This paper will appear in *Proc. IEEE/ACM ASONAM*, Vancouver, Canada, Aug. 2019. The definitive version is published in IEEE Xplore Digital Library and ACM Digital Library via http://dx.doi.org/10.1145/3341161.3342933.

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

Jesper Holmström, Daniel Jonsson, Filip Polbratt, Olav Nilsson
Linnea Lundström, Sebastian Ragnarsson, Anton Forsberg, Karl Andersson, Niklas Carlsson
Linköping University, Sweden
{jesho280, danjo390, filpo653, olani913, linlu607, sebra023, antfo325, karan369}@student.liu.se, niklas.carlsson@liu.se

***Abstract*—News and information spread over social media can have big impact on thoughts, beliefs, and opinions. It is therefore important to understand the sharing dynamics on these forums. However, most studies trying to capture these dynamics rely only on Twitter's open APIs to measure how frequently articles are shared/retweeted, and therefore do not capture how many users actually read the articles linked in these tweets. To address this problem, in this paper, we first develop a novel measurement methodology, which combines the Twitter steaming API, the Bitly API, and careful sample rate selection to simultaneously collect and analyze the timeline of both the number of retweets and clicks generated by news article links. Second, we present a temporal analysis of the news cycle based on five-day-long traces (containing both clicks and retweet over time) for the news article links discovered during a seven-day period. Among other things, our analysis highlights differences in the relative timelines observed for clicks and retweets (e.g., retweet data often lags and underestimates the bias towards reading popular links/articles), and helps answer important questions regarding differences in how age-based biases and churn affect how frequently news articles shared on Twitter are accessed over time.**

***Index Terms*—Social media, News and information sharing, Temporal click dynamics, Twitter, Bitly**

I. INTRODUCTION

While information sharing over social media started out innocently, organizations and individuals are today using social media as a tool to sway thoughts and opinions for their own benefit, including to undermine other organizations, individuals, or even to weaken entire countries. To provide any chance to counter these negative trends that threaten our society, as the spreading of (mis)information is becoming an increasing concern, it is therefore important to further our understanding of the underlying sharing dynamics on these social media.

A popular metric when studying the information reach on social media, such as Twitter, is the number of retweets. While Twitter's open APIs provide easy ways to measure the number of retweets, this metric does not capture how many users actually read (or click on) articles linked in these tweets. With many bots and other orchestrated efforts contributing retweets, using only such first-order metrics can therefore result in the wrong conclusions and estimates of the articles' relative reach.

To address this problem and further the understanding of the extent to which news articles shared over Twitter are read over time, we first present a novel measurement methodology that allows us to simultaneously track news article clicks and retweets over time. Our methodology combines the use of the Twitter APIs with use of the Bitly API (a popular link shortener frequently used on Twitter and other social media services). Second, we present a temporal characterization based on this methodology, in which we analyze the clicks and retweets of the news articles promoted on Twitter between April 12-19, 2018, and address a number of previously not addressed research questions. While we are not the first to combine the use of the Twitter API and the Bitly API [1], we are the first to present a longitudinal measurement framework and analysis of how the number of clicks changes over time.

Our methodology continuously identifies all Bitly links associated with seven major news websites and then tracks both the clicks and retweets of all newly discovered links for a five-day period. By breaking the timeline into discrete 20-minute blocks and carefully tune the sampling frequency with which we measure how many clicks have been associated with each link thus far, we are able to track each such link for five days, while ensuring that we keep the number of calls to Bitly's API within its rate limits.

Our temporal analysis is based on the five-day traces for each of the news article links discovered during a seven-day period, uses the dates when the articles were published to timestamp the age of news articles, and highlights interesting click dynamics. First, we observe noticeable differences in the relative fraction of the clicks vs retweets that occur during different parts of the news cycle and that the retweet data often underestimates the bias towards clicking popular links/articles. Second, while some care should be taken when interpreting the clicks-per-retweet ratio of individual links, we note significant differences in this ratio, including (most alarmingly) a significant number of links for which there are more retweets than clicks, suggesting that some people (or bots) retweeted these links without actually clicking the link.

