SCHOOL OF INFORMATION CENTER FOR
SOCIAL MEDIA RESPONSIBILITY
UNIVERSITY OF MICHIGAN

Iffy Quotient: A Platform Health Metric for Misinformation

Paul Resnick[1], Aviv Ovadya[2], and Garlin Gilchrist[3]

v1: October 10, 2018 (link)
Latest version can always be accessed at http://umsi.info/iffy-quotient-whitepaper

Executive Summary

Social media sites and search engines have become the de facto gatekeepers of public communication, a role once occupied by publishers and broadcasters. With this new role come public responsibilities, including limiting the spread of misinformation.

Externally maintained metrics offer a way to measure the progress of media platforms at meeting their public responsibilities. By contrast with the current environment of accountability by anecdotes, Platform Health Metrics can focus attention on the overall performance of platforms rather than on bad outcomes in individual cases.

The Center for Social Media Responsibility at the University of Michigan School of Information has developed the Iffy Quotient, a metric for how much content from “Iffy” sites has been amplified on Facebook and Twitter. We use the term “Iffy” to describe sites that frequently publish misinformation. It is a light-hearted way to acknowledge that our categorization of the sites is based on imprecise criteria and fallible human judgments. We are publishing a web-based dashboard that charts the Iffy Quotient since early 2016. The dashboard enables comparisons over time and between platforms. This report describes the calculation of the Iffy Quotient in detail, discusses some of its potential limitations, and analyzes some of the trends.

Consonant with previous claims that there was a major uptick in sharing of misinformation in the lead up to the 2016 U.S. Presidential election, on both Facebook and Twitter the Iffy Quotient approximately doubled from January to November of 2016. The Iffy Quotient was higher at Facebook than Twitter in 2016 and into 2017. However, there has been gradual improvement in Facebook’s Iffy Quotient since mid-2017, with a substantial cumulative impact. Facebook has now returned to its early 2016 levels and is at or below Twitter’s level most days. The contrast between Facebook and Twitter is even more pronounced in an engagement-weighted version of the Iffy Quotient, which we can think of as a proxy for the fraction of total user attention. In 2016 the Iffy sites’ share of attention was about twice as high on Facebook as Twitter; now it is 50% higher on Twitter.

[1] Paul Resnick is a Professor and Associate Dean for Research at the University of Michigan School of Information. He has been a consultant to Facebook but this work is independent.
[2] Aviv Ovadya served as Chief Technologist for the Center for Social Media Responsibility with primary responsibility for architecting the Iffy Quotient. He continues to serve as a consultant to the Center.
[3] Garlin Gilchrist is the Executive Director of the Center for Social Media Responsibility, currently on unpaid leave while he campaigns for political office. He helped to draft this report prior to taking leave.

SCHOOL OF INFORMATION CENTER FOR
SOCIAL MEDIA RESPONSIBILITY
UNIVERSITY OF MICHIGAN

Report

Introduction

Social media sites and search engines have become the de facto gatekeepers of public communication, a role once occupied by publishers and broadcasters. With this new role come public responsibilities, beyond the commercial responsibilities that a company has to please customers and reward shareholders. Among these public responsibilities are limiting the spread of misinformation, ensuring a level playing field in the free competition of ideas, and promoting interpersonal connections that heal rather than aggravate societal divisions.

The Center for Social Media Responsibility is developing Platform Health Metrics, which track how well social media sites and search engines (which we refer to collectively as media platforms) are meeting these public responsibilities. A metric reduces an abstract ideal, such as limiting the spread of misinformation, to a concrete measurement that can be taken repeatedly, enabling comparisons over time and between platforms. This report introduces the Iffy Quotient, one such metric. It computes the fraction of the most popular URLs that come from Iffy sites - sites that have frequently published misinformation and hoaxes in the past. Our website reports the Iffy Quotient for Facebook and Twitter going back to 2016 and it will be updated on an ongoing basis.

For both Twitter and Facebook, there was an increase in attention to Iffy sites in the runup to the 2016 U.S. elections. The Iffy Quotient nearly doubled on each site from April to November, peaking right around election day.

Key Performance Metrics are powerful management tools for media companies. Mature consumer-facing technology platforms already maintain internal suites of metrics such as monthly page views, clickthrough rates, dwell times, customer acquisition and retention, and ad revenue. These metrics strongly influence decisions about changes to products and policies. Typically, product managers are rewarded for improving some primary metric, subject to the constraint that there is at most a modest decline in other metrics.

Externally maintained metrics offer two advantages over internal metrics maintained by the platforms. First, they can draw attention to issues that platforms may either not be tracking themselves or not prioritizing as much as the public would like. This form of public accountability is preferable to the current environment of accountability by gotcha anecdotes. It focuses attention on the overall performance of platforms rather than on bad outcomes in individual cases; some bad outcomes may be inevitable given the scale on which the platforms operate.

Second, external metrics can create public legitimacy for claims that platforms make about how well they are meeting public responsibilities. Even if Facebook actually reduces the audience share for Iffy content, the public may be skeptical if Facebook defines the metric, conducts the measurement without audit, and chooses whether to report it.

Of course, metrics are not a panacea. The thing that is measured is often a proxy for the thing that really matters. Managerial efforts to improve the metric (the proxy) may not similarly improve the true quantity of interest; in extreme cases this is referred to as “gaming the metric”. Externally-maintained metrics may be especially susceptible to such problems. Due to limited access to proprietary data, an external metric may involve compromises that make it a weaker proxy. This report also describes some of the limitations and compromises involved in defining and measuring the Iffy Quotient, and the potential risks that come from them.

