Reverse Engineering of Content Delivery Algorithms for Social Media Systems

Ittipon Rassameeroj, S. Felix Wu
Department of Computer Science
University of California
Davis Davis, CA, USA
Email: {itras,sfwu}@ucdavis.edu

***Abstract*—Social media systems have become a primary platform to consume and exchange information nowadays. As a black box, social algorithms were designed and trained to pick up, filter, and rank the most relevant and desired content to be delivered to each individual one of us. However, these modern social algorithms were typically complicated, and its development involved easily more than one hundred software engineers. Therefore, we, including the social media providers, do not really know how these social algorithms work such that we are unsure about the quality of the delivered content or the existence of any bias or disparity? In this paper, we explore content delivery algorithms on Facebook fan pages and communities. We empirically defined four hypotheses, which originally were from Facebook features, social network concepts, and what we have observed from our SINCERE data set. We took a heuristic approach to statistically analyze our data set to uncover, probabilistically, how Facebook push user-created contents among users. Each hypothesis is represented by an explainable logical expression rule. We used each of the top 100 posts with the largest number of comments in ABC News, CNN, and Fox News fan pages. Our main contribution is to validate each hypothesis against empirical data such that together we provided a partial explanation for the content delivery algorithm used by Facebook.**

***Index Terms*—content delivery algorithms, social algorithms, news media, reverse engineering, Facebook fan pages**

I. Introduction

We nowadays have a huge amount of online contents on the Internet. In the past, we mostly used online search engines as an interface to search and bring us to desired contents. However, our information consumption behavior has been changed. In addition to search engines, many people have currently used social media to consume information. Social media systems normally provide us with online socialization, information dissemination and consumption, and content creation. In particular, we have enabled social media to push us content instead of searching and going to information sources.

Social media systems have some mechanisms to select and rank a bunch of different contents for each user based on personalization, which we call *social algorithms*. Basic techniques of social algorithms are based on users social network data (such as relationship), personalization, and online activities. Unfortunately, we do not really know how social algorithms actually work, which is like a black box. Also, we do not have any formal or standard methodology for content delivery algorithms in social media systems, and there is no existing work about that so far, especially for Facebook fan pages and communities. Thus, it is really challenging to evaluate whether contents filtered from the systems are the most desired and suitable for us.

Misinformation in social media is one of good application examples that really needs a good model of content delivery algorithms. Since a lot of misinformation, such as fake news, is disseminated in online social network, basic interesting research questions are how does social media systems deliver misinformation contents to users? In addition, how many users are infected or trapped from those contents? It is very challenging due to lack of content delivery models to estimate feasibly infected users from those contents. Besides, if we have a good content delivery model, we will be able to make use of it to understand or analyze strategies of misinformation distributors, and to learn how they take advantages of social algorithms to spread misinformation contents to get attention from users.

In this work, we do reverse-engineering of user-created content delivery algorithms on Facebook fan pages and communities. We focus on only user comments as user-created content. Our research question is how does Facebook push the contents among users who interact with a post and among them? We heuristically define hypotheses to do reverse engineering to answer the research questions. Our hypotheses empirically are from what we have noticed from Facebook features, social network theories, and what we have observed from our data set. We also present our hypotheses by partial explanations of logical expression. We use top 100 posts that contained the largest number of comments from each of three news media Facebook fan pages: ABC News, CNN, and Fox News. We present our algorithms to analyze each of total 300 posts independently to evaluate and validate all hypotheses.

The remainder of this paper is organized as follows. We further motivate our study with additional related work and backgrounds in Section II. In Section III, we present some related technical backgrounds, scope, definitions, and an overview of this work. Also, we introduce our SINCERE data and the data set we used in this paper. In Section IV, we present our hypotheses, definition and explanation of logical expression for each hypothesis, measurements for

978-1-7281-2946-4/19/$31.00 ©2019 IEEE

each hypothesis, algorithms that we analyze our data set to evaluate the hypotheses, and experiment results. We discuss our limitations for this work and our plan for the future work in Section V. Finally, we summarize all of this paper in Section VI.

II. Related Work

As we investigate social algorithms, let us start with social algorithm concepts. Lazer [2] has presented some ideas, overview, and brief definition of social algorithms. The author has also given some examples to show how social algorithm has an impact on choosing and ranking all content on the Internet for users. On the other hand, Yang [3] has presented social algorithms in terms of computational algorithms, which is a special class of algorithms for solving optimization problems. The author has shown the algorithms that does not focus on only social media systems, but they could be applied for general context, mainly nature-inspired, population-based algorithm for optimization.

Social media systems have become one of the main channels to diffuse information rapidly. Basic techniques that have been used for their content delivery algorithms include personalization, interests, and online activities. In addition, the strength of weak ties theory [4] is the most popular techniques applied to content delivery algorithm in most of social media services. Gilbert and Karahalios [5] have presented a predictive model that maps social media data to tie strength in online social networks. For example, contents that are created from friends whom we frequently interact with should be picked and shown first. Bakshy et al [6] have presented how strong and weak ties are influent in terms of information propagation in social networks.