Third, we observe 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

ASONAM '19, August 27-30, 2019, Vancouver, Canada

ACM ISBN 978-1-4503-6868-1/19/08...$15.00
http://dx.doi.org/10.1145/3341161.3342933

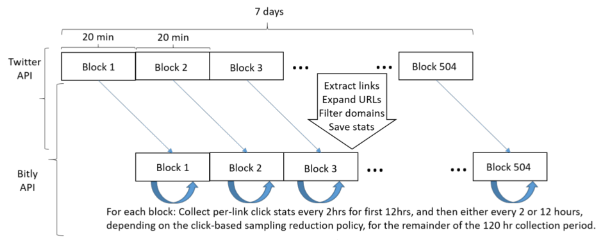


Fig. 1. Methodology overview: 7+5 day measurement campaign.

articles. Fourth, we provide insights into how age-dependent popularity skews and age-dependent churn impact the clicks classes of links obtain over time and how the number of clicks of individual links within a short initial time window is often a good indicator of the (future) lifetime clicks for that link (e.g., compared to the lifetime retweets).

Finally, we validate our findings using both data from May 2017 and a per-website-based analysis. While there are substantial differences between the websites (e.g., with regards to their articles' age distribution, clicks-per-link distribution, and the relative skew with which clicks happen early in the news cycle; all aspects characterized here), the relative age-biases and the general popularity skews (e.g., young popular links often see a significant drop in clicks, and older long-term popular links often see relatively smaller changes in the click rates) are consistent for all considered websites.

Outline: Section II outlines our methodology. Section III compares the timelines of clicks and retweets, and analyzes the impact of article age. Section IV analyzes age-dependent popularity skews and churn, and relates our temporal results with corresponding lifetime metrics. Section V presents a per-site analysis that validates that our observations and conclusions hold on a per-site basis. Finally, we discuss related work (Section VI) and summarize our conclusions (Section VII).

II. Methodology

We first present a novel collection framework that carefully tracks the number of clicks that Bitly links embedded in a series of (re)tweets generate over time.[1]

Figure 1 illustrates the steps taken during an example measurement campaign. Here, we followed all Bitly links that are tweeted during a seven-day period and that point to news articles associated with a set of pre-selected news websites. To allow efficient tracking, we split the collection into 20-minute blocks and track the clicks associated with the links of interest in each such block over five days (120 hours)

Collection of Bitly links to selected news websites: First, during each 20-minute block, we use Twitter's streaming API to extract all tweets with Bitly links posted during that block, and save the corresponding tweets to a block-specific JSON file. Second, using the Bitly API, we extract the full URLs associated with the Bitly links not yet observed. Third, using the full URLs, we identify links to seven pre-selected news domains: BBC, The Times, The Guardian, Huffington Post, CNN, Fox News, and Breitbart. These sites cover a wide spectrum of political views/biases and have been suggested to publish fake or politically biased news at vastly different rates. In this step, we also collect our first "0-hour" measurement for the number of clicks associated with the links.

Longitudinal click statistics collection: For each 20-min block, we use the Bitly API to periodically collect the number of clicks associated with each link for the next 120 hours.

Sample frequency selection: An important aspect here is the selection of per-link sample frequencies that take into account the constraints of the Bitly API (e.g., 100/minute, 1000/hour, and using at most 5 parallel connections).

After some initial investigation, we decided to use the following sampling policy. First, for the first 12 hours, all links in a block are measured every 2 hours (plus/minus a few minutes). This is the time period when most of the clicks occur. Second, at the 12 hour mark, we select two subsets of links that we continue to track every 2 hours, while we sample the other links again at the 24 hour mark, and then every 24 hours after that. The first subset (called "top") contains the links that have obtained more than a threshold C_{12} clicks after 12 hours and the second subset (called "random") is just a random subset. The higher sampling frequency of "top" is motivated by these links being more likely to generate more additional clicks (and therefore are more likely to capture interesting events), whereas "random" provides a baseline that we use to evaluate the extent information is lost when down sampling.