Limiting the Reach of Misinformation is a Public Responsibility of Media Platforms

Media platforms should not be expected to prevent the publication of misinformation, or prevent people who seek access to it from finding it (unless the misinformation is also harmful in some way that is prohibited by law, such as by directly inciting violence). We think, however, that media platforms should not amplify misinformation—as Renee Diresta concisely describes “Free speech is not the same as free reach”[4]. Misinformation, if widely shared, can influence public opinion, create social divisions, and even stir up violence, as has been documented in India, Sri Lanka, and Myanmar.[5] It can degrade trust, which is necessary for society to function well. And it can drive government actions that benefit special interests rather than public interests. It is especially important that special interests, including foreign actors, not be able to manipulate media platforms so that they spread misinformation.

Facebook and Twitter have accepted responsibility for countering deliberate manipulation. For example, Twitter executive Colin Crowell wrote on the company blog in 2017, “We’re working hard to detect spammy behaviors at source, such as the mass distribution of Tweets or attempts to manipulate trending topics.”[6] Facebook reported that it removed 583 million fake accounts in the first quarter of 2018.[7] The two seem to diverge somewhat, however, in whether they accept responsibility for not amplifying misinformation in the absence of deliberate manipulation. Facebook appears to implicitly accept this responsibility, with their announcement that they will take action to reduce the audience for items that journalist fact-checkers judge to be false[8]. Crowell’s post, on the other hand, suggests that Twitter thinks its only responsibility is to ensure the spread of counter-information to misinformation.

[4] https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/

[5] https://www.nytimes.com/2018/07/18/technology/facebook-to-remove-misinformation-that-leads-to-violence.html

[6] https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html

[7] https://newsroom.fb.com/news/2018/05/enforcement-numbers/

[8] https://www.facebook.com/help/1952307158131536

Whether or not they accept responsibility for preventing the amplification of misinformation, executives from both sites argue that they should not become “arbiters of truth”. This still leaves room, however, for other kinds of counter-measures that could reduce the platforms’ amplification of misinformation. One approach focuses entirely on process, eschewing assessments of truth: for example, identifying and eliminating bot accounts. Another is to outsource judgments of information reliability, to journalists, third-party fact-checkers, media watchdogs, or platform users. This report does not argue for or against any particular counter-measures; the Iffy Quotient merely provides a way to measure the effectiveness of any counter-measures that may have been implemented.

Defining Misinformation

Exactly what counts as misinformation has become politically contested, and is an active front in the ongoing culture wars. Many factual claims are embedded in contextual spin that may invite misinterpretation without actually making a factual claim. Even with factual claims, only limited evidence may be available. Not everyone may agree, given the available evidence, whether the claim is true or false.

We begin with a recipient-centric, subjective definition of misinformation: information that *the recipient would judge* to be false or misleading *if they took the time to carefully consider all of the evidence about it*. With that in mind, the platform’s responsibility, we argue, should be to spread a piece of information only to those people who would, in full knowledge of available evidence, access and spread it.

However, media platforms take many actions at an aggregate, non-personalized level. That makes it useful to adopt a collective, but still subjective, definition of misinformation: information that *most recipients would judge* to be false or misleading *after taking the time to carefully consider all of the evidence about it*. With that definition, the platform’s responsibility is to support the amplification of a piece of information if a majority of the potential audience, in full knowledge of available evidence, would access and spread it.

These definitions are based on a counterfactual: how individuals would judge an item *if* they considered the evidence for and against it. In an ideal world, there would be a process that resolved that counterfactual for enough people to yield a good estimate of the collective judgment. In the future, automated classifiers might be used as a proxy, with verification against human judgments on a sample of items. In current practice, we rely on external entities as a proxy to speak for what the collective judgment would be.

Site-level Proxy Judgments

In our case, we rely on two external entities, Media Bias/Fact Check[9], and Open Sources[10]. They make judgments not on individual items but on entire sites. We treat their judgements as a proxy for whether a particular content site frequently publishes information that the majority of people would consider false or misleading, after consider all evidence. Because these external entities make their judgments based on imprecise criteria, we adopt the whimsical term “Iffy” rather than the more definitive term “misinformation”.

Site-level judgements alone would not be appropriate for platforms to use in making decisions about whether to amplify the audience for particular items, for two reasons. First, even sites that frequently publish misinformation may also publish some things that are not misleading. Second, the judgments made by Media Bias/Fact Check and Open Sources, according to the criteria they define and articulate, may not match what the majority of people would consider false or misleading. This reasoning is evident in the announced practices of media platforms. For example, Facebook takes action to reduce the audience for items that journalist fact-checkers judge to be false[11], but, to our knowledge does not rely on the site-level judgments of Media Bias/Fact Check or Open Sources.

It is reasonable, however, to use site-level judgments in calculating the Iffy Quotient. Treating all items from a site as iffy will give an overestimate of the absolute amount of misinformation distributed by the platform. The amount of overestimation should be fairly stable, however, and thus should not affect comparisons between sites and comparisons over time. Similarly, mistaken judgments about whole sites could lead to errors in the absolute percentage of Iffy content as recorded in an Iffy Quotient measurement, but are unlikely to have a large effect on comparisons between measurements. Thus, the Iffy Quotient is suspect as a measure of the absolute amount of misinformation that is spread by platforms, but is a reasonable way to judge whether Twitter or Facebook spread more misinformation at a particular point in time, whether Twitter spread more or less Iffy content in July 2018 vs. July 2017, or whether a change in media platform moderation policy enacted on a particular date led to less amplification of Iffy content the following month.