We receive contents filtered by social algorithms as a black box, but how do we know whether a received content is the most desired we want to get? There are existing researches investigating bias in social algorithms. Pariser has presented *filter bubble*, which personalization makes people see different viewpoint of content [1]. [7] is the first paper to measure the filter bubble effect by exploring content diversity of a filter-based recommender system, MovieLens. In addition, since we all do not really know how social algorithms work, there are some concerns about bias of algorithmically selected content. Notwithstanding, Bozdag [8] has presented another side of those processes. The author has shown that online services that filter information are not only algorithms, but humans can also manually influence the filtering process. Furthermore, ideological diverse issue is another concern. Flaxman et al [9] have pointed out that both search engines and social media may use ideological distance as a factor to pick a content up and push to users. Also, these tools are associated with an increase in the mean ideological distance between individuals. Bakshy et al [10] have studied impact of social algorithms on Facebook by exploring whether personalized news feed on Facebook hide some content to prevent users from accessing posts contained conflicting political views. Flaxman et al have explored an impact of polarizing channels on news consumption for ideological segregation in [11]. Bernstein et al [12] have done reverse-engineer to estimate invisible audiences on Facebook. What they have done is to survey active users to ask them to estimate their audience size, then they have compared their estimates to their actual audience size from server logs. However, apart from News Feed algorithm, no one explores social algorithms of content in Facebook pages or communities.

III. Methodology

A. Background

Social media systems have some mechanisms to rank and choose contents for each user. However, we do not really know how they do, and even social media providers probably do not know that because they may use some machine learning techniques as a black box to deal with that. Facebook News Feed algorithm is quite straightforward because contents are ranked based on user relationship links, such as close friends and interaction frequency. However, Facebook fan pages or communities is different because all users can interact (like, react, or reply) with any contents regardless of relationship as long as they are members in the same communities or the same pages. Thus, algorithms for fan pages are also different from News Feed. It is very challenging because we have no clue how Facebook chooses, ranks, and pushes user-created contents (such as user comments) among users without any relationship links. Moreover, there is no existing work studying about social algorithms in communities or Facebook fan pages.

B. Conceptual approach

The user-created contents we focus on this work are user comments in posts in Facebook fan pages and communities. We explore how Facebook delivers those contents among users who participate (like, react, and comment) in a post or interact with (like, react, or reply) other user comments. We define four hypotheses to do reverse engineering for content delivery algorithms in Facebook fan pages. Our hypotheses empirically are from what we have noticed in Facebook features, some social network concepts, Facebook News Feed algorithms, and what we have observed in our data set. We also present each hypothesis as explainable logical expression rule. For each hypothesis, we statistically analyze each Facebook post in each fan page independently from the data set by our proposed algorithms to understand how content delivery algorithms work.

In this paper, what we explore content delivery algorithms in Facebook fan pages is basically to understand patterns of which users see which comments based on their interactions. We use like, reaction, and reply as indicators of which contents are seen by which users. Also, in this work, ”interaction” means like/react/comment for a post and like/react/reply for a comment. When we know that which users see which contents, we assume Facebook pushes those contents to those users. Therefore, we try to come up with heuristic approaches to show why and how Facebook chooses specific contents to particular users.

TABLE I
THE TOTAL NUMBER OF THE DATA SET

Data	The number of data
Posts	300
Comments	13,031,988
Likes	29,902,546
Reactions	23,391,556
Replies	1,380,399

C. Data

We use our social interactive networking and conversation entropy ranking engine (SINCERE) data [13]. It has been crawled via Facebook API from the public interaction data on many Facebook fan pages and public communities from 2008 to early 2018. The SINCERE data is composed of almost all posts, comments, likes, reactions (love, haha, wow, anger, sad), and shares. In this work, as we focus on user comments, from the SINCERE data, we choose top 100 posts having the largest comments in each of three news media Facebook fan pages–ABC News, CNN, and Fox News as the data set, to do experiment, evaluate, and prove our hypotheses. From total 300 posts in the data set, the amount of comments in each post in range of 4,991 - 330,605 comments in ABC News page, 7,997 - 70,213 comments in CNN page, and 56,278 - 412,621 comments in Fox News. In addition, Table I shows the total amount of data set we used in this work.

IV. HYPOTHESES AND EXPERIMENTS

In this section, we describe each of our four hypotheses represented by logical expression explanation. In addition, we present measurements and our algorithms to analyze the data set to evaluate and validate each hypothesis. Finally, we show results from running the algorithms on the data set.

A. Hypothesis 1.

This hypothesis is that when a user just creates a comment in a post, other users who have earlier created comments in this post will see this comment. Let us represent this hypothesis by a formal definition or a rule. Suppose comment c is created by user x at time t in post p on page pg. The logical expression explanation of a content delivery rule for the hypothesis is:
$((\exists\ comment\ pg.p.c) \wedge (author(pg.p.c) = x) \wedge (time(pg.p.c) = t)) \wedge ((\exists\ comment\ pg.p.\overline{c}) \wedge \neg(author(pg.p.\overline{c} = x)) \wedge (time(pg.p.\overline{c}) < t))$

Initially, this hypothesis is from what we have noticed Facebook notification system, which is when we create a comment in a post, then another user does that too later, Facebook will send us a notification that someone also made the comment in the post. Thus, we believe that the notification system may essentially encourage users who make comment earlier to see further comments if it is still in their online time. In addition, in terms of psychology, when people create a content or leave a comment in a post, they like to see feedback [14].