For our experiments, links were added to "random" with a probability 0.25 and we picked a threshold C_{12} so that "top" would be of a similar size. Assuming independence between the sets, the combined set (sampled every 2 hours) was therefore forecasted to include $\approx$43.75% of the links. This fraction was selected to ensure that we stayed below an upper bound on the fraction $f<0.57$ that we can continue to sample at a high frequency while satisfying the Bitly rate constraint (1000/hour). This bound is obtained by solving the following inequality: $24(60f+11(1-f))<1,000$, where we have used that on average 24 new unique Bitly links are added per hour (to the full set of followed links), and that the links in the union of "top" and "random" are sampled 60 times and the others 11 times. In the end, for the dataset analyzed here, "random" contained 24.8% of the links, "top" 25.5% of the links, and 7.1% of the links were in both sets (resulting in f=0.43).

Complementing tweet statistics: To compare the evolution of clicks and retweets for a particular article, we extracted all tweets associated with the links of interest, the follower statistics of each tweeter, and rebuilt the tree structures that the tweets form. For the purpose of this analysis, we focus only on the timing of the tweets and refer to all tweets that occur after the first observed tweet as "retweets", even though some may be of other types (e.g., quotes). To match the click data, we simply extracted and compared the cumulative numbers for every 2 hours in the dataset.

Limitations: Our methodology is limited to short-links using the Bitly API and only covers a subset of all articles shared on Twitter. We do not try to distinguish between links/clicks generated by bots vs humans, but note that bots

[1]Code+datasets available: www.ida.liu.se/~nikca89/papers/asonam19.html

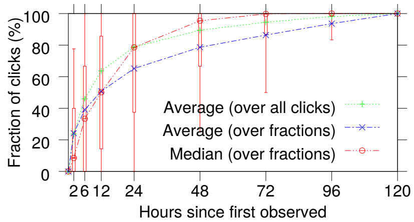


Fig. 2. Distribution statistics and averages for the cumulative number of clicks as a function of the time since the link was first observed.

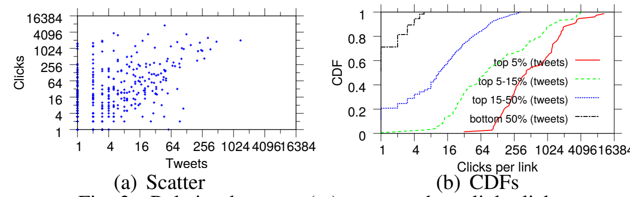
trying to influence opinion are more likely to (re)tweet than follow links, and crawlers are likely to spread more evenly across links. Similarly, just because a link is clicked does not mean that a user reads the article in full. Finally, we note that the Twitter API limits the total tweet volume to 1%, and the Bitly API limits the request rate and number of concurrent connections (to five). The Twitter constraints did not impact our results since our stream (containing only tweets with Bitly URLs) did not reach the 1% threshold, monitored during the collection. The Bitly API constraints, on the other hand, forced us to use a 2-hour spacing between calls for each link of interest, combined with down sampling of some links after 12 hours, and the use of 20-minute blocks to spread the potential missing data points in the case that not all links could be resolved in a particular 20-minute block. In the dataset analyzed here, we do not have any such missing data points.

III. Clicks over time

Let us first consider the rate that new clicks are observed as a function of the time since each link were first observed in the tweet stream. Figure 2 shows distribution statistics calculated across all individual links, as well as average statistics calculated both as an aggregate across all clicks (ignoring differences between links) and as an average of the fraction of clicks that each individual link has accumulated. For the distribution statistics, we show the 10-percentiles (bottom markers), 25-percentiles (bottom of boxes), medians (middle red markers), 75-percentiles (top of boxes), and 90-percentiles (top markers). All values are normalized between 0 and 1, with 0 and 1 corresponding to the clicks at the start and end of each measurement campaign, respectively.

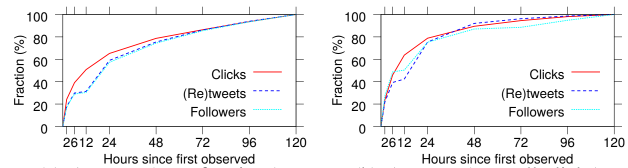
Overall, approximately 80% of all observed clicks occur within the first 24 hours of the five-day period, suggesting that most clicks occur early in the life of a link. However, there are large variations across links. The significantly slower increase when considering the average over all fractions (e.g., approximately 60% within the first 24 hours) suggests that the tail of less popular links (with fewer clicks) sees a more even spread of clicks. This is interesting as it suggests that studies that focus 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.

