Do we Read what we Share?
Analyzing the Click Dynamic of News Articles Shared on Twitter

Motivation

• News and information spread over social media can have big impact on thoughts, beliefs, and opinions
  • Important to understand the sharing dynamics on these forums ...

• Most studies trying to capture these dynamics rely only on Twitter’s open APIs to measure how frequently articles are shared/retweeted
  • They do not capture how many users actually read the articles linked in these tweets ...

... here, we instead focus on the clicks leading to linked articles ... 
... and measure + analyze these over time.
Motivation

• News and information spread over social media can have big impact on thoughts, beliefs, and opinions
  • Important to understand the sharing dynamics on these forums ...

• Most studies trying to capture these dynamics rely only on Twitter’s open APIs to measure how frequently articles are shared/retweeted
  • They do not capture how many users actually read the articles linked in these tweets ...

... here, we instead focus on the clicks leading to linked articles ...
... and measure + analyze these over time.
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

• Example observations from temporal analysis
  • Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
  • Retweet data often underestimates biases towards clicking popular links/articles
  • Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
  • Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
  • Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

• We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

tweets
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

Example observations from temporal analysis

• Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
• Retweet data often underestimates biases towards clicking popular links/articles
• Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
• Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
• Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis tweets
Contributions at a glance

- Two main contributions
  - A novel longitudinal measurement framework
  - The first analysis of how the number of clicks changes over time

- Example observations from temporal analysis
  - Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
  - Retweet data often underestimates biases towards clicking popular links/articles
  - Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
  - Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
  - Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

- We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis.
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

  tweets

  fake news

  time

  clicks

  bit.ly
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

![Diagram showing tweets and clicks over time]
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

---

Example observations from temporal analysis

• Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle

• Retweet data often underestimates biases towards clicking popular links/articles

• Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks

• Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles

• Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

• We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

- Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
- Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
- Retweet data often underestimates biases towards clicking popular links/articles
- Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
- We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

- Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
- Retweet data often underestimates biases towards clicking popular links/articles
- Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
- Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
- Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis.
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

Example observations from temporal analysis
• Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
• Retweet data often underestimates biases towards clicking popular links/articles
• Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
• Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
• Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis.
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

- Example observations from temporal analysis
  • Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
  • Retweet data often underestimates biases towards clicking popular links/articles
  • Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
  • Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
  • Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

- We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis
Contributions at a glance

- Two main contributions
  - A novel longitudinal measurement framework
  - The first analysis of how the number of clicks changes over time

Example observations from temporal analysis
- Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
- Retweet data often underestimates biases towards clicking popular links/articles
- Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
- Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
- Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis of tweets.
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

- Example observations from temporal analysis
  • Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
  • Retweet data often underestimates biases towards clicking popular links/articles
  • Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
  • Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
  • Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

- We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis.
Contributions at a glance

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time

• Example observations from temporal analysis
  • Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
  • Retweet data often underestimates biases towards clicking popular links/articles
  • Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
  • Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
  • Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis.
Contributions at a glance

- Two main contributions
  - A novel longitudinal measurement framework
  - The first analysis of how the number of clicks changes over time

- Example observations from temporal analysis
  - Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
  - Retweet data often underestimates biases towards clicking popular links/articles
  - Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
  - Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
  - Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links

- We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis
Methodology
Methodology

- Collection of Bitly links to 7 pre-selected news website
  - 20-minute blocks (with latest tweets) collected over 7 days
Methodology

- Collection of Bitly links to 7 pre-selected news website
  - 20-minute blocks (with latest tweets) collected over 7 days
Methodology

- Collection of Bitly links to 7 pre-selected news website
  - 20-minute blocks (with latest tweets) collected over 7 days
Methodology

- Collection of Bitly links to 7 pre-selected news website
- 20-minute blocks (with latest tweets) collected over 7 days
Methodology

- Collection of Bitly links to 7 pre-selected news website
  - 20-minute blocks (with latest tweets) collected over 7 days
• Collection of Bitly links to 7 pre-selected news website
  • 20-minute blocks (with latest tweets) collected over 7 days

• Longitudinal click statistics
  • Each block scheduled for collection every 2 hours for 5 days
• Collection of Bitly links to 7 pre-selected news website
  • 20-minute blocks (with latest tweets) collected over 7 days
• Longitudinal click statistics
  • Each block scheduled for collection every 2 hours for 5 days
Methodology

• Collection of Bitly links to 7 pre-selected news website
  • 20-minute blocks (with latest tweets) collected over 7 days
• Longitudinal click statistics
  • Each block scheduled for collection every 2 hours for 5 days
• Careful sample frequency selection
  • To stay within rate limits, not all links sampled each time
  • Three sets per block: “top”, “random”, and “rest”
  • Sets sampled at different frequencies
  • Sets and their sizes designed to stay within rate limits (details/eqns. in paper)
Methodology

