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)
v2 (this version): July 23, 2019 (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.

In 2019, we changed the methodology for calculating the Iffy Quotient. We now deem a site as Iffy or OK based on ratings from NewsGuard, falling back on rating from Media Bias/Fact Check only for sites not rated by NewsGuard. In addition, because NewsGuard has higher coverage of sites that were popular in 2019 than those in 2016, and for sites that were popular on Facebook than those that were popular on Twitter, we now define the Iffy Quotient as the fraction of URLs from rated sites rather than the fraction of URLs from all sites. This has increased the absolute values of the Iffy Quotient on particular days in comparison to our version 1 numbers, but the trends over time remain largely unchanged.

---

[1] Paul Resnick is a Professor and Associate Dean for Research at the University of Michigan School of Information, and Director of the Center for Social Media Responsibility. 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.
[3] Garlin Gilchrist helped to draft the initial version of this report while he was Executive Director of the Center for Social Media Responsibility. He currently serves as Lieutenant Governor of the State 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 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 to 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 argues: "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]

---

[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

Crowell's post, on the other hand, goes on to suggest that Twitter thinks its only responsibility is to ensure the spread of counter-information to misinformation.

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 seemingly factual claims are in fact contextual spin that may invite misinterpretation without actually making any factual claims. 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, NewsGuard[9] and Media Bias/Fact Check.[10] They make judgments not on individual items but on entire sites. We treat their judgments as a proxy for whether a particular content site frequently publishes information that the majority of people would consider false or misleading, after considering 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 judgments 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 NewsGuard and Media Bias/Fact Check, 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 NewsGuard or Media Bias/Fact Check.

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 then download site judgments from NewsGuard and Media Bias/Fact Check about which sites are Iffy. Finally, we calculate the

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

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.

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.

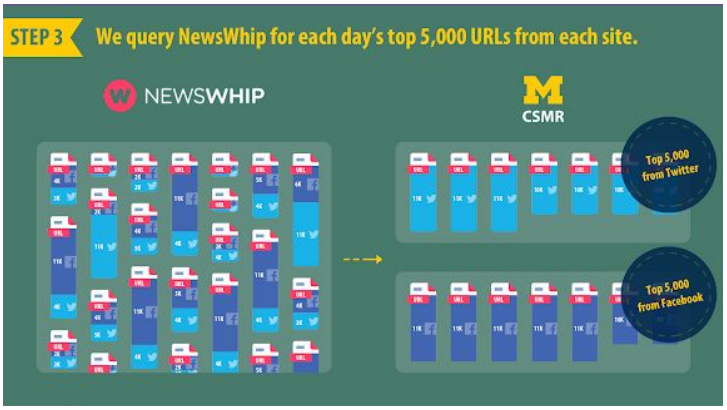
NewsWhip tracks new URLs added to the sites it monitors. Whenever a new page is added to one of these sites, NewsWhip checks for social engagement with that URL on Facebook and Twitter as described below. NewsWhip also continues to check social engagement repeatedly, though more and more rarely as the 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 27, 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.



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. We continue to update the Iffy Quotient for a date for 90 days, and then freeze it to provide an archival score for that date.

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.