With a significant tail of links with relatively few clicks, there is a substantial number of links that transition from 0 to 1 relatively quickly. This is also reflected by the (compared to the average values) relatively sharp increase in the median fraction. As this point can be relatively sensitive to the specific



(a) Scatter (b) CDFs

Fig. 3. Relation between (re)tweets and per-link clicks.



(a) Average (over fractions) (b) Average (over all clicks)

Fig. 4. Comparison of timeline data for clicks, (re)tweets, and aggregated followers associated with (re)tweeters.

selection of links, the average values are more stable and provide a relatively more fair first order comparison when comparing subsets of links. In the following, we primarily focus on the two average metrics.

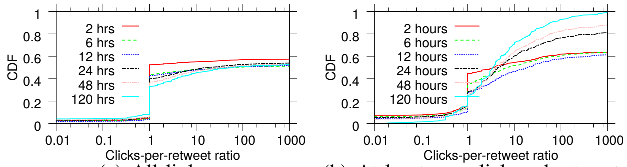
A. *Comparison with tweet data*

Retweets have often been used to estimate popularity and information reach of news articles. However, just because a tweet is retweeted a lot (little) does not mean that it is read a lot (little). This is, for example, captured in Figure 3(a), where we plot the number of retweets against the number of clicks observed to each article (over 120 hours). To help readers, Figure 3(b) shows the corresponding cumulative distribution functions (CDFs) of the number of clicks associated with links that fall into different percentile ranges as determined by the tweets. While there are significant correlations between the metrics, there are also many outliers.

Part of the above differences are due to only a subset of all people reading an article also retweeting the article; resulting in relatively fewer retweets than clicks. The larger click volumes can also be explained by us potentially missing tweets (of the same link) that occurred before our measurements started or the link being posted on other forums. However, interestingly, we also observe links for which the opposite is true (i.e., there are more retweets than clicks), suggesting that some people (or bots) retweet the links without actually clicking the link. While this may be due to people having accessed the article via some other means than clicking the link, it appears that many of these cases are due to people (or bots) actually not reading some articles they retweet. This is clearly not good, as these instances indicate lack of human sanity checking before these news articles are shared, and raises concerns as careful human sanity checking is an important tool to reduce the spreading of fake news.

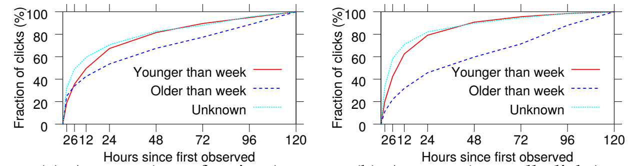
We have also found that the clicks typically progress somewhat faster than the retweets. For example, Figure 4 shows that the retweets have a slower initial rise than the clicks. There are also some other subtle differences. For example, note that the relative fraction of retweets over the last three days are less than the relative fraction of clicks when accounting for all

This paper will appear in *Proc. IEEE/ACM ASONAM*, Vancouver, Canada, Aug. 2019. The definitive version is published in IEEE Xplore Digital Library and ACM Digital Library via http://dx.doi.org/10.1145/3341161.3342933.



(a) All links (b) At least one click and retweet

Fig. 5. CDFs of the ratio between clicks and retweets, observed over the first T hours, where T=2, 6, 12, 24, 48, 120.



(a) Average (over fractions) (b) Average (over all clicks)

Fig. 6. Clicks over time for articles of different age categories.

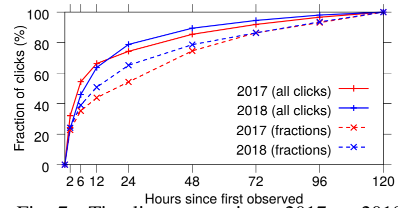


Fig. 7. Timeline comparison: 2017 vs 2018.

clicks (Figure 4(b)), but that the rates are similar when looking at a per-link basis (Figure 4(a)). This captures that retweet data underestimate the bias towards reading popular links/articles. These links are often highly shared in the beginning, but also accumulate reads/clicks (at a much slower rate) later.

