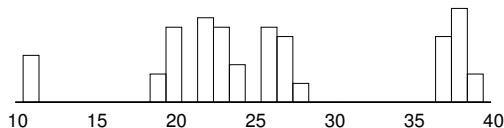

- ▶ We chose one period per name.

A Name's Period of Fame

- ► Plan: look at occurrences of names in Google's news archive to study fame durations now and in the past.
- ► We used two heuristics to identify periods of fame.
  1. *Spike method*
     - ► The spike around a news story: extends from week with most mentions to 10% threshhold.

  2. *Continuity method*
     - ► Continuous public interest: longest stretch without a 7-day gap.

- ► We chose one period per name.

A Name's Period of Fame

- ▶ Plan: look at occurrences of names in Google's news archive to study fame durations now and in the past.
- ▶ We used two heuristics to identify periods of fame.
  1. *Spike method*
     - ▶ The spike around a news story: extends from week with most mentions to 10% threshhold.
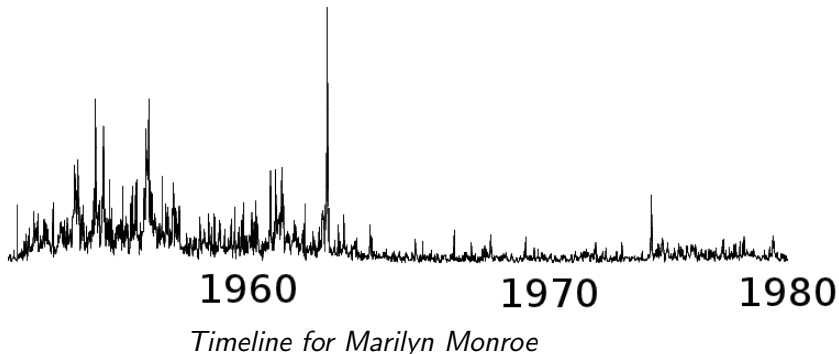

  2. *Continuity method*
     - ▶ Continuous public interest: longest stretch without a 7-day gap.
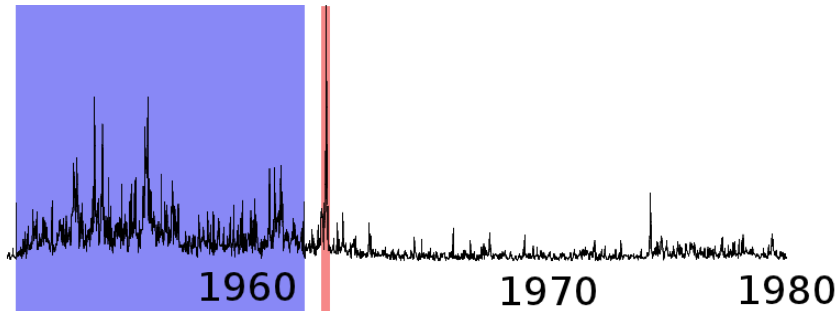

- ▶ We chose one period per name.

A Name's Period of Fame

- *Spike method:* One news story: extends from peak to 10% of peak.
- *Continuity method:* Continuous interest without a 7-day gap.

Timeline for Marilyn Monroe

A Name's Period of Fame

- *Spike method:* One news story: extends from peak to 10% of peak.
- *Continuity method:* Continuous interest without a 7-day gap.

Timeline for Marilyn Monroe
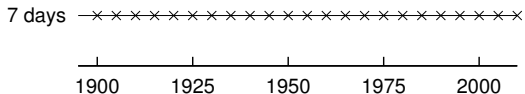
Results

- The median duration of fame is one week for the entire period of study (1895-2011).
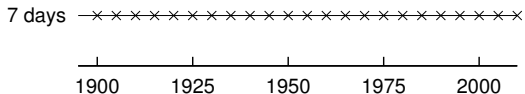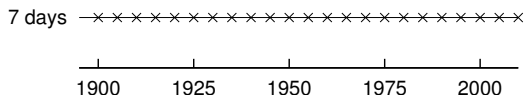
Results