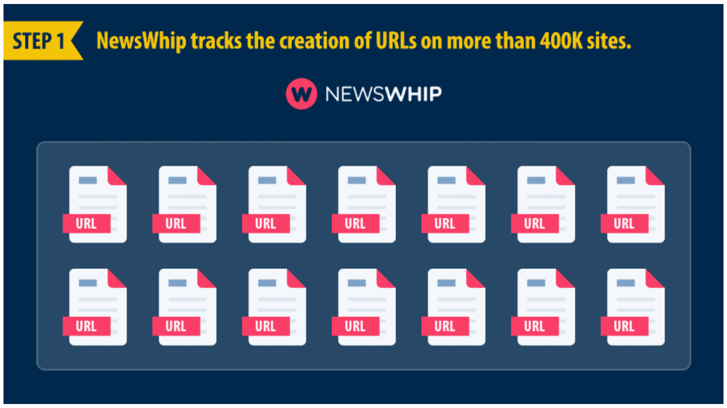
Calculating the Iffy Quotient for English Language News

For each day, we download the most popular URLs on Facebook and Twitter from NewsWhip[12], a commercial social media monitoring company. We download site judgments from Media Bias/Fact Check and Open Sources about which sites are Iffy, and compute “inferred

[9] https://mediabiasfactcheck.com
[10] http://www.opensources.co
[11] https://www.facebook.com/help/1952307158131536
[12] https://www.newswhip.com

judgments” for urls that redirect to and from those sites. Finally, we calculate the fraction of urls from Iffy sites and report a 7-day moving average to smooth out daily fluctuations. Details follow.

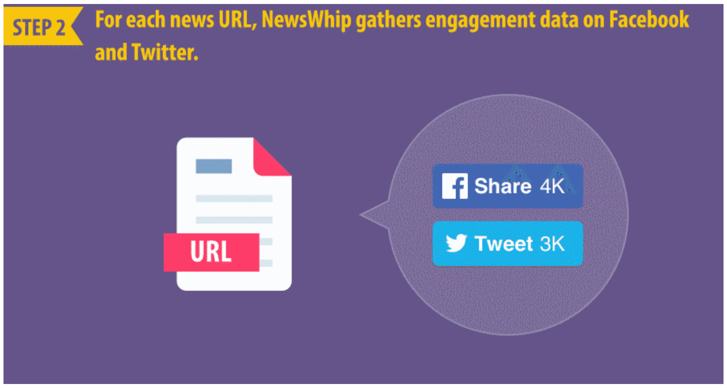
1. NewsWhip tracks the creation of URLs on more than 400,000 sites every day.


STEP 1
NewsWhip tracks the creation of URLs on more than 400K sites.
NEWSWHIP
URL

NewsWhip maintains a list of sites that it monitors. This includes traditional news sites, sites that people treat like news, and other sites that are relevant to NewsWhip's clients. This does not include all websites, but it is quite broad.

Limitation: It is possible that the incompleteness of NewsWhip’s tracking could lead to incorrect estimation of the true Iffy Quotient. For example, if NewsWhip is less effective at discovering high-engagement fly-by-night sites (e.g., Macedonian traffic arbitrage) than it is at tracking established reliable sources, a higher fraction of the missed sites may be Iffy, leading to our measured value being an underestimate. Or it could go the other way, if NewsWhip misses more URLs from untracked reliable sites (perhaps small niche sites) than from Iffy sites.

2. For each news URL, NewsWhip gathers engagement data on Facebook and Twitter.

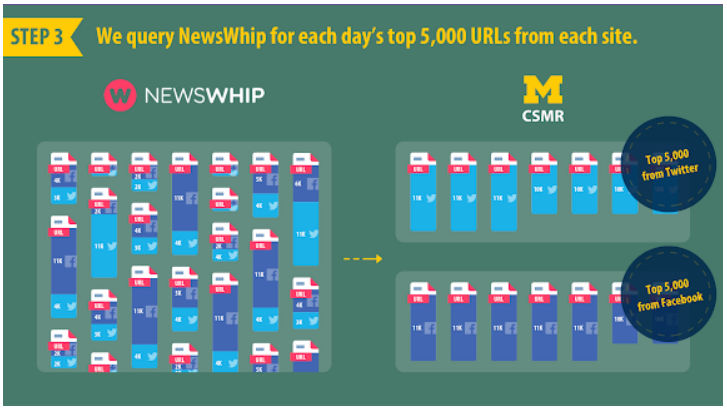

STEP 2
For each news URL, NewsWhip gathers engagement data on Facebook and Twitter.
URL
Share 4K
Tweet 3K

NewsWhip tracks new URLs added to the sites it monitors. Whenever a new page is added to one of these sites, NewsWhip checks for social engagement with that URL on Facebook and Twitter as described below. NewsWhip also continues to check social engagement repeatedly, though more and more rarely as the number of engagements appears to stabilize.

NewsWhip provides a Facebook engagement score for each URL, which appears to closely track data available through the public Facebook Graph API. It provides an indicator of aggregate engagements (likes and other reactions, shares, comments) with a URL without revealing the identity of any particular person who engages with it. NewsWhip associates all engagement data with the URL and its publication date, regardless of when the engagement occurred on Facebook.

NewsWhip also provides a "Twitter Influencer Shares" value for URLs. This is the sum of tweets and retweets mentioning a URL by the "influencers" that NewsWhip tracks. NewsWhip tracks at least 300,000 accounts, including verified Twitter accounts and other accounts that are useful to NewsWhip's clients. This approach has been used starting around November 27th 2017. Prior to that, NewsWhip used a different method to estimate the number of tweets and retweets. Exact numbers before and after the changeover may not be comparable. The change should, however, affect URLs from Iffy sites and other sites in a similar way, and thus should not appreciably affect the Iffy Quotient.