To investigate the individual differences further, Figure 5 shows CDFs of the ratio between clicks and retweets, as observed over different time periods T, across all analyzed links (Figure 5) and across the 380 (out of 1,448 links) with at least one click and at least one retweet during the 120 hour period. We note that an infinite ratio corresponds to links without any retweets (but some clicks) and set the ratio equal to 1 whenever we have not observed any clicks or retweets over T. Although this fraction reduces over time, it is still non-negligible after 120 hours. Finally, while some of the initial observations of links with more retweets than clicks (i.e., with ratios below 1) may be attributed to clicks being missed due to the clicks taking place before we first observe the link, we note that the fraction of standouts in this category increases over time. For example, across the 1,448 links analyzed here the number of links with more retweets within the T hour time period progresses as 65 (first 2 hours), 62 (6 hours), 49 (12 hours), 80 (24 hours), 84 (48 hours), and 119 (120 hours).

B. *Impact of article age*

For this analysis, when possible, we extracted the article's publication date from the long URL. Figure 6 compares the relative timeline of clicks to articles "younger than 1 week" (633 articles) and "older than 1 week" (440 articles) when first observed by us, as well as to the remaining articles with "unknown" age (433 articles). Large differences are observed, with the "older" articles accumulating clicks at a much more uniform rate over the measurement duration. Furthermore, for this subset, most of the clicks are associated with articles that do not appear to fade; e.g., seen by the much flatter "older than 1 week" curve in Figure 6(b) compared to Figure 6(a). This is interesting as the opposite is true for the overall set of links; e.g., as seen by the sharper knee for the "over all clicks" in Figure 2, or more specifically, for the "younger than 1 week" category, by the sharper knee for this curve in Figure 6(b) compared to Figure 6(a). These observations are consistent with our previous results (above), as the "older" category contains more long-term popular articles that have clicks more evenly spread over time.

C. *Validation on year-old data*

The early peaks, the skew towards a subset of highly popular links, and the differences between links to articles of different age appear invariant when comparing with data collected roughly one year earlier (first week of May 2017). This is exemplified in Figures 7 and 8. Figure 7 compares the average fraction of clicks accumulated as a function of time using both average metrics and for both datasets. Figure 8 shows the same age-based breakdown for the 2017 dataset as was shown for the 2018 dataset in Figure 6. We note that the both the relative gap between the two average curves (Figure 7) and the age-based differences (Figure 8 vs Figure 6) have similar characteristics, suggesting that our observations are more generally applicable and have remained valid over time.

IV. AGE AND LONG-TERM CHURN

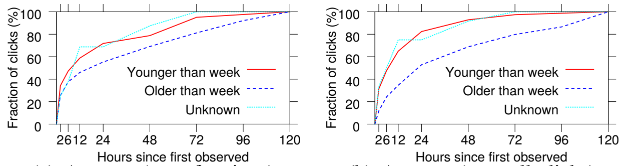
A. *Age-dependent popularity skew*

We next compare and contrast the total views accumulated by links associated with different aged contents. Figures 9(a) and 9(b) show the CDFs and Complementary CDFs (CCDFs) of the clicks observed, respectively, for links to articles of different age. To provide some additional depth, we break down the two previous age categories "younger" and "older" (than a week) further: 315 "less than 1 day", 318 "between 1 day and 1 week", 49 "between 1 week and 1 month", 187 "between 1 month and 1 year", 204 "older than year", and 433 "unknown". However, while we observe some differences within these two sub-classes (especially the "older" category), the most substantial differences are between the categories themselves. For example, even with the x-axis on log scale, the "younger" (than a week) articles see a general shift in the curves to the right compared to the "older" articles. Furthermore, for the two oldest categories, the CCDFs shows relatively straight-line behavior, suggesting a power-law-like skew. For the "younger" articles, there is relatively higher churn in the popularity.

