ESTABLISHING ACCOUNTABILITY TO ACCOUNTS ON SOCIAL MEDIA AND ONLINE ADVERTISING PLATFORMS

Liang-Yun Wang
ESTABLISHING ACCOUNTABILITY TO ACCOUNTS ON SOCIAL MEDIA AND ONLINE ADVERTISING PLATFORMS

Social media platforms and online advertising platforms may facilitate the sharing of ideas and information between users. For example, a user may create an account on a social media platform and share content items, such as posts, photographs, news articles, etc., with other users on the social media platform. Unlike traditional media (e.g., newspapers, magazines, ratio, TV, etc.), social media and online advertising platforms are able to target very narrow audiences at a low cost. The ability to target specific audiences with content items at a relatively low cost has made social media and online advertising platforms more attractive to some advertisement and content providers than traditional media.

However, due to the large amount of users and information being shared via a social media or online advertising platform, it is difficult to identify the activities of a particular account. As a result, accounts may be created on social media platforms for the purpose of sharing false content items, such as fake news, political propaganda, false advertisements, etc., with targeted groups of users on the social media platform. This can hurt the credibility of authorized advertisement and content providers on social media platforms since it is difficult for a user to know whether a shared item of content is real or fake. This can also decrease user engagement with the social media platforms, as users of the platforms may become frustrated with large amounts of false content items being posted to users’ news feeds. False content items are used herein to refer to information provided with the intent to mislead in order to damage an agency, entity, or person, and/or gain financially or politically.

We propose a mechanism that uses a reach index to identify accounts of interest on a social media or online advertising platform. The reach index for an account can be defined by a
number of users that have consumed a content item originated from the account over a particular
period of time. For example, if an account posts an article on a social media platform, the reach
index can be the number of users who read the article during a 24-hour period of time.

When the reach index for an account exceeds a threshold value, the social media or
online advertising platform can limit the exposure of the account to other users of that social
media or online advertising platform. The exposure of the account to other users of a social
media platform can be limited by no longer posting content items originating from the account to
the news feeds of the other users. For example, if the reach index of an account exceeds a
threshold, a social media platform may no longer post content originating from the account to
users’ news feeds. The exposure of the account to other users of a social media platform can also
be limited by lowering the priority of content items originating from the account in search
results. For example, if a user searches for a particular type of content that matches the type of
content shared by the account, the content shared by the account may not be displayed as
prominently or may be displayed further down in the search results, or may not be displayed
within the search results.

After limiting the exposure of the account, the social media or online advertising platform
can transmit a request for verification to the account. As will be discussed in more detail below,
the verification may refer to a response confirming that the representation of the account and/or
the account holder on the social media or online advertising platform is accurate (i.e., not false).
Upon receiving the verification from the account, the social media or online advertising platform
can remove the limitations placed on the account.

Therefore, in order to reach a larger target audience, an account will need to first be
verified. This verification adds a level of accountability to accounts on social media or online
advertising platforms in that accounts can no longer be created anonymously to share false content items with a large amount of users. Furthermore, by limiting the exposure of an account to other users when a reach index exceeds a threshold, the number of users exposed to content from a potentially fraudulent account is reduced.

Figure 1 depicts a flow diagram of a method to determine whether to limit exposure of an account on a social media platform based on a reach index of a content item originating from the account. One or more operations of Figure 1 may be performed by a social media or online advertising platform. First, at step 101, an account of a social media platform may be identified.

Next, at step 102, a reach index may be determined for the account identified at step 101. As previously described, the reach index can correspond to the number of users that have consumed a content item originated from the account over a particular period of time. The period of time can be based on the type of content posted. For example, if the content item is a sports video, the period of time can be a day, whereas if the content item is a news article, the period of time can be a week.

Subsequently, at step 103, a determination as to whether the reach index for the account exceeds a threshold may be made. The threshold value can be a predefined value for any content item shared by any unverified account on the social media platform. For example, the threshold value can be set at 1000 users that have consumed a content item over a period of time. Alternatively, the threshold value can be based on the type of content item being shared. For example, if the shared content item is a sports video, the threshold value can be set at 5,000 users, whereas if the content item is a news article, the threshold value can be set at 1,000 users.