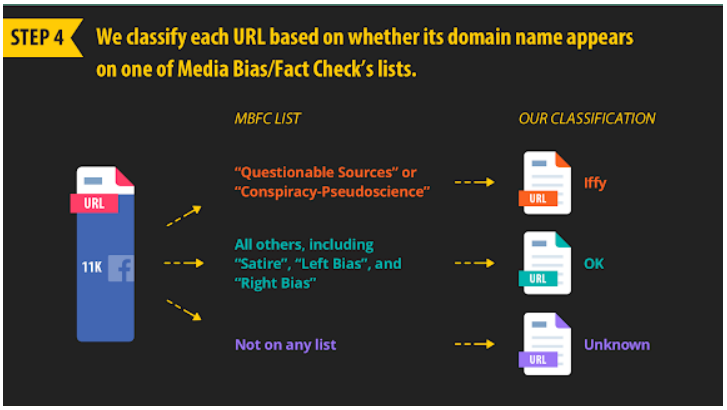
3. We query NewsWhip for each day's top 5,000 URLs from each site.


STEP 3
We query NewsWhip for each day's top 5,000 URLs from each site.
NEWSWHIP
CSMR
Top 5,000 from Twitter
Top 5,000 from Facebook

We query NewsWhip for 5,000 URLs published on each date. Since engagements are still trickling in for recently published content, the list, and thus the Iffy Quotient, may be slightly unstable for the most recent few days.

Limitation: Fake accounts may be used to inflate the engagement counts on Twitter and Facebook, as a way to make them appear more popular and thus drive real traffic. Twitter and Facebook try to root out such fake accounts, and NewsWhip tries to avoid inclusion of fakes among the Twitter accounts that it tracks, but those attempts can never be completely successful. Fake accounts are perhaps more likely to be used for content from Iffy sites than other sites. This could lead to an overestimate of the true Iffy Quotient, by causing more Iffy sites to creep into the top 5,000 than would be there if no fake accounts were included in engagement scores.

4. We classify each URL based on whether its domain name appears on one of MBFCs lists.



STEP 4 We classify each URL based on whether its domain name appears on one of Media Bias/Fact Check's lists.

MBFC LIST

OUR CLASSIFICATION

"Questionable Sources" or "Conspiracy-Pseudoscience" → Iffy

All others, including "Satire", "Left Bias", and "Right Bias" → OK

Not on any list → Unknown

Given a URL, we classify it by comparing the hostname to the hostnames listed in the *source lists* curated and maintained by Media Bias/Fact Check and Open Sources. Our primary source is Media Bias/Fact Check, because it has broader coverage and has continued to be updated; we also compute an Iffy Quotient based on the other source, as a robustness test.

Media Bias/Fact Check (hereafter "MBFC") evaluates sites and puts them on one or more of the lists, based on criteria they describe for each of the lists. We classify as Iffy any site that is on either the "Questionable Sources" list or the "Conspiracy-Pseudoscience" list. MBFC's website describes the criteria for each as follows:

> A questionable source exhibits *one or more* of the following: extreme bias, overt propaganda, poor or no sourcing to credible information and/or is fake news. Fake News is the *deliberate attempt* to publish hoaxes and/or disinformation for the purpose of profit or influence. Sources listed in the Questionable Category *may* be very untrustworthy and should be fact checked on a per article basis.
>
> Sources in the Conspiracy-Pseudoscience category *may* publish unverifiable information that is *not always* supported by evidence. These sources *may* be untrustworthy for credible/verifiable information, therefore fact checking and further investigation is recommended on a per article basis when obtaining information from these sources.

MBFC also provides explicit listings of sites it has evaluated and judged *not* to be appropriate for one of those lists. Other lists they provide include "Left-Bias", "Left-Center Bias", "Least Biased", "Right-Center Bias", "Right Bias", "Pro-Science", and "Satire". We classify as "OK" any site that is not Iffy and is on one of these other lists. If a site is not on any of MBFC's lists, we classify it as Unknown.

Complete listing of the MBFC judgments are available on its website. To give a flavor for the judgments:

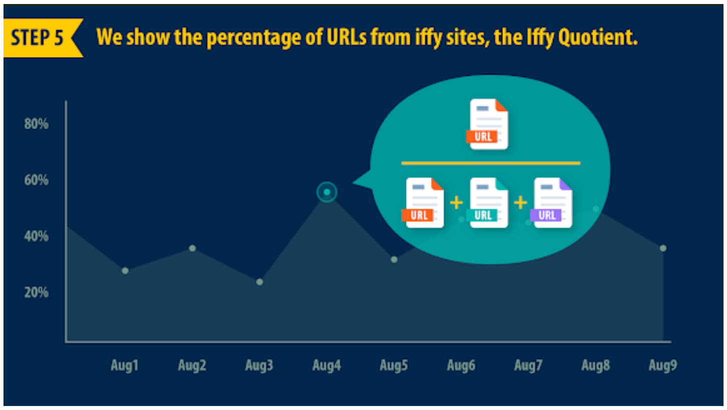
- Vox and the Upworthy are classified as OK (Left Bias)
  but Learn Progress and Occupy Democrats are Iffy (Questionable Source).
- Fox News and the Drudge Report are classified as OK (Right Bias)
  but Breitbart and TruthFeed are Iffy (Questionable Source);

Open Sources, like MBFC, tags sites that it evaluates according to a variety of categories. For the IffyQuotient computed using Open Sources, we classify sites as Iffy if Open Sources labels them as any one of: bias, conspiracy, hate, fake, junksci, or rumor. Open Sources has evaluated many fewer sites than MBFC, and the last update prior to this report was April 28, 2017.[13]

Publishers sometimes change their domain names. This is especially true for Iffy publishers who may change domain names to get a fresh start with search engines and social media sites that may have started to demote or demonetize them based on prior complaints and investigations. MBFC and Open Sources may not always notice these name changes right away. Moreover, when they do, they may put the new site on their lists but remove the old sites.