B. *Age-dependent churn*

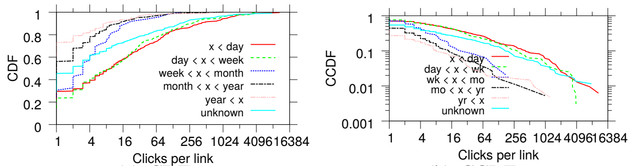
In contrast to for YouTube videos (where long-term popularity have been found to reduce the churn over time [2]), we have found increasing churn among both the "younger" and "older" articles. For example, we observe much higher correlation between the clicks within hours 0-to-2 and hours 2-to-6 (Figure 10(a)) than between during hours 0-to-6 and hours

This paper will appear in *Proc. IEEE/ACM ASONAM*, Vancouver, Canada, Aug. 2019. The definitive version is published in IEEE Xplore Digital Library and ACM Digital Library via http://dx.doi.org/10.1145/3341161.3342933.



(a) Average (over fractions) (b) Average (over all clicks)

Fig. 8. Age-based breakdown for 2017 dataset.



(a) CDF (b) CCDF

Fig. 9. Distributions of the number of clicks during the 120 hour period, conditioned on the age of the article when first observed.

6-to-24 (Figure 10(b)), and between during hours 0-to-24 and hours 24-to-120 (Figure 10(c)). In fact, a short initial interval have been found to be a good predictor of the clicks over the remainder of the time period. This is shown in Figures 10(d). During this time period many links obtain a significant fraction of their clicks, and the time period may be short enough that the best selection remains good also for the next time interval.

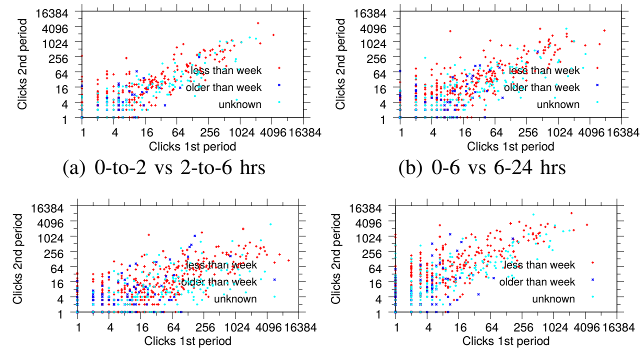
As expected, the clicks during the first two hours is an even better predictor of the total number of clicks a link over the full 0-120 hour window. Figure 11, shows the corresponding correlation results. Comparing with Figure 3, we note that the number of clicks after just 2 hours is a much better predictor of the clicks over the first 120 hours than the retweets over the full 0-to-120 period. This suggests that even the clicks observed over a very short interval provides more insights into the actual information reach (over a longer time period) than a retweet based analysis can do, highlighting the value of the methodology presented in this paper.

C. Lifetime clicks

Thus far we have focused on the clicks that occur during the 120 hour measurement window of each link. However, some clicks occurs before we first observe the link. Since links to "younger" articles on average obtaining more clicks during the 120 hour interval, we expect these "younger" links to gain on the lifetime clicks observed for the "older" links. To capture how the gap between the overall lifetime clicks reduces over the 120 hour window that we track each link, Figure 12 plots the CDFs of the lifetime clicks as observed when we first discover the links (Figure 12(a)) and 120 hours later (Figure 12(b)). As expected, despite being on log-scale, the gap reduce substantially. However, we also note that the gap still is substantial at the end of the 120 hour period, highlighting that the links in the "older" category in fact include many links to long-term popular articles.

V. Per-site-based analysis

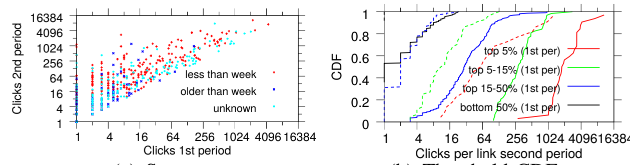
We have observed significant differences in the relative click-per-link distributions (capturing the reach) and the age distributions for the articles of the different news websites considered here. Figure 13(a) shows the CCDF of the number



(a) 0-to-2 vs 2-to-6 hrs (b) 0-6 vs 6-24 hrs

(c) 0-24 vs 24-120 hrs (d) 0-2 vs 2-120 hrs

Fig. 10. Scatter plots of clicks during non-overlapping intervals.