At step 104, in response to determining that the reach index for the account exceeds the threshold, the exposure of the account to users of the social media platform is limited. Limiting
the exposure of the account may prevent content items originating from the account from being posted to users’ news feeds. In addition or alternatively, the exposure of the account to other users of a social media platform can be limited by lowering the priority of content items originating from the account in search results. To remove the limitations of exposure that are placed on the account, a verification process may need to be completed by the user or organization associated with the account.

In embodiments, although the exposure of the account on a social media platform can be limited, content shared by the account can still be accessed via a direct link (e.g., hyperlink). For example, a first user of the social media platform can send a link to a content item originating from the account to a second user of the social media platform. However, upon selection of the link by the second user, a notification can be provided to the second user indicating that the account has not been verified.

Furthermore, at step 105, a request for verification may be transmitted to the account on the social media platform. The request may be a notification provided to the account via the social media platform, an email sent to an email address associated with the account, a text message provided to a phone number associated with the account, etc. The notification may prompt a user or organization associated with the account to verify the authenticity and ownership of the account. For example, if an account claims to be affiliated with a political party, a request may be transmitted for the account to verify the account’s affiliation with the political party.

Various methods may be used to verify an account. One method may be requesting and receiving a scanned copy of a government identification of the user of the account. Another method may be charging a small amount of money to a credit card associated with the account,
and then subsequently refunding the amount of money. Another method may be requesting an organization to provide an official business document of the organization associated with the account. It should be noted that any other methods of verifying an account on a social media or online advertising platform may be used. In some embodiments, a request to make the ownership information for the account publicly available may also be transmitted to the account along with the request for verification.

At step 106, upon receiving the verification (and confirming that the verification is valid), the limitations on the exposure of the account to users of the social media platform may be removed. Once the limitations are removed, content items originating from the account may be posted on users’ news feeds and promoted in search results. Since the ownership of the account has been verified, the owner of an account may be held accountable for any content or activity associated with the account. Thus, the posting of fake news, propaganda, false advertising, etc. by anonymous parties may be reduced. Furthermore, in embodiments, the ownership information associated with the account may be made publicly available. This may enable third-parties, such as government agencies, to monitor, identify and hold accountable the owners of the accounts on social media or online advertising platforms that may be providing false content items.

As discussed above, a disseminator of misinformation can create a single account on a social media or online advertising platform to provide false content items to a large number of users. Alternatively, many accounts may be created to provide false content items to a smaller number of users. For example, rather than using a single account to provide false content items to 1000 users, ten accounts may be used that can each provide false content items to 100 users. In such a scenario, utilizing a reach index threshold may not prevent the false content items from
being provided to a larger number of users, since the reach index for each of the accounts may be relatively small.

Therefore, to prevent the above scenario of multiple accounts, content matching may be performed to identify the multiple accounts being used to provide the same false content items. The content matching may identify that a same or similar content item is being provided by multiple accounts on a social media platform during a period of time. For example, the content matching may identify a particular news article that has been provided on a social media platform by 10 different accounts over a period of time of one day. Upon identifying the content item provided by multiple accounts, the social media platform may aggregate the reach index for each of the accounts. For example, if each of the 10 different accounts has a corresponding reach index of 100, then the aggregate reach index for the 10 accounts is 1,000. The social media platform can then determine whether the aggregate reach index exceeds a threshold. If the aggregate reach index exceeds a threshold, then the exposure of each of the accounts (e.g., the 10 different accounts) may be limited.
ABSTRACT

A mechanism that uses a reach index to identify accounts of interest on a social media or online advertising platform is described. In particular, an account of the social media platform may be identified. A reach index for the account may be determined. A determination as to whether the reach index for the account exceeds a threshold may be performed. In response to determining that the reach index for the account exceeds the threshold, the exposure of the account to users of the social media platform may be limited. A request for verification may be transmitted to the account on the social media platform. Upon receiving the verification, the limitation on the exposure of the account to the users of the social media platform may be removed.

**Keywords:** social media, online advertising, verification, fake news, propaganda, accountability
Identify an account of a social media platform

Determine a reach index for the account

Determine whether the reach index for the account exceeds a threshold

In response to determining that the reach index for the account exceeds the threshold, limit the exposure of the account to users of the social media platform

Transmit a request for verification to the account on the social media platform

Upon receiving the verification, removing the limitation on the exposure of the account to the users of the social media platform

FIG. 1