To account for incompleteness of some of our classifications that might result from sites changing their domain names, we test for automatic redirects and compute some inferred labels. In particular, if a URL published some time ago now redirects to a site that MBFC or Open Sources has listed, we give it the same classification as the new site. Similarly, if one of the sites that MBFC or Open Sources has listed now redirects to a new site, we also give the new site the same classification as the original.

5. We show the percentage of URLs from Iffy sites.


STEP 5
We show the percentage of URLs from iffy sites, the Iffy Quotient.
80%
60%
40%
20%
URL
URL
URL
URL
Aug1
Aug2
Aug3
Aug4
Aug5
Aug6
Aug7
Aug8
Aug9

Given the classification of each URL as Iffy, Unknown, or OK, we define the Iffy Quotient as the fraction of all sites that are Iffy. We calculate this separately for Facebook and Twitter, for each date. To smooth the graph, we apply a seven-day moving average. That is, we compute the raw Iffy Quotient for each date, but report the average of the raw Iffy Quotients for the previous seven days.

The fraction of URLs classified as Unknown is substantial, in the range of 40-60% for most dates. Some sites may never be queued for judgment by MBFC, notably user-generated content platforms such as Medium and YouTube, sites that do not publish news and information at all, and sites that publish only domestic news in countries outside the U.S. The inclusion of popular URLs from such sites inflates the denominator and leads to an artificially lower value for the Iffy Quotient. This is another reason not to place too much emphasis on the absolute value of the Quotient, but rather to use it primarily for comparisons across platforms and across time, where the inflation of the denominator happens in all values that are compared.

[13] https://github.com/BigMcLargeHuge/opensources/commits/master

Limitation: Some of the Unknown URLs may be from sites that MBFC would judge as Iffy but hasn't gotten around to judging yet. This could make comparisons over time be misleading. For example, if MBFC fails to keep up with newly-emerging sources of misinformation, the Iffy Quotient could decline over time even though the fraction of popular URLs from truly Iffy sites did not decline. Or, if MBFC clears a backlog, the Iffy Quotient could go up just as a result of better detection. For this reason, we publish a second chart that shows not only the fraction of Iffy sites but also the fractions of OK and Unknown sites. If the OK fraction stays steady or increases, while the Iffy fraction declines, we can be more confident that the decline reflects a real phenomenon and not just MBFC falling behind in its evaluation of sites.

As a supplementary analysis, we also compute an engagement-weighted version of the Iffy Quotient. Rather than treating all popular URLs as equal, we weight them by the estimated engagement scores that NewsWhip provides us. The denominator, then, is the sum of engagement scores for the top 5,000 URLs, and the numerator is the sum of engagement scores for those URLs that are from Iffy sites. Arguably, this version is a closer proxy for the real quantity of concern, the fraction of human attention to unreliable information. However, it depends more heavily on the accuracy of the engagement estimates than the unweighted Iffy Quotient, which depends on the estimates only to select the top 5,000. Moreover, the weighted version is more volatile from day to day, and more sensitive to the classification decisions for the few most engaged with URLs.

In summary, here's how we create our main Iffy Quotient chart.

1. NewsWhip provides the 5,000 most engaged-with URLs each day, on Facebook and Twitter.
2. Media Bias/Fact Check provides lists of domain names they have judged.
    a. We define as Iffy those sites listed as Questionable Sources or Conspiracy/Pseudoscience
    b. We define as OK those sites listed in other categories, including Left Bias and Right Bias
3. We check for automatic redirects to infer categorization of additional domain names.
4. For each site, the Iffy Quotient is the fraction of the day's 5,000 URLs that are from domain names categorized as Iffy.
5. We report a seven-day moving average to smooth the chart.

Analysis

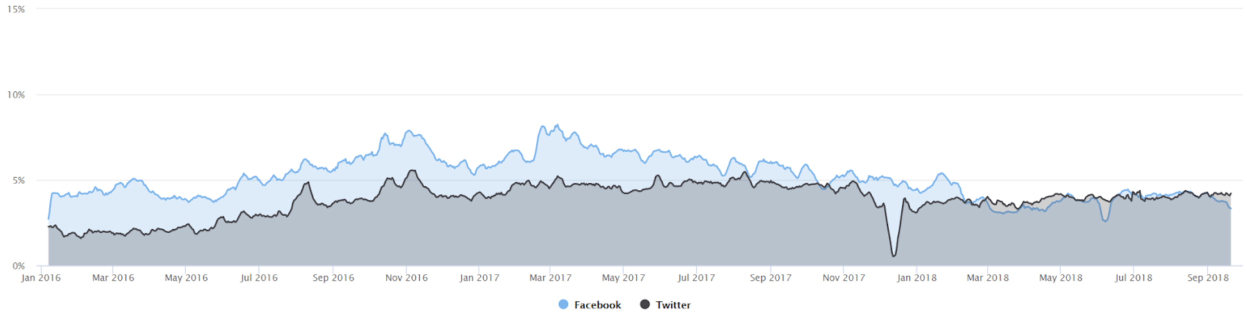


Figure 1. The Iffy Quotient for Facebook and Twitter, dating back to 2016. See the website for an up-to-date, dynamic version where you can hover over points to see data for particular dates.