(a) Scatter (b) Threshold CDFs

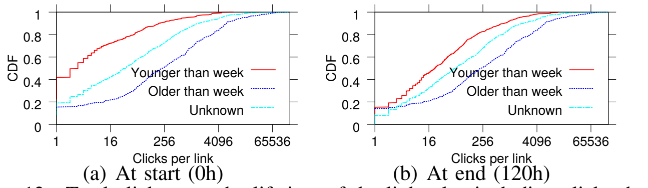
Fig. 11. Correlation between early counts (0-to-2) and full-period counts (0-to-120). Note that the second interval here contains the first. (Solid CDFs shows "younger" and dotted CDFs shows "older" articles.)

of clicks to the six websites with the most links in the dataset: Guardian (432), Breitbart (376), Huffington Post (227), BBC (174), CNN (108), and Fox (101). Of these websites, our simple URL-based age-classifier could classify the majority of websites for four of the services: Guardian (431), Breitbart (343), CNN (104), and Fox (97). Figure 13(b) shows the age distribution (on a per-day granularity) of the articles of these four websites. In general, Fox News (23 vs 44) and CNN (7 vs 63) have the largest fraction of "older" articles and The Guardian (276 vs 61) and Breitbart News (211 vs 66) the largest fraction of "younger" articles in our dataset.

Figure 14 highlights that there a substantial differences in how quickly the websites gain the majority of their clicks. We note that BBC and Huffington Post articles attract most of their clicks (across article catalogue) faster than the other websites (Figure 14(a)), suggesting that a larger fraction of their articles contains news that people find relatively more interesting when current, but for which interest fades over time. With Fox News replacing Huffington Post in the top two (together with BBC) when considering the time to peak when calculating the averages over all clicks (Figure 14(b)), it appears that Fox News have some highly read articles that are short-lived. This corresponds to the right-most tail in Figure 13(a).

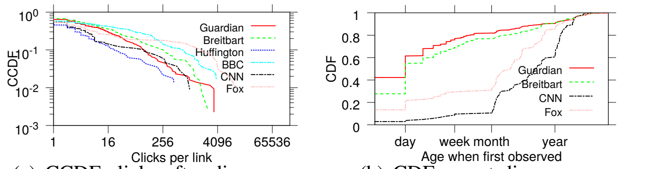
Finally, and perhaps most importantly, despite the above differences observed between the different websites, we have found that our age-based conclusions are consistent for each of the news websites individually, further validating our previous claims. Figure 15 presents the corresponding per-website-based comparisons of the clicks to links pointing to articles "older" or "younger" than a week. Again, we note a substantial gap between the "younger" and "older" articles, with a significantly flatter distributions for the links to the "older" articles. This confirms that our prior observations also holds on a per-website bases, across the different sites.

This paper will appear in *Proc. IEEE/ACM ASONAM*, Vancouver, Canada, Aug. 2019. The definitive version is published in IEEE Xplore Digital Library and ACM Digital Library via http://dx.doi.org/10.1145/3341161.3342933.



(a) At start (0h) (b) At end (120h)

Fig. 12. Total clicks over the lifetime of the link, also including clicks that occurred before we first observed the link.



(a) CCDF clicks after discovery (b) CDF age at discovery

Fig. 13. Based on website (new only).

VI. Related Work

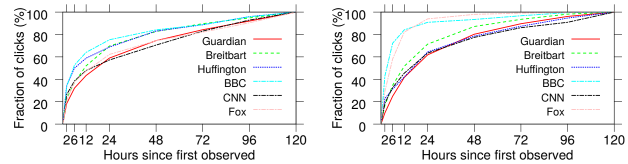
Twitter data have been used to study a wide range of popularity and information sharing dynamics, including the 24-hour news cycle [3], the impact of retweets [4], and to quantify influence [5], [6]. Twitter data have also been used to study the spreading of fake news [7], [8] or the online diffusion in general [9]. With surprising little work having used click data, we believe that the methodology and the tools presented here will greatly help towards better understanding the actual news consumption related to some of the above aspects.

Most closely to our work is the work by Gabielkov et al. [1]. Combining Twitter and Bitly data, they show that a snapshot of the number of times that Bitly links are retweeted do not map well to the number of clicks for the same links. In contrast, our methodology uses periodic Bitly calls to collect longitudinal statistics and to gain insights into differences in how the clicks changes over time for links to articles of different age and/or that are associated with different news websites.