Algorithm 1 explains our evaluate process on the data set in each post. We have two measurements for this experiment:

(i) The number of comments that have interactions (like/ react/reply) from users who have earlier created comment(s) in the same post.
(ii) From all interactions in each comment in (i), how many of them are from users who created comments earlier.

Algorithm 1 Evaluation for hypothesis 1

$C \leftarrow$ {all comments having reactions or replies}
for each comment $c_i \in C$ **do**
  $t_{c_i} \leftarrow$ timestamp of comment c_i
  $l_{c_i} \leftarrow$ {all users who react or reply in comment c_i}
  $Cb \leftarrow$ {all comments that their timestamp $< t_{c_i}$}
  for each comment $c_j \in Cb$ **do**
    $u_{c_j} \leftarrow$ an author of comment c_j
    if $u_{c_j} \in l_{c_i}$ **then**
      $num_{overlap} \leftarrow num_{overlap} + 1$
    end if
  end for
  if $num_{overlap} > 0$ **then**
    $num_{comment_{overlap}} \leftarrow num_{comment_{overlap}} + 1$
    $total_{overlap} \leftarrow total_{overlap} + \frac{num_{overlap} * 100}{|L_{C_i}|}$
    $num_{overlap} \leftarrow 0$
    $n \leftarrow n + 1$ ▷ # comments with reactions from users who created comment before
  end if
  $m \leftarrow m + 1$ ▷ # comments having like/react/reply
end for
$percent_{num_{comment}} \leftarrow \frac{num_{comment_{overlap}} * 100}{m}$
$avg_{overlap} \leftarrow \frac{total_{overlap}}{n}$

We run Algorithm 1 with each of 300 posts in three fan pages independently to find percentages of the first (i) and second (ii) measurements, which are in the last two line in the Algorithm. The result is shown in Figure 1. Each dot on boxes represents each of 100 posts for each page, and each box represents each fan page: ABC News, CNN, and Fox News. The three left boxes (dark blue, orange, and light blue) are the result of the first measurement (i) while the right three boxes (gray, yellow, and green) are the result the second measurement (ii) for ABC News, CNN, and Fox News pages respectively.

In Figure 1, from 100 posts in each page on the data set, percentage averages of comments having interactions from users who earlier created comments (measurement i) are 39.22% for ABC News, 48.83% for CNN, and 68.91% for Fox News. In addition, in each of those comments, average percentage of interactions from users who earlier created comments (measurement ii) are 31.06% in ABC News, 34.86% in CNN, and 54.72% in Fox News. For ABC News (dark blue and gray boxes), it is obvious that there are two groups of posts: high and low percentages. Almost all of high percentages are live video posts while others are other kinds of post. Also, Fox News has much higher percentages than others. We figure out

that top 100 posts from Fox News are live video posts mostly (73 of 100). We strongly believe that users usually spend more time in a live video post, so they will see streaming of other comments as a real-time. That is why live video posts have very high percentage for this hypothesis.

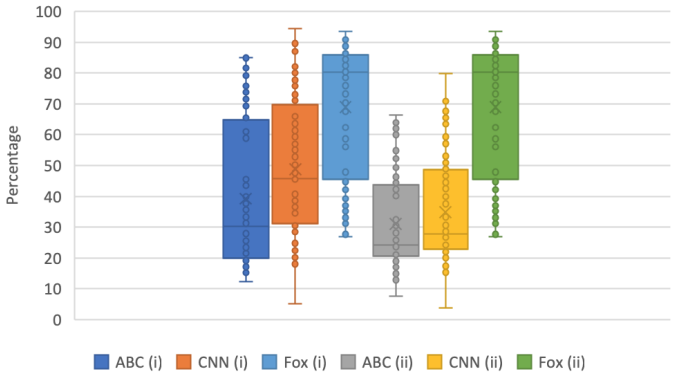

Fig. 1. Percentage of comments that have interactions from users who created comments earlier (the left three boxes) and percentage of interactions from those users in each of those comments (right three boxes) for top 100 posts in ABC News, CNN, and Fox News pages.

Another interesting question for this hypothesis is, when a comment c is just created in a post at time t, how many other users who have earlier created comments in the post will see the comment c in terms of time difference? For example, if comment c gets most reactions from users who have created comments in last five minutes, which their comments have been created from time $t-5$ to t, we can assume the comment c is mostly seen by users who have created comments in last five minutes. We analyze all 100 posts in each page to find average time duration windows, and we have five time duration windows: 1 - 10 minutes, 11 - 20 minutes, 21 - 30 minutes, 31 minutes - 1 hour, and more than 1 hour. We present the result in Figure 2 as average percentage of 100 posts for ABC News, CNN, and Fox News. From the result, it is very obvious that most users who create comments will participate with other users' comment that will be created in 10 minutes. In other words, we can assume Facebook may push the most recent contents that are just created to users who have created comments in last 10 minutes. In other words, most users will be online for 10 minutes after they create a comment.