Figure 1 shows the Iffy Quotient dating back to early 2016, for both Twitter and Facebook. For reasons described in the previous sections, the absolute percentage is not particularly meaningful. An Iffy Quotient of 5% does not mean that 5% of all URLs contained misinformation, but that 5% of URLs came from sites that MBFC judged to be in a category that we have labeled Iffy. Moreover, NewsWhip might have missed some URLs and MBFC might have failed to label some sites or mislabeled them. The Iffy Quotient is far more meaningful as a way to make comparisons across time and between platforms, especially focusing on stable trends rather than individual dates.

First, notice the temporal trends. For both Twitter and Facebook, there was an increase in attention to Iffy sites in the runup to the 2016 U.S. elections. The Iffy Quotient nearly doubled on each site from April to November, peaking right around election day. On Facebook, there has been a clear downward trend since about March of 2017, with the Iffy Quotient in September 2018 slightly lower than it was early in 2016. On Twitter, the Iffy Quotient remained near record highs through 2017, declining slightly in 2018 but still at nearly twice the level of early 2016.[14]

There are several factors that plausibly contribute to these temporal trends. First, in the run-up to the elections, and afterwards when people were unusually politically-activated, the general public had more interest in political news - especially in sensational political news - than in other time periods. URLs from Iffy sites may have been able to appeal to that audience interest better than URLs from other sites. This explanation is consistent with reports of Macedonian sites with no political agenda earning advertising revenue by posting invented or copied political stories.[15] Second, both domestic and foreign publishers with political agendas may have expended more time and money to spread misinformation during the run-up to the election.

[14] We think that the big, temporary dip for Twitter at the end of 2017 is just an artifact of poor data quality for that period.
[15] https://www.wired.com/2017/02/veles-macedonia-fake-news/

After the election, in addition to demand and supply declining, the platforms took some actions to reduce the spread of Iffy content. For example, in December 2016, Facebook announced a partnership with third-party fact-checkers, sending them questionable stories and showing lower in the feed those that the fact-checkers labeled as false[16]. On January 11, 2018, Facebook announced that it would reduce the reach of all public external content, in favor of native posts from friends and family.[17] On its own, that wouldn't affect the Iffy Quotient, which is based on whatever public content is most popular. However, that announcement and one the following week also implied other changes that might have affected the Iffy Quotient.[18] One was prioritizing content around which people interacted with friends; it could be that people interact less around content from Iffy sites. Another was prioritizing news that the community rates as trustworthy, that people find informative, and that is local. Without knowledge of exactly when particular product features or policy changes were rolled out, we are not able to assess the impacts of particular initiatives on the Iffy Quotient. However, there has been a long-term decline in Facebook's Iffy Quotient since March 2017. On August 6, Facebook announced a ban on several pages associated with InfoWars host Alex Jones. That did not have an immediate impact on the Iffy Quotient, which hovered right around 4% in the weeks before and after. Twitter waited until September 6 to take similar steps; again there was not an immediately measurable effect on Twitter's Iffy Quotient.

Next, notice the difference between Twitter and Facebook. From 2016 through the middle of 2017, more Iffy sites gained attention on Facebook than on Twitter. By early 2018, they had equalized, and in March, April, and September Twitter had more of its popular URLs coming from Iffy sites than Facebook did. Again, it is not entirely clear why the two platforms have changed positions. Presumably, however, there should have been similar fluctuations in supply and demand for Iffy content at the two sites, with the major difference being policies and technical features. Facebook may have been more successful at detecting and countering fake accounts and manipulation campaigns, more aggressive in discounting ranking signals that are associated with Iffy sites, or more aggressive in demoting particular articles and sources.

[16] https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/
[17] https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/
[18] https://newsroom.fb.com/news/2018/01/trusted-sources/

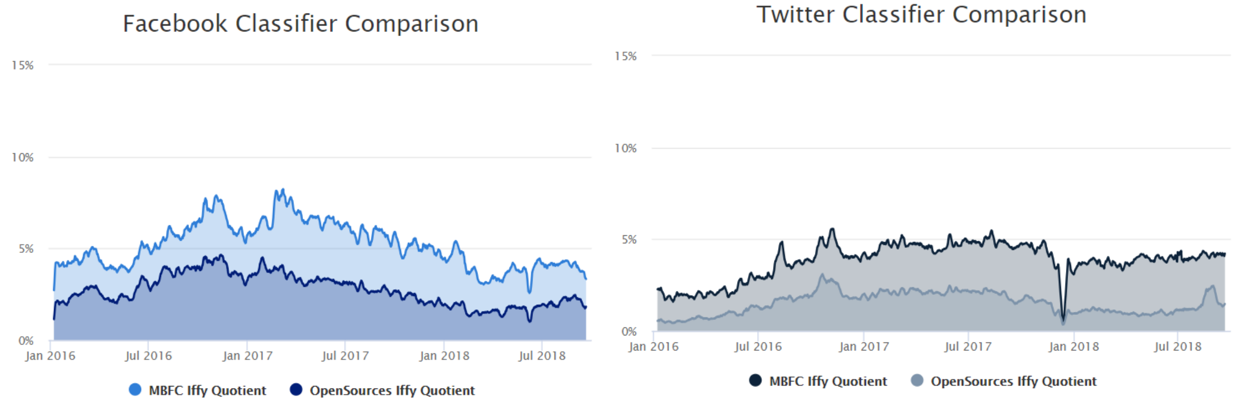


Figure 2. Comparison of Iffy Quotient computed using MBFC and Open Sources lists. On the left side, the dark blue area is computed using Open Sources' categorizations while the light blue is the same as the light blue line in Figure 1, Facebook's Iffy Quotient computed with the MBFC categorization). Similarly, on the right side, the light grey area with the black line is the same as that charted in Figure 1, for Twitter using MBFC categorization.