- Collection of Bitly links to 7 pre-selected news website
  - 20-minute blocks (with latest tweets) collected over 7 days
- Longitudinal click statistics
  - Each block scheduled for collection every 2 hours for 5 days
- Careful sample frequency selection
  - To stay within rate limits, not all links sampled each time
  - Three sets per block: “top”, “random”, and “rest”
  - Sets sampled at different frequencies
  - Sets and their sizes designed to stay within rate limits (details/eqns. in paper)
- Complementing tweet statistics
Clicks over time
Clicks over time

- 80% of all observed clicks occur within the first 24 hours
- Large variations across links
- Tail of less popular links sees a more even spread of clicks
- Suggests that studies focusing only on popular articles (or tweets) may underestimate the duration of the news cycle and the time that many news articles are read after they are first published
• 80% of all observed clicks occur within the first 24 hours
• 80% of all observed clicks occur within the first 24 hours
• Large variations across links
Clicks over time

- 80% of all observed clicks occur within the first 24 hours
- Large variations across links
- Tail of less popular links sees a more even spread of clicks
  - Suggests that studies focusing only on popular articles (or tweets) may underestimate the duration of the news cycle and the time that many news articles are read after they are first published
Correlations between retweets and clicks, but significant differences

• Generally larger click volumes
• Expected: Only subset of readers retweet what they read
• We may also miss earlier tweets (resulting in clicks)
• Significant set of tweets with more retweets than clicks
• Some people (or bots) retweet the links without actually clicking the link.
• This is clearly not good, as human sanity checking is an important tool to reduce the spreading of fake news
Comparison with tweet data

- Correlations between retweets and clicks, but significant differences
- Generally larger click volumes
- Expected: Only subset of readers retweet what they read
- We may also miss earlier tweets (resulting in clicks)
- Significant set of tweets with more retweets than clicks
- Some people (or bots) retweet the links without actually clicking the link.
- This is clearly not good, as human sanity checking is an important tool to reduce the spreading of fake news
Comparison with tweet data

- Correlations between retweets and clicks, but significant differences
- Generally larger click volumes
  - Expected: Only subset of readers retweet what they read ...
  - We may also miss earlier tweets (resulting in clicks)
Comparison with tweet data

- Correlations between retweets and clicks, but significant differences
- Generally larger click volumes
  - Expected: Only subset of readers retweet what they read ...
  - We may also miss earlier tweets (resulting in clicks)
Comparison with tweet data

- Correlations between retweets and clicks, but significant differences
- Generally larger click volumes
  - Expected: Only subset of readers retweet what they read ...
  - We may also miss earlier tweets (resulting in clicks)
- Significant set of tweets with more retweets than clicks
  - Some people (or bots) retweet the links without actually clicking the link.
  - This is clearly not good, as human sanity checking is an important tool to reduce the spreading of fake news
Comparison with tweet data

• Clicks typically progress somewhat faster (at start) than retweets
Comparison with tweet data

- Clicks typically progress somewhat faster (at start) than retweets

- Also some subtle difference (see paper) that indicates retweet data underestimate bias towards popular links/articles

- These links are often highly shared in the beginning, but also accumulate reads/clicks (at a much slower rate) later …
Comparison with tweet data

• Clicks typically progress somewhat faster (at start) than retweets
• Also some subtle difference (see paper) that indicates that retweet data underestimate bias towards popular links/articles
Comparison with tweet data

• Clicks typically progress somewhat faster (at start) than retweets
• Also some subtle difference (see paper) that indicates that retweet data underestimate bias towards popular links/articles
  • These links are often highly shared in the beginning, but also accumulate reads/clicks (at a much slower rate) later ...
Impact of age

- “Older” articles accumulating clicks at a much more uniform rate over the measurement duration
Impact of age

- “Older” articles accumulating clicks at a much more uniform rate over the measurement duration.
Impact of age

• “Older” articles accumulating clicks at a much more uniform rate over the measurement duration

• For “older” articles, most of the clicks are associated with articles that do not appear to fade (as quickly) in popularity
  • The opposite is true for the “overall” set and for “younger” articles
  • Again, the “older” set contains more long-term popular articles
Impact of age

- “Older” articles accumulating clicks at a much more uniform rate over the measurement duration
- For “older” articles, most of the clicks are associated with articles that do not appear to fade (as quickly) in popularity
  - The opposite is true for the “overall” set and for “younger” articles
  - Again, the “older” set contains more long-term popular articles
Validation year-old data (from 2017)

- Identify invariants
  - The early peaks, the skew towards a subset of highly popular links, and the differences between links to articles of different age appear invariant
Validation year-old data (from 2017)

- Identify invariants
  - The early peaks, the skew towards a subset of highly popular links, and the differences between links to articles of different age appear invariant
Validation year-old data (from 2017)

- Identify invariants
  - The early peaks, the skew towards a subset of highly popular links, and the differences between links to articles of different age appear invariant
Age and long-term churn
Age-dependent popularity skew

- While some differences within the two age-based sub-classes, the most substantial differences are between the categories themselves.
Age-dependent popularity skew

While some differences within the two age-based sub-classes, the most substantial differences are between the categories themselves.