- The median duration of fame is one week for the entire period of study (1895-2011).
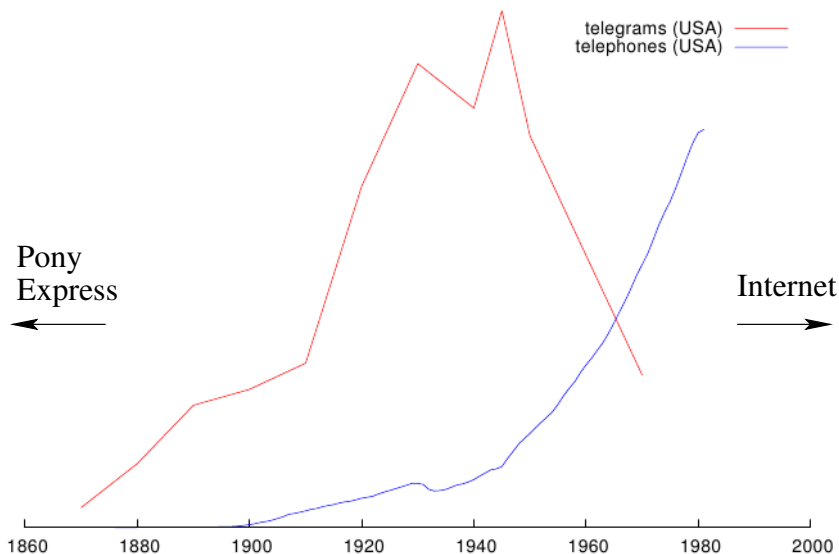

- ≥99% of bootstrap samples give exactly 7 days.

Results

- The median duration of fame is one week for the entire period of study (1895-2011).

- $\geq$99% of bootstrap samples give exactly 7 days.
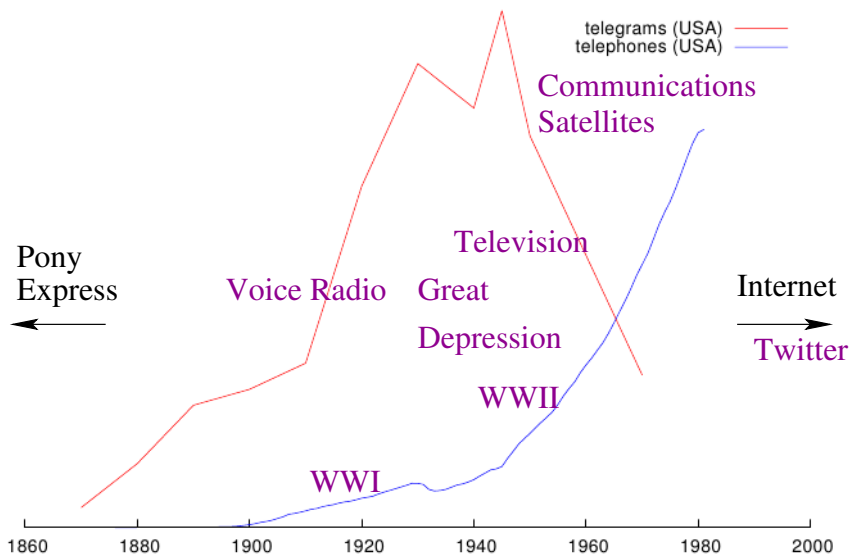- In a side study of public Blogger posts between 2000 and 2010, the median duration was also one week.

Results

telegrams (USA)
telephones (USA)

Pony
Express

←

Internet

→

1860   1880   1900   1920   1940   1960   1980   2000

U.S. Census via http://eh.net/encyclopedia/article/nonnenmacher.industry.telegraphic.us

FCC stats via http://www.galbithink.org/telcos/early-telephone-data.htm.

Results

telegrams (USA)
telephones (USA)

Pony
Express

Voice Radio

Television

Great

Depression

WWII

WWI

Communications
Satellites

Internet

Twitter

1860 1880 1900 1920 1940 1960 1980 2000

U.S. Census via http://eh.net/encyclopedia/article/nonnenmacher.industry.telegraphic.us

FCC stats via http://www.galbithink.org/telcos/early-telephone-data.htm.
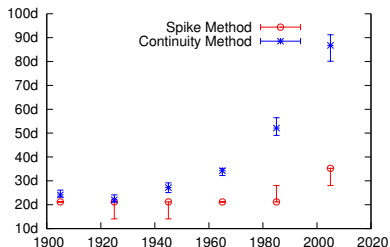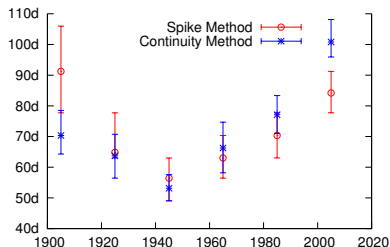
Results

What happens when we focus on the most famous names?

- ▶ If we look at the 99th percentile of duration instead of the median, then we see an increasing trend since the 1940s. (left)

- ▶ The same thing happens if we look at the 1000 most-mentioned names in each year. (right)

Future Work

- Beyond names, e.g. news stories
- Use geo data – newspapers have location tags!
- Were communications the driving force here? Try inferring the telegraph network from news propagation.
- Measure attention across dimensions other than time/fame: different countries, languages, levels of education.
- More nuanced statistical analysis.
- What are the causes? (Modelling? Control for diversity of sources?)
- What else can 100 years of news tell us? (Culturomics: using big data to measure cultural trends.)

Thanks!

Questions?