We get similar trends, though the absolute values are smaller, when we classify sites according to the Open Sources lists, rather than Media Bias/Fact Check as shown in Figure 2.

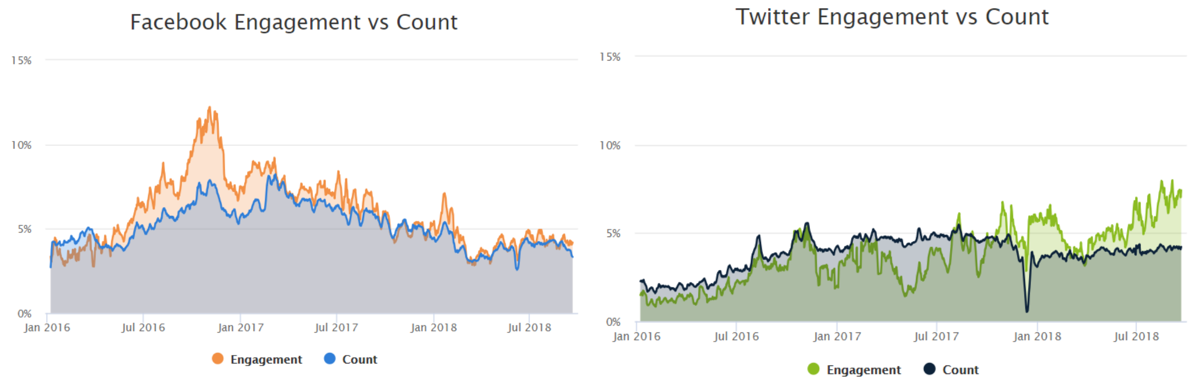


Figure 3. Engagement-weighted vs. count-based Iffy Quotients. Again, on the left the light blue line is the same as the corresponding line in Figure 1, the count-based Iffy Quotient for Facebook. The orange line shows the engagement-weighted Iffy Quotient. On the right side, for Twitter, the black line matches that in Figure 1, while the green shows the engagement-weighted version.

Figure 3 compares the engagement-weighted and count-based Iffy Quotients for Facebook and Twitter. For Facebook, in late 2016 and early 2017, the engagement-weighted Iffy Quotient was noticeably higher than the count-based. This suggests that not only were more URLs from Iffy sites getting into the daily top 5,000 during this period, they also were getting more than the average engagement among those 5,000. After the first quarter of 2018, that difference disappeared. That is, not only did fewer of the popular URLs comes from Iffy sites; they also lost their advantage in average engagement among the popular URLs. For Twitter, on the other hand, while the fraction of popular URLs from Iffy sites went down slightly in 2018 compared to 2017, the fraction of total engagement went up. At first glance, this might appear to be related to the change in how NewsWhip calculated its engagement estimates for Twitter on November 27, 2017, but the lines cross earlier, in mid October. Over the past eight months, the fraction of popular URLs on Twitter from Iffy sites was fairly stable, but the Iffy sites gained a larger share of the total tweets and retweets.

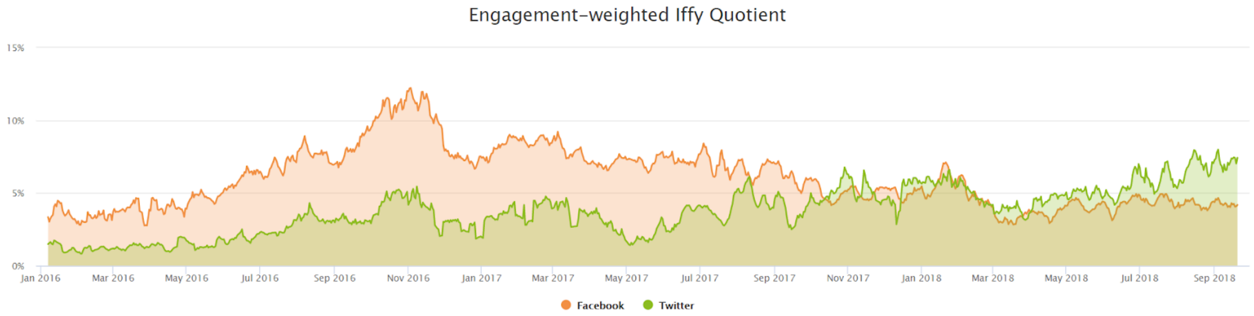


Figure 4. Engagement-weighted Iffy Quotient for Facebook and Twitter.

The engagement-weighted Iffy Quotient (Figure 4) presents an even starker contrast between the trajectories of Facebook and Twitter than the count-based Iffy Quotient. In 2016, Iffy sites' share of engagement on Facebook was nearly twice their tweet share. In late 2018, the positions had reversed, with Iffy sites' tweet share about 50% higher than their share of Facebook engagements.

This picture is consonant with the recent finding of Alcott, Gentkow, and Yu.[19] They used another commercial service, BuzzSumo, to estimate the total monthly tweets and Facebook engagements for a set of 570 sites that "have been identified as producers of false stories", analogous to our notion of Iffy sites. They found that total Facebook engagements dropped by nearly two-thirds from the end of 2016 to July 2018, while the tweets about URLs from those Iffy sites increased slightly. One advantage of our approach is that, by expressing the attention share of Iffy sites as a fraction of that for all popular sites on each platform, we are able to compare the Iffy Quotient for the two platforms directly at any point in time.