“young” (≤ 1 week)
• While some differences within the two age-based sub-classes, the most substantial differences are between the categories themselves
Age-dependent popularity skew

- While some differences within the two age-based sub-classes, the most substantial differences are between the categories themselves.
- For the two “oldest” classes, the CCDFs shows relatively straight-line behavior, suggesting a power-law-like popularity skew.
Age-dependent popularity skew

- While some differences within the two age-based sub-classes, the most substantial differences are between the categories themselves.
- For the two “oldest” classes, the CCDFs shows relatively straight-line behavior, suggesting a power-law-like popularity skew.
- For the “younger” articles, there is relatively higher popularity churn.
Age-dependent churn

- Increasing churn among both the “younger” and “older” articles
- In contrast, for YouTube videos, long-term popularity has been found to reduce the churn over time [Borghol et al., 2011]
Age-dependent churn

- Increasing churn among both the “younger” and “older” articles
  - In contrast, for YouTube videos, long-term popularity has been found to reduce the churn over time [Borghol et al., 2011]
Age-dependent churn

- Increasing churn among both the “younger” and “older” articles
  - In contrast, for YouTube videos, long-term popularity has been found to reduce the churn over time [Borghol et al., 2011]

- A short initial interval have been found to be a good predictor of the clicks over the remainder of the time period
Even the clicks observed over a very short interval (e.g., 2 hrs) provides better insight into the actual information reach over a longer time period (e.g., 120 hrs) than the retweet do (even if using the same 120 hrs)
Age-dependent churn

Even the clicks observed over a very short interval (e.g., 2 hrs) provides better insight into the actual information reach over a longer time period (e.g., 120 hrs) than the retweet do (even if using the same 120 hrs)
Age-dependent churn

Even the clicks observed over a very short interval (e.g., 2 hrs) provides better insight into the actual information reach over a longer time period (e.g., 120 hrs) than the retweet do (even if using the same 120 hrs)
Life-time clicks

- “Younger” links gain on the lifetime clicks observed for the “older” links, closing the gap
• “Younger” links gain on the lifetime clicks observed for the “older” links, closing the gap
“Younger” links gain on the lifetime clicks observed for the “older” links, closing the gap.
Life-time clicks

- “Younger” links gain on the lifetime clicks observed for the “older” links, closing the gap
Life-time clicks

- “Younger” links gain on the lifetime clicks observed for the “older” links, closing the gap.
- However, the gap is still substantial at the end of the 120 hour period.
  - This highlighting that the “older” category includes many links to long-term popular articles.
Per-site-based analysis
Invariants despite significant differences

- Age-based invariants (across websites) despite significant differences observed between the different websites (e.g., age, speed clicks are obtains, and click distribution)
- Our age-based conclusions are consistent for each of the news websites individually, further validating our previous claims
Invariants despite significant differences

- Age-based invariants (across websites) despite significant differences observed between the different websites (e.g., age, speed clicks are obtained, and click distribution)

- Our age-based conclusions are consistent for each of the news websites individually, further validating our previous claims
Invariants despite significant differences

• Age-based invariants (across websites) despite significant differences observed between the different websites (e.g., age, speed clicks are obtains, and click distribution)

• Our age-based conclusions are consistent for each of the news websites individually, further validating our previous claims
Conclusions
Conclusions and summary

• Two main contributions
  • A novel longitudinal measurement framework
  • The first analysis of how the number of clicks changes over time
• Example observations from temporal analysis
  • Noticeable differences in the relative number of clicks vs. retweets occurring at different parts of the news cycle
  • Retweet data often underestimates biases towards clicking popular links/articles
  • Significant differences in the clicks-per-tweets ratio, including (alarmingly) many links with more retweets than clicks
  • Significant age biases, including relatively high initial click rates for articles younger than a week and much more stable click rates for older and long-term popular articles
  • Insights into how age-dependent popularity skews and age-dependent churn impact the clicks observed by different classes of links
• We validate our findings (and identify invariants) using both data from May 2017 and a per-website-based analysis
Do we Read what we Share? Analyzing the Click Dynamic of News Articles Shared on Twitter

Jesper Holmstrom, Daniel Jonsson, Filip Polbratt, Olav Nilsson, Linnea Lundstrom, Sebastian Ragnarsson, Anton Forsberg, Karl Andersson, and Niklas Carlsson

Niklas Carlsson (niklas.carlsson@liu.se)