B. Hypothesis 2.

Suppose user x and y like or react post p on page pg, and reaction r of post p is $Pg.p.r$ contained x and y. If x creates comment c in post p, then comment c is supposed to be delivered to y. The explanation of the logical expression rule for this hypothesis is:
$((\exists\ reaction\ \ pg.p.r) \wedge (author(pg.p.r) = x) \wedge author(pg.p.r) = y)) \wedge ((\exists\ comment\ \ pg.p.c) \wedge author(pg.p.c) = x))$

This hypothesis is from a concept of social network recommendation systems, which is information will be

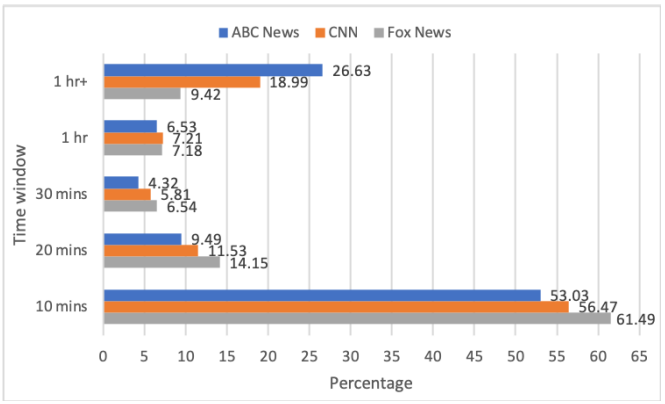

Fig. 2. Time duration windows between the time users who created comments earlier and the time they interacted with other comments later for each page.

diffused among users based on their common interests. Our basic idea is that if two users click like/react the same content (a post), we assume that they have a common interest. Then, if one of them (x) creates a content (a comment in the post), another user (y) may be interested in x's comment. Thus, our hypothesis is that Facebook may push x's content to y.

Essentially, our data analysis approach firstly is to get all users who liked/reacted a post and created comment(s) in the post from our data set first. Then, from those users, we find two measurements for this hypothesis:

(i) Percentage of users who got interactions on their comments from other users who also reacted the post. In other words, we can see the proportion of how many comments, created by users who reacted the post, are seen by other users who also reacted the post.

(ii) For each user in (i), percentage of interactions on his/her comments from other users who reacted the post. Finally, we find average percentage of that for all users.

Algorithm 2 explains all steps to find two measurements for each post on the data set.

Figure 3 shows results from running Algorithm 2 for each of 100 posts in each page. The first three boxes from the left side (dark blue, orange, and gray) are the first measurement (i), and the three right boxes (yellow, light blue, and green) are the second measurement (ii) for ABC News, CNN, and Fox News respectively. We can see both ABC News and CNN are very similar for both matrices while Fox News is much higher than others because Fox News has many live video posts in the data set. For the first measurement (i), the average percentage is 42.12% for ABC News, 43.69% for CNN, and 68.45% for Fox News. For the second measurement (ii), the average of percentage is 31.06% in ABC News, 30.04% in CNN, and 50.23% in Fox News.

C. Hypothesis 3.

Suppose user x creates comment c in post p on page pg, and user y reacts or replies to comment c. After that, if user y creates comment $\bar{c}$, then comment $\bar{c}$ will be delivered to user x. Thus, logical expression explanation of the hypothesis

Algorithm 2 Evaluation for hypothesis 2

$L_P \leftarrow$ {all users who like or react a post}
$U \leftarrow$ {all users who react and comment in the post}
for each user $u \in U$ **do**
  $L_u \leftarrow$ {all users who react any comments of u}
  if $|L_u| > 0$ **then**
    $intersect_{mutuallike} \leftarrow |L_p \cap L_u|$
    $percent_{overlap} \leftarrow \frac{intersect_{mutuallike} * 100}{|L_u|}$
    $tot_{mutuallike} \leftarrow tot_{mutuallike} + percent_{overlap}$
    $n \leftarrow n + 1$ ▷ # users having reacts in their comment
    if $intersect_{mutuallike} > 0$ **then**
      $m \leftarrow m + 1$ ▷ # users having likes from users liking the post
    end if
  end if
end for
$percent_{mutuallike} \leftarrow \frac{m*100}{|U|}$ ▷ the 1^{st} measurement (i)
$avg_{overlap} \leftarrow \frac{num_{mutuallike}}{n}$ ▷ the 2^{nd} measurement (ii)

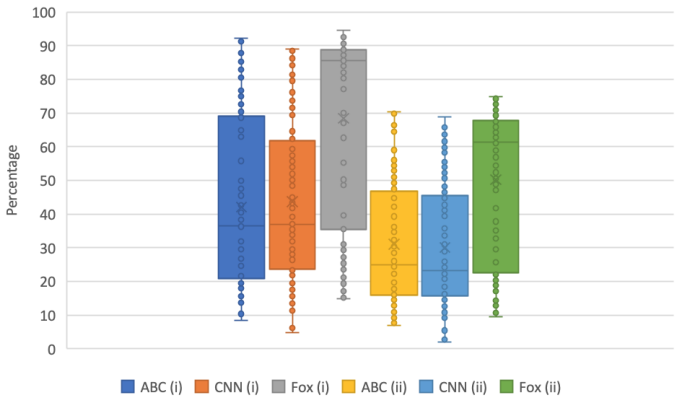

Fig. 3. Percentage of users who liked/reacted a post and like/react/reply a comment created by other users who also liked/reacted the post in ABC News, CNN, and Fox News pages.

can be represented by:
$((\exists\ comment\ pg.p.c) \wedge (author(pg.p.c = x)) \wedge (((\exists\ reaction\ pg.p.c.r) \wedge (author(pg.p.c.r = y)) \vee ((\exists\ reply\ pg.p.c.rp) \wedge (author(pg.p.c.rp = y))) \wedge ((\exists\ comment\ pg.p.\bar{c}) \wedge (author(pg.p.\bar{c} = y))$

We call this hypothesis as mutual interaction. We apply an idea of strength tie [4] and Facebook News Feed algorithm. In online social networks, if two users interact with each other frequently, that means they have strength tie [5]. For example, if one of them creates a content, Facebook will usually pushes his content to another user on her news feed first. Therefore, we apply these concepts to our hypothesis based on user interactions in each post on our data set. Our approach basically is to find probability that users interact with each other comments. For example, if user x likes user y's comment, and y likes x's comment, then the probability will be equal to 1. Algorithm 3 explains our approach to find the probability for each post on our data set.

Algorithm 3 Evaluation for hypothesis 3

for each user u who comments in a post and has reactions or replies on his/her comment **do**
  $L_u \leftarrow$ {all users who react/reply all comments of u}
  for each user $v \in L_u$ **do**
    $C_v \leftarrow$ the number of all comments made by v
    if $C_v > 0$ **then**
      $num_{liker} \leftarrow num_{liker} + 1$
      $L_v \leftarrow$ {users who react all comments of v}
      if $u \in L_v$ **then**
        $num_{mutuallike} \leftarrow num_{mutuallike} + 1$
      end if
    end if
  end for
  if $num_{liker} > 0$ **then**
    $sum_{prob} \leftarrow sum_{prob} + \frac{num_{mutuallike}}{num_{liker}}$
    $n \leftarrow n + 1$
    $num_{liker} \leftarrow 0$
    $num_{mutuallike} \leftarrow 0$
  end if
end for
$avg_{prob} \leftarrow \frac{sum_{prob}}{n}$

Figure 4 - 6 present results from running Algorithm 3 on the data set, which is probabilities that users interact with each other comments for each 300 posts in ABC News, CNN, and Fox News. In each figure, probability is represented on y-axis for each of 100 posts independently (on x-axis) that ranked by the largest number of comments. On x-axis, the leftmost (post 1) means the first largest number of comments while the rightmost (post 100th) means the 100th largest. All posts that have very low probability are live video posts. On our data set, Fox News has 73 live video posts of 100 posts while ABC News and CNN have 28 and 24 live video posts respectively. That is why there are many posts having very low probabilities in Figure 10 because of many live video posts in top 100 posts in Fox News page.

We also present probability comparison of all 100 posts in each page in Figure 7. The leftmost box (blue) is from all probabilities of 100 posts in Figure 4 for ABC News, the middle box (orange) is from the all 100 posts in Figure 5 for CNN, and the rightmost box (gray) is from the all 100 posts in Figure 6 for Fox News.

Figure 7 also shows comparison of average probability for each page. Average probability is 0.43 for ABC News, 0.26 for CNN, and 0.08 for Fox News. We figure out live video posts have low probabilities. Fox News has very low average probability due to many live video posts from the data set.

D. Hypothesis 4.

Suppose user x creates comment c_1 in post p on page pg at time t, and user y reacts or replies to comment c_1 at time $t + m$. After that, user x creates another comment c_2 at time

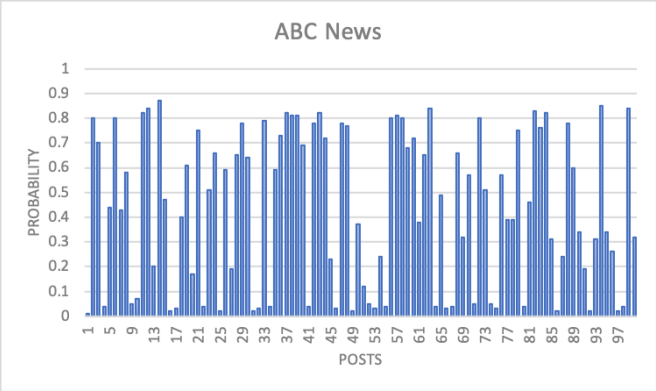

Fig. 4. Probabilities of users mutual interaction for top 100 posts in ABC News page

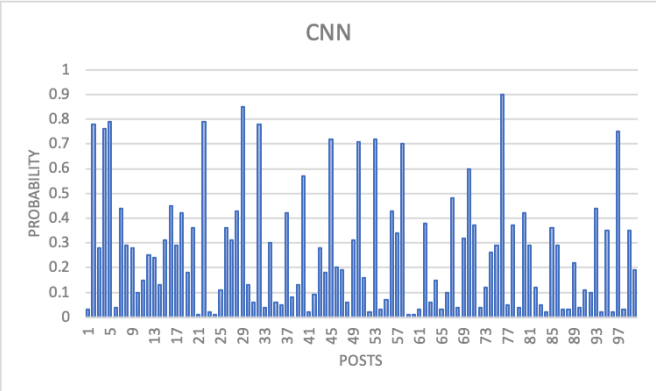

Fig. 5. Probabilities of users mutual interaction for top 100 posts in CNN page

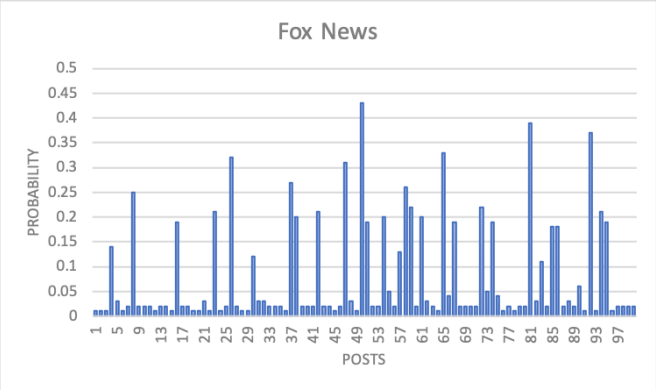

Fig. 6. Probabilities of users mutual interaction for top 100 posts in Fox News page

$t + m + n$, where m and n > 0, so comment c_2 should be delivered to user y. We represent an explanation rule by:
$((\exists\ comment\ \ pg.p.c_1)\ \wedge\ (author(pg.p.c_1)\ =\ x)\ \wedge\ (time(pg.p.c_1)\ =\ t))\ \wedge\ (((\exists\ reaction\ \ pg.p.c_1.r)\ \wedge\ (author(pg.p.c_1.r)\ =\ y)\ \wedge\ (time(pg.p.c_1.r)\ =\ t\ +\ m))\ \vee\ (((\exists\ reply\ \ pg.p.c_1.rp)\ \wedge\ (author(pg.p.c_1.rp)\ =\ y)\ \wedge\ (time(pg.p.c_1.rp)\ =\ t + m)))\ \wedge\ ((\exists\ comment\ \ pg.p.c_2)\ \wedge\ (author(pg.p.c_2)\ =\ x)\ \wedge\ (time(pg.p.c_2)\ =\ t + m + n))$

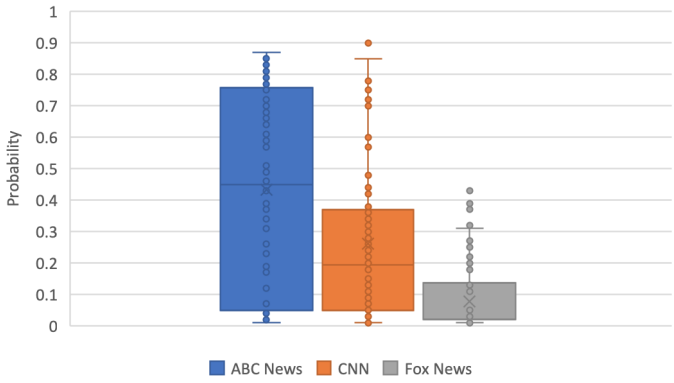

Fig. 7. Probability comparison of mutual user interactions for top 100 posts on ABC News, CNN, and Fox News pages

We are interested in this hypothesis because of misinformation on Facebook. In our SINCERE data, we have found some comments contained fake news or malicious contents could draw a lot of attentions. One possibility is that a user may create a good content to try to draw attention from many participants first. After that, he may post a malicious content. If our hypothesis is true, Facebook will probably deliver that malicious content to most users who participate in his first content. If so, attackers may use this to be one of their strategies to disseminate malicious contents to users.

The following procedure explains our conceptual approach to analyze all comments in each of total 300 posts in each page to validate our hypothesis:

1) Get all users who made more than one comment in a post. Let v be the number of those users.
2) For each user u in step 1, find all users who like/react/reply all u's comments.

$$L_u = \bigcup_{i=1}^{m} l_{u_i} \tag{1}$$

where l_{u_i} is users who like/react/reply comment i created by user u, and m is the number of u's comments that have like/react/reply.

3) For each of u's comment i that has like/react/reply, we find an intersection between users who like/react/reply this comment i and the union of all users from (1), then divided by (1). Repeat this step for all comments of user u to find an average for that.

$$U_k = \frac{\sum_{i=1}^{m}(\frac{|l_{u_i} \cap L_u|}{|L_u|})}{m} \tag{2}$$

4) For each comment j created by other users ($\overline{u}$), we find an intersection between users who like/react/reply comment j and the union of all users from (1), then divided by (1).

Repeat this step for all comments of other users $\overline{u}$ to find an average for that.

$$\overline{U_k} = \frac{\sum_{j=1}^{n}(\frac{|\overline{l_{u_i}} \cap L_u|}{|L_u|})}{n} \quad (3)$$

where n is the number of comments, that have like/react/reply, created by $\overline{u}$.

5) Repeat back to step 2 for the next user until end of users in step 1. Finally, each user u will have U and $\overline{U}$ from step 3 and 4.
6) Find average of U and $\overline{U}$ for all users in step 1. Let x and y be average of all Us and all $\overline{U}$s respectively.

$$x = \frac{\sum_{k=1}^{m} U_k}{v} \quad (4)$$

$$y = \frac{\sum_{k=1}^{n} \overline{U_k}}{v} \quad (5)$$

If the hypothesis is true, in the last step, x will be much larger than y. In our procedure, x is like the probability that users who interact with a comment will interact with another comment created by the same author later while y is opposite. In addition, Algorithm 4 presents the whole process to analyze for all comments in each post in the data set to validate hypothesis 4.

Algorithm 4 Validation of hypothesis 4

for each user u who makes more than one comment in a post **do**
  $L_u \leftarrow$ {all users who react/reply all u's comments}
  for each comments i of user u **do**
    $l_{u_i} \leftarrow$ {all users who react comment i of u}
    if $|lu_i| > 0$ **then**
      $sum_1 \leftarrow sum_1 + \frac{|l_{u_i} \cap L_u|}{|L_u|}$
      $num_1 \leftarrow num_1 + 1$
    end if
  end for
  $avg_1 \leftarrow \frac{sum_1}{num_1}$
  $total_1 \leftarrow total_1 + avg_1$
  $m \leftarrow m + 1$
  for each user v who comments in this post **do**
    for each comment j of user v **do**
      $l_{v_j} \leftarrow$ {all users who react comment j of v}
      if $|l_{v_j}| > 0$ **then**
        $sum_2 \leftarrow sum_2 + \frac{|l_{v_j} \cap L_u|}{|L_u|}$
        $num_2 \leftarrow num_2 + 1$
      end if
    end for
  end for
  $avg_2 \leftarrow \frac{sum_2}{num_2}$
  $total_2 \leftarrow total_2 + avg_2$
  $n \leftarrow n + 1$
end for
$x \leftarrow \frac{total_1}{m}$
$y \leftarrow \frac{total_2}{n}$

Figure 8 - 10 show the results from running Algorithm 4 on the data set for ABC News, CNN, and Fox News pages. X-axis represents each of 100 posts, which the leftmost (post no. 1) is the post that has the largest number of comments and so on, and y-axis is the probability. The blue line is x, and the orange line is y, which is what we have just mentioned in the procedure and the algorithm. It is very obvious the blue line is always larger than the orange line for all cases. Therefore, we conclude that Facebook may deliver a recent comment to users who have interacted with an earlier comment created by the same author.

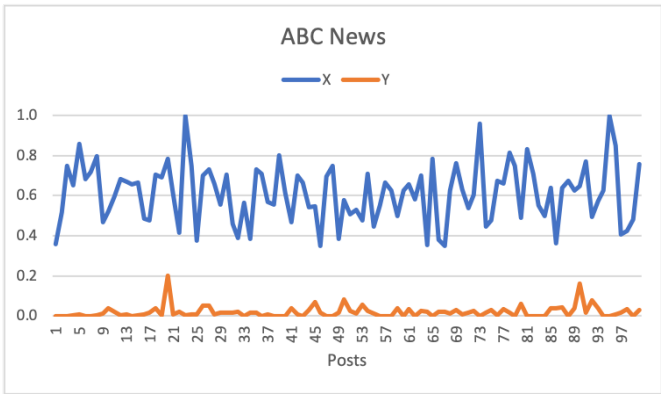

Fig. 8. Validation of Hypothesis 4 for top 100 posts in ABC News page

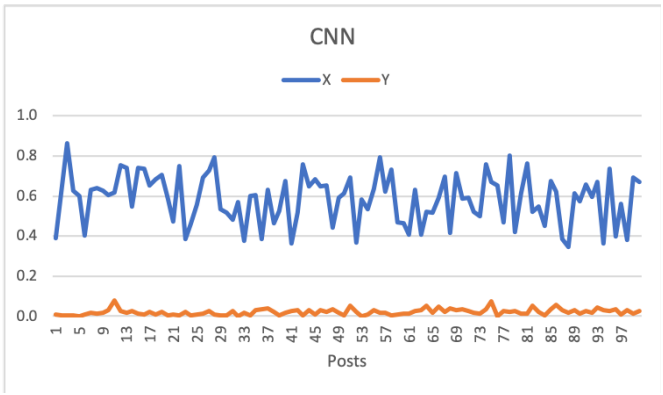

Fig. 9. Validation of Hypothesis 4 for top 100 posts in CNN page

V. Limitations and Future Work

The main contribution for this work is to investigate how Facebook chooses content (comments) to users in fan pages. In other words, we present how which users see which content. We use like, reaction, and reply as indicators of seeing content. These indicators vary for a few reasons. Firstly, many people see contents, but they do not want to interact with other people. Secondly, as like or reaction is emotional interaction, that really varies and depends on people personality. Some people click like a lot while some other do not. Thirdly, some users interact with other's content a lot because they spend more time to keep browsing the content, not only because of the content delivery algorithms.

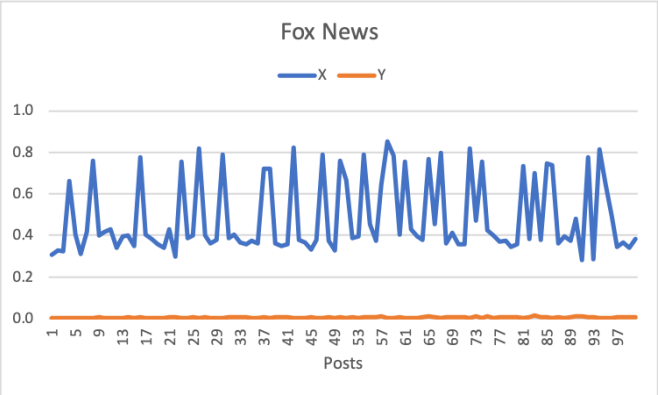

Fig. 10. Validation of Hypothesis 4 for top 100 posts in Fox News page

The data set we have used in this work has the huge number of interactions. There were so many real-time comments around that time, and Facebook had to pick up just only a few comments for each user at a particular time. Hence, they might miss many contents at that time. We think their interactions may be very spread. In addition, the Facebook API we have used to crawl the data does not provide timestamp of likes and reactions. We cannot know exactly when users click like or reaction. That is a reason why time has not been considered in the second hypothesis. Moreover, We have done the experiments for each post in each page independently because, from the API, each user has different Facebook IDs for different posts.

For future work, we will perform a larger scale of experiments by testing our hypotheses with a higher volume of data. Also, several social algorithms related to content types (e.g. live, picture, URL), categories (politics, sport, etc.), date/time (social algorithms should be changed by periods of time), and pages (each page or community may have different algorithms) are significant to be taken into the experiment.

Due to a lot of disseminated misinformation in social media, we are going to explore how Facebook treats those contents and how malicious users or attackers make use of content delivery algorithms to disseminate misinformation contents.

Furthermore, we will improve our existing hypotheses and define more hypotheses to cover most common cases. Ultimately, we will find a way to form all of hypotheses to be a formal method or explaination of content delivery algorithms for Facebook, which is like the explainable AI.

VI. Conclusion

We have done reverse engineering of content delivery algorithms for Facebook fan pages and communities. We have focused on only user comments as the contents in this work. We have empirically defined four hypotheses and use heuristic approach to statistically analyze our data set to see how which users see which comments based on their interaction so that we understand how Facebook delivers user contents among them. We have defined and represented each hypothesis by an explainable logical expression rule. We have used each of the top 100 posts with the largest number of comments in ABC News, CNN, and Fox News fan pages as our data set. We have presented measurements matrices and algorithms to analyze the data set to evaluate and validate for each hypothesis. In addition, we have presented evaluation results from running our proposed data analysis algorithms with the data set. We conclude Facebook deliver contents to users based on their interaction patterns as our four hypotheses. We strongly believe our contribution will be a good model to improve content delivery algorithms in social media systems. Finally, we also believe that our work can bring about enhanced social network applications, such as analysis of fake news dissemination, recommendation systems, and community detection.

References

[1] E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Group , The, 2011.

[2] D. Lazer, "The rise of the social algorithm," *Science*, vol. 348, no. 6239, pp. 1090–1091, 2015. [Online]. Available: http://science.sciencemag.org/content/348/6239/1090

[3] X.-S. Yang, *Social Algorithms*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 1–15. [Online]. Available: https://doi.org/10.1007/978-3-642-27737-5_678-1

[4] M. S. Granovetter, "The strength of weak ties," *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973. [Online]. Available: https://doi.org/10.1086/225469

[5] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '09. New York, NY, USA: ACM, 2009, pp. 211–220. [Online]. Available: http://doi.acm.org/10.1145/1518701.1518736

[6] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in *Proceedings of the 21st International Conference on World Wide Web*, ser. WWW '12. New York, NY, USA: ACM, 2012, pp. 519–528. [Online]. Available: http://doi.acm.org/10.1145/2187836.2187907

[7] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan, "Exploring the filter bubble: The effect of using recommender systems on content diversity," in *Proceedings of the 23rd International Conference on World Wide Web*, ser. WWW '14. New York, NY, USA: ACM, 2014, pp. 677–686. [Online]. Available: http://doi.acm.org/10.1145/2566486.2568012

[8] E. Bozdag, "Bias in algorithmic filtering and personalization," *Ethics and Information Technology*, vol. 15, no. 3, pp. 209–227, Sep 2013. [Online]. Available: https://doi.org/10.1007/s10676-013-9321-6

[9] S. Flaxman, S. Goel, and J. M. Rao, "Filter bubbles, echo chambers, and online news consumption," *Public Opinion Quarterly*, vol. 80, no. S1, pp. 298–320, 2016. [Online]. Available: http://dx.doi.org/10.1093/poq/nfw006

[10] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015. [Online]. Available: http://science.sciencemag.org/content/348/6239/1130

[11] S. R. Flaxman, S. Goel, and J. M. Rao, "Ideological segregation and the effects of social media on news consumption," 2014.

[12] M. S. Bernstein, E. Bakshy, M. Burke, and B. Karrer, "Quantifying the invisible audience in social networks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013, pp. 21–30. [Online]. Available: http://doi.acm.org/10.1145/2470654.2470658

[13] F. Erlandsson, R. Nia, M. Boldt, H. Johnson, and S. F. Wu, "Crawling online social networks," in *2015 Second European Network Intelligence Conference*, Sept 2015, pp. 9–16.

[14] N. Grinberg, P. A. Dow, L. A. Adamic, and M. Naaman, "Changes in engagement before and after posting to facebook," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 564–574. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858501