Others have studied the actual accesses to news articles (published by Al Jazeera English) and the redirects to these articles via social media [10]. Content sharing, likes, and actual content viewing have also been evaluated in in other contexts, including for YouTube and other UGCs that allow the view counts (similar to clicks) to be easily collected for each video [2], [11], [12]. Click-through-rates and similar metrics have also been used to measure the quality of ads [13], [14].

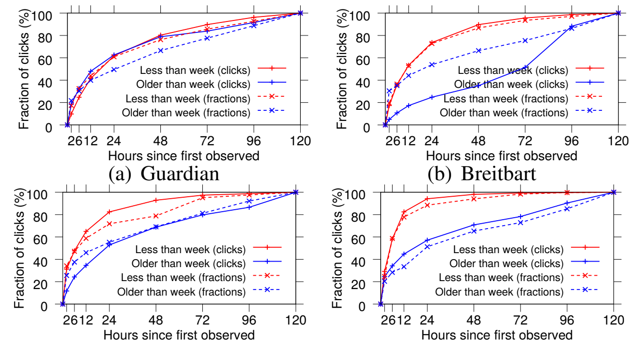
VII. Conclusions

This paper makes several contributions. First, we develop a novel measurement framework for collecting and analyzing the timelines of all clicks and retweets associated with the Bitly links pointing to articles published by seven popular news websites. Second, we use the methodology to collect temporal five-day datasets (per link) that contain periodically collected click and retweet statistics associated the links discovered during a seven-day period (April 12-19, 2018). Third, we present a temporal characterization and analysis based on the dataset. Interesting observations include (i) significant differences in the clicks-per-retweet ratios of individual links, including a significant number of links for which there are more retweets than clicks, (ii) significant age biases, including relatively



(a) Average (over fractions) (b) Average (over all clicks)

Fig. 14. Cumulative click over time for different news websites.



(a) Guardian (b) Breitbart (c) CNN (d) Fox

Fig. 15. Per-website-based age analysis: young vs old articles.

high initial click rates for articles younger than a week and much more stable click rates for older, long-term popular articles, and (iii) significant age-dependent popularity skews and churn rates. These example findings have implications on the spreading of fake news (e.g., links that are retweeted more than they are clicked suggest that some people/bots retweet links without reading them), and highlights the importance of not using retweets as a proxy for clicks/reads, of distinguishing where in the lifecycle different links are, and the impact that this has on the rate they generate clicks, for example. Finally, temporal findings are validated using both data from May 2017 and a per-website-based analysis.

References

[1] M. Gabielkov *et al.*, "Social clicks: What and who gets read on twitter?" in *Proc. ACM SIGMETRICS*, 2016.
[2] Y. Borghol *et al.*, "Characterizing and modelling popularity of user-generated videos," *Perform. Eval.*, vol. 68, no. 11, pp. 1037–1055, 2011.
[3] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proc. ACM KDD*, 2009.
[4] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proc. WWW*, 2010.
[5] M. Cha *et al.*, "Measuring user influence in twitter: The million follower fallacy," in *Proc. ICWSM*, 2010.
[6] E. Bakshy *et al.*, "Everyone's an influencer: quantifying influence on twitter," in *Proc. ACM WSDM*, 2011.
[7] D. M. J. Lazer *et al.*, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.
[8] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.
[9] S. Goel *et al.*, "The structural virality of online diffusion," *Management Science*, vol. 62, no. 1, pp. 180–196, 2015.
[10] C. Castillo *et al.*, "Characterizing the life cycle of online news stories using social media reactions," in *Proc. ACM CSCW*, 2014.
[11] M. Cha *et al.*, "Analyzing the video popularity characteristics of large-scale user generated content systems," *IEEE/ACM Trans. on Netw.*, vol. 17, pp. 1357–1370, 2009.
[12] S. Mitra *et al.*, "Characterizing web-based video sharing workloads," *ACM Trans. on the Web*, vol. 5, pp. 8:1–8:27, 2011.
[13] A. Farahat and M. C. Bailey, "How effective is targeted advertising?" in *Proc. WWW*, 2012.
[14] H. B. McMahan, G. Holt, D. Sculley, M. Young, and D. Ebner, "Ad click prediction: a view from the trenches," in *Proc. ACM KDD*, 2013.