[19] http://web.stanford.edu/~gentzkow/research/fake-news-trends.pdf

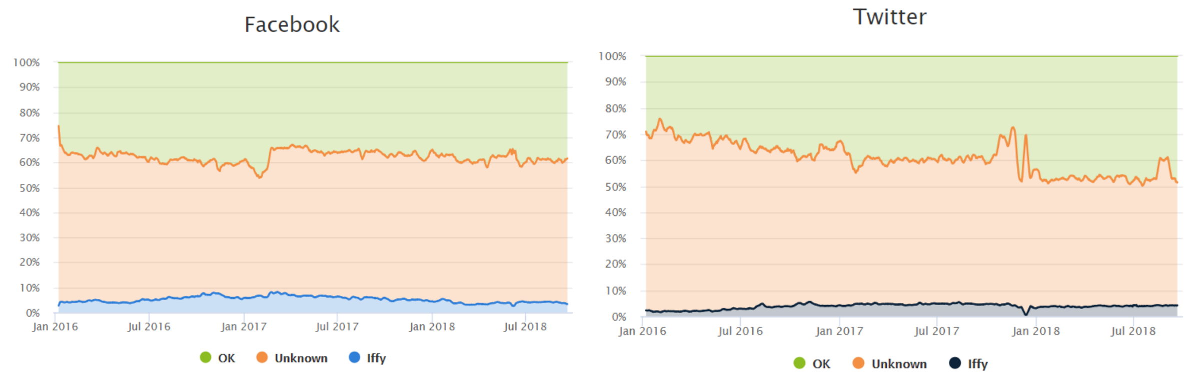


Figure 5: The share of OK and Unknown sites, as well as Iffy sites.

As described in the previous section, an observed decline in the Iffy Quotient could reflect the URL lists becoming stale. Figure 5 shows that the fraction of popular URLs coming from known OK sites stayed stable during or slightly increased during 2017 and 2018. The stability of the OK fraction throughout 2018, for both Twitter and Facebook, suggests that MBFC's lists have not become stale.

Tracking Changes in the Iffy Quotient

Beyond the basic trends we have identified above, how should readers make use of the Iffy Quotient? One way is to try to track the impact of external events and of platform technology and policy changes. This can be done retrospectively, as we did for the 2016 elections and Facebook and Twitter's sanctions against Alex Jones. We are happy to provide data in other formats to anyone who would like to produce infographics.

Tracking can also be done prospectively. When Twitter or Facebook announce a counter-measure, such as deletion of a large number of fake accounts, journalists can start to track the Iffy Quotient and see if it changes.

Or you can sign up for alerts. Join our email list by sending a request to csmr-info@umich.edu. We will send out an alert when there is a significant change in the Iffy Quotient for either Facebook or Twitter that is sustained over several days, as well as other general announcements from the Center for Social Media Responsibility such as the release of additional Platform Health Metrics.

What’s Next for the Iffy Quotient?

Our website that tracks the Iffy Quotient will automatically update daily. We will be making iterative improvements and welcome partnerships with organizations that can help us validate or improve the metric. In the URL collection phase, we would welcome other sources that could be combined with what we get from NewsWhip. In the classification phase, we would like to reduce the large number of Unknown sites. We would be happy to include other classifiers besides Media Bias/Fact Check and Open Source; these could be either human or automated.

We would also like to be able to filter out popular URLs that have nothing to do with news, politics, public affairs, science or health. That would make the absolute value of the Iffy Quotient be a more meaningful quantity, the fraction of popular URLs from Iffy sites, among those where reliability of the information matters. It would be especially useful to have an automated classifier that operated on individual URLs rather than entire sites. We would be happy to partner with anyone who has trained a news and public affairs classifier.

We would also like to expand to other platforms beyond Twitter and Facebook. These could include Google search results and YouTube search results and recommendations. If you have data on another platform, or even just a suggestion of how we might collect it, we’d be happy to hear from you.

Conclusion

Platform Health Metrics are a way to provide constructive accountability to the media platforms at a meaningful scale. The current environment, based on identifying and reporting individual bad outcomes, is less constructive form of accountability for the platforms because they can never show that they are doing well, only make it harder for watchdogs to catch mistakes.

By contrast, the Iffy Quotient tracks trends over time rather than reporting on individual problems. It provides quantitative evidence in support of the claim that Facebook and Twitter did a poor job during the 2016 election season; they amplified the distribution of information from Iffy sites at double the rate that they did earlier that year. But the Iffy Quotient can also tell a more positive story of progress when progress is made, as happened in late 2017 and 2018, especially on Facebook.

Acknowledgements

We gratefully acknowledge the NewsWhip service and the work of Media Bias/Fact Check and Open Sources to assess news sites and share the assessments publicly.

The Iffy Quotient is an adaptation of Aviv Ovadya's previous work toward a dashboard for measuring attention toward unreliable sources, which he developed prior to joining the Center for Social Media Responsibility. Ovadya's work began in late 2016 and was also funded through his 2017 Knight News Innovation Fellowship at the Tow Center for Digital Journalism at Columbia University. The Iffy Quotient was also inspired in part by Craig Silverman's analysis of engagements for the top stories from hoax and mainstream news sites.[20]

Vitaliy Lyapota was the primary software developer for Iffy Quotient. Yuncheng Shen created the images and animations describing the steps in the calculation of the Iffy Quotient. Fernand Pajot and Ashwin Rajadesingan conducted data analyses that contributed to quality assurance.

[20] https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook