How often should I access my online social networks?

Eduardo Hargreaves
Department of Computer Science
Federal University of Rio de Janeiro
Rio de Janeiro, RJ, Brazil
eduardo@hargreaves.tech

Daniel Sadoc Menasché
Department of Computer Science
Federal University of Rio de Janeiro
Rio de Janeiro, RJ, Brazil
sadoc@dcc.ufrj.br

Giovanni Neglia
Inria, Université Côte D'Azur
Sophia Antipolis, France
giovanni.neglia@inria.fr

Abstract—Users of online social networks are faced with a conundrum of trying to be always informed without having enough time or attention budget to do so. The retention of users on online social networks has important implications, encompassing economic, psychological and infrastructure aspects. In this paper, we pose the following question: what is the optimal rate at which users should access a social network? To answer this question, we propose an analytical model to determine the value of an access (VoA) to the social network. In the simple setting considered in this paper, VoA is defined as the chance of a user accessing the network and obtaining new content. Clearly, VoA depends on the rate at which sources generate content and on the filtering imposed by the social network. Then, we pose an optimization problem wherein the utility of users grows with respect to VoA but is penalized by costs incurred to access the network. Using the proposed framework, we provide insights on the optimal access rate. Our results are parameterized using Facebook data, indicating the predictive power of the approach.

Index Terms—Facebook, model, sampling, online social networks, age of information, value of access, News Feed, timelines

I. Introduction

Motivation: Users of online social networks are faced with a conundrum of trying to be always informed without having enough time or attention budget to do so. The latest version of iOS (12) and beta versions of Android send reports to users about the time they spend on apps installed on smartphones, and allow users to limit their screen time [8]. The time spent in social networks has important implications, which go beyond the social field, encompassing economic, psychological and network infrastructure aspects. From the economic point of view, in a world where "a wealth of information creates a poverty of attention" [34], attention becomes the scarce resource in the *attention economy* [11], [33].

Challenges: On social networking platforms, the dispute for attention occurs through the counting of clicks and the time spent viewing publications. Users of social network platforms typically consume content created by the sources they follow. If users do not access online social network frequently, important posts might be missed. On the other hand, if users access too often, the chance that there will be no new relevant content created in the interval between accesses is high. Ultimately, frequent accesses to social networks can be a source of challenges for users (e.g., reflecting addictions or consuming resources unnecessarily) and for social network managers (e.g., lack of new information can lead to a drop in the quality of the experience of users impairing engagement with the platform). Thus, determining the optimal frequency with which users should access a social network, taking into account the multiple and complex factors involved, is an important open problem.

Gap on prior art: Most social media platforms get their revenue by displaying ads to their users. For this reason, the literature on user retention in social networks focuses primarily on the effect of retention on profit, leveraging sponsored advertisements and publications [22]. In essence, it is assumed that the more frequently a user accesses the network, the more likely he is to consume a sponsored product [14], [37]. Users incur monetary and attention-related costs to access social networking platforms, and if they access the network too often they end up either receiving repeated content or content that does not necessarily suit their needs. *We are unaware of previous work that analyzed the optimal frequency with which users should access their social networks, considering the intrinsic value associated with each access.*

Methodology and goals: In this paper, we consider the problem of determining the optimal rate of access to a social network. To this purpose, we propose an analytical model that allows us to determine, depending on the rate at which sources generate content, the chance of a user accessing the network and obtaining new content. The proposed model yields a closed form expression for the value of an access (VoA), which we use as basis for an optimization problem. In the optimization problem, the utility of each user increases with respect to VoA, and decreases with respect to the cost associated to an access to the social network. The solution to the optimization problem provides insights on the optimal access rate. Our results are parameterized using Facebook data, which indicate the predictive power of the proposed approach.

Contributions: In summary, our main contributions are listed below.

- **New metric:** We develop a new metric, referred to as *value of access*, or VoA, to measure the amount of new information transmitted by a social network platform per

access. The metric is evaluated using data collected from Facebook.

- **Analytical model:** We present an analytical model to estimate the expected number of new impressions for each view of a *timeline*. The model produces simple closed form expressions for the metric of interest.
- **Optimal access frequency:** We pose an optimization problem that produces as result the optimal frequency of access, balancing between the cost per access and the utility derived from it. Using the model, we analytically indicate properties of the optimal solution.

Paper structure: The remainder of the paper is organized as follows. In Section II we introduce the analytical model that is validated in Section III. The problem of the optimal frequency of access is posed in Section IV, and numerical examples are reported in Section V. Related work and a discussion of the practical implications of the work follow in Sections VI and VII and Section VIII concludes.

II. MODEL

Next, we describe the proposed model. We consider sources that create posts that appear in users timelines. A user *access*, also referred to as *view*, is the act of refreshing a timeline. A *timeline*, in turn, is a sorted list of posts presented to a user. Every time a user accesses the network, a timeline is prepared. The timeline may contain posts that have already been presented to the user before. Nonetheless, there are no repeated posts in a given timeline. An *impression* is a post appearing in a timeline of a user.

Let $\mathcal{I}$ and $\mathcal{J}$ be a set of users and sources, respectively. Let $i \in \mathcal{I}$ be a user following a set of sources $\mathcal{J}_i$. The social network platform chooses the publications from $\mathcal{J}_i$ to be displayed in user's i timeline. Except otherwise noted, we focus on a given user i, and drop subscript from $\mathcal{J}_i$ whenever it is clear from context.

We assume that the set of sources followed by i creates publications following a Poisson process with an aggregate rate of λ publications per unit time. A timeline of size K contains at most K publications, i.e., the *timeline size* is the maximum number of elements in a timeline.

Let T be a random variable characterizing the interval between accesses of a given user to the social network. The average interval between views is given by $\overline{T}$. In addition, we denote by $f_T(t)$ the probability density function of T. We remove the subscript when clear from the context, denoting the density simply as $f(t)$.

The *access rate* (or view rate) is the rate at which a user accesses the network. We denote by μ be the user access rate, i.e., the frequency with which the user of interest accesses the social network. The access rate (or view rate) is our main quantity of interest.

We denote by ρ the average number of posts created between two user accesses. Then, we have

$$\mu = \frac{1}{\overline{T}}, \qquad \rho = \frac{\lambda}{\mu}. \tag{1}$$

A. Value of Access (VoA) and utility

At every access, a user receives a timeline possibly containing novel impressions. A *novel impression* is an impression corresponding to a post that has not yet been presented to that user, i.e., an impression is considered *new* if it is included in the timeline of that user for its first time. A *repeated impression* is an impression corresponding to a post that has already been presented to that user.

The *value of access (VoA)* is the expected number of novel impressions in a timeline. We assume that the number of new impressions serves as a proxy for the amount of valuable information received, and we interchangeably refer to the VoA and to the expected number of new impressions per access as synonyms.

Let V be the random variable denoting the number of novel impressions in a timeline, and $\overline{V}$ its expected value, i.e., $VoA = \overline{V}$. In addition, let A be the random variable characterizing the number of publications created by the sources in the interval between two views. Table I together with Appendix A summarize the notation used throughout this work.

The *access cost* is the cost of accessing a social network, e.g., measured in monetary units, but reflecting the attention span dedicated to the access. The *utility of an access* is the value of the access minus the associated costs.

The analytical model that follows produces the expected number of new impressions per view, $\overline{V}$, and the utility per time unit, as a function of

- the size of the timeline (K),
- the publication creation rate (λ) and
- the rate at which the user accesses the social network and refreshes the timeline (μ).

TABLE I
TABLE OF NOTATION (ACCESSES AND VIEWS ARE SYNONYMS).

variable	description
V	Value of Access (VoA); in this work, we assume equal to number of *new* impressions per access with fixed timeline size
$\overline{V}^{(P)}$	VoA assuming Poisson timeline size number of *new* impressions per access
K	size of timeline (number of impressions per access)
λ	rate of creation of publications (publications per unit time)
A	number of publications created between accesses
μ	average access rate
$1/\mu$	average interval between accesses
ρ	average amount of new publications per access

B. Assumptions

The analytical model is based on the following three assumptions:

(A1) Content is generated according to a Poisson process and the aggregate rate of content generation is λ.

(A2) The timeline displays as many new content as possible. If there is at least one new post that has not yet been displayed, and the user has space in the timeline to include such impression, i.e., if the number of newly created posts is smaller than or equal to K, the content will be displayed

generating a new impression. Note that the model is agnostic with respect to the order at which the contents are displayed in the timeline

(A3) Content already displayed will be repeated if there are not enough novelties. The motivation for repeating impressions across timelines is to fill the timeline if there are no new impressions available.

C. Quantifying the value of access

To compute the VoA, we first condition on the time between two accesses. Let τ be the time between accesses (fixed and given). In this case, given the assumptions of the model, the number of new impressions per access is given by a truncated Poisson distribution, Note that a user can receive a maximum of K new impressions per refresh of his timeline,

$$V = \begin{cases} A, & \text{if } A < K, \\ K, & \text{otherwise.} \end{cases} \tag{2}$$

where A is the number of publications created during τ,

$$P(A = a|\tau) = \frac{e^{-\lambda\tau}(\lambda\tau)^a}{a!}, \tag{3}$$

Hence,

$$P(V = i|\tau) = \begin{cases} P(A < i|\tau), & \text{if } i < K \\ P(A \geq i|\tau), & \text{if } i = K \\ 0, & \text{if } i > K \end{cases} \tag{4}$$

The three conditions above correspond to situations in which (i) the generated content is insufficient to fill the timeline and (ii) the timeline is filled exclusively with new content. The third and last condition (iii) captures the fact that the timeline has space for up to K impressions.

It follows from (4) that $E(V|\tau)$ is given by

$$E(V|\tau) = \left(\sum_{i=0}^{K} iP(A = i|\tau)\right) + KP(A > K|\tau). \tag{5}$$

As $P(A > K|\tau) = 1 - P(A \leq K|\tau)$, we have:

$$E(V|\tau) = \left(\sum_{i=0}^{K} ie^{-\lambda\tau}\frac{(\lambda\tau)^i}{i!}\right) + K\left(1 - \sum_{i=0}^{K} e^{-\lambda\tau}\frac{(\lambda\tau)^i}{i!}\right) \tag{6}$$

$$= K - \sum_{i=0}^{K}(K - i)e^{-\lambda\tau}\frac{(\lambda\tau)^i}{i!} \tag{7}$$

Finally, by applying the law of total probability,

$$\overline{V} = \int_{\tau=0}^{\infty} E(V|\tau)f(\tau)d\tau. \tag{8}$$

Next, we consider special cases of the expression above, as a function of the distribution of time between users' accesses.

D. Special case I: deterministic time between accesses

Let κ be the deterministic time between samples. Then,

$$f(\tau) = \delta_\kappa(\tau) \tag{9}$$

where $\delta_\kappa(\tau)$ is a Delta Dirac function, with all mass at point κ.

Note that in this case the number of posts created between two accesses is distributed according to a Poisson variable with mean $\kappa\lambda$. Then,

$$\overline{V} = \int_{\tau=0}^{\infty} E(V|\tau)f(\tau)d\tau = E(V|\tau) \tag{10}$$

and (10) is given by (7).

E. Special case II: exponentially distributed interval between accesses

Next, we consider an exponentially distributed time between accesses. Let T be an exponentially distributed random variable with mean $1/\mu$, i.e., $f(\tau) = \mu e^{-\mu\tau}$. Substituting $f(\tau)$ into (8),

$$\overline{V} = \int_{\tau=0}^{\infty} \left(K - \sum_{i=0}^{K}(K - i)e^{-\lambda\tau}\frac{(\lambda\tau)^i}{i!}\right)\mu e^{-\mu\tau}d\tau. \tag{11}$$

After algebraic manipulation (see Appendix B), we get a compact closed form expression for $\overline{V}$,

$$\overline{V} = \frac{\lambda}{\mu}\left(1 - \left(\frac{\lambda}{\lambda + \mu}\right)^K\right). \tag{12}$$

a) Probabilistic interpretation: As the interval between publications and the interval between accesses are exponentially distributed with rates equal to λ and μ, the ratio $\lambda/(\lambda + \mu)$ is the probability that a new publication occurs before the user accesses the social media [13]. Then, due to the lack of memory of the exponential distribution, $(\lambda/(\lambda+\mu))^K$ is the probability that at least K publications have occurred between two accesses. In this case, such publications are sufficient to fill the timeline with novelties.

Let $p = 1 - (\lambda/(\lambda + \mu))^K$ be the probability that the user accesses the system before the arrival of K consecutive new publications. When $p \approx 1$, the user will receive, on average, λ/μ new impressions per view.

b) Limiting probabilities: The following limits apply:

Size of the timeline $\lim_{K\to\infty} \overline{V} = \lambda/\mu$. In this scenario, the user can read all the novelties created by the sources between two accesses, and the average number of novelties per access is equal to λ/μ;

Content generation $\lim_{\lambda\to\infty} \overline{V} = K$ and $\lim_{\lambda\to0} \overline{V} = 0$ In this scenario, sources generate content at high (resp., low) rate, and the VoA is maximum (resp., minimum), and equal to K (resp., 0);

Access frequency $\lim_{\mu\to0} \overline{V} = K$ and $\lim_{\mu\to\infty} \overline{V} = 0$ In this scenario, users access the system with low (resp., high) frequency, and the number of news per view is maximum (resp., minimum), and equal to K (resp., 0).

In the upcoming sections, we validate the proposed model using data collected from Facebook during the Brazilian elections in 2018.

III. EXPERIMENTAL RESULTS

The proposed analytical model was validated in a two-step process using Facebook data. First, we assess the accuracy of the model in a trace-driven simulation using only the posts created by a set of selected sources. Then, we incorporate the effect of the Facebook News Feed algorithm [15] by measuring the VoA directly at the browser's of a set of virtual users.

During the experiment, we selected the publications of the pages of the candidates for the Presidency of the Republic of Brazil on Facebook, as well as those of their respective political parties, during the first round of the electoral campaign of the presidential elections of 2018. We also selected a set of 30 representative Brazilian's media outlets. The data were obtained both through the Facebook API and through a browser extension named Facebook Tracking Exposed [1]. We collected the publications created between September 27, 2018 and October 10, 2018. The pages considered, and their respective addresses, are summarized in the Appendix C.

A. Trace-driven FIFO simulations without Facebook News Feed filtering

We start by validating the model before the influence of the News Feed algorithm. To this aim we envisioned collecting all the publications from the selected sources during the experiment period, through the Facebook API. Unfortunately, we could not obtain posts created by all sources because the Facebook API only retrieved posts from the candidates and their respective political parties during this period. For this reason, we restrict to only those set of publishers in the trace-driven experiment.

Figure 1(a) shows the total number of publications per day for all selected pages during the experiment period. The aggregate number of publications by all sources varied between 30 and 193 publications per day. The number of publications ranged from a minimum of 4 to a maximum of 222 posts per page.

After selecting the sources, the behavior of a FIFO timeline of size $K = 10$ was simulated using the collection of the considered posts. Each snapshot is an instance of a timeline of size K. The simulation consists of building a sequence of snapshots. A snapshot is built selecting a post uniformly at random from the available posts and the $K - 1 = 9$ immediately preceding posts. We refer to the selected post as the *reference post* of the snapshot. For further details on the behavior of FIFO timelines, we refer the reader to [17].

Given the desired sample interval, $1/\mu$, measured in hours, we determined the number of snapshots we took from the system. If $\mu = 1$, for instance, in a period of 14 days we draw 24×14 snapshots. The snapshots are then ordered in time, based on their reference posts, and from this list of snapshots we assess the value of each access (VoA).

In the experiments considered, $1/\mu$ ranged from 1 to 24. For each value of μ, 30 simulation rounds were performed. The results are shown in Figure 1(b). The red circles are the data predicted by the model through equation (12), with $\lambda = 4.487$ and $K = 10$. The blue triangles are the averages of the values obtained in the simulations. The deviation from the simulation is due to the fact that posts are created following a real (possibly non-Poisson) generation process. Nonetheless, this small difference between the data predicted by the model and the data obtained through simulation, indicates the Poisson process is a good approximation for the generating process. Note that as $1/\mu$ increases, VoA tends to the asymptotic value given by $K = 10$, that is, the timeline will be filled completely with new information, with high probability, and the marginal gains due to an increase in the access rate are smaller.

B. Facebook Tracking Exposed (FTE) experimental setup

Next, our goal is to assess the effect of the Facebook News Feed personalization algorithm on the VoA. For this purpose, we ran an experiment similar to the one presented in [17]. We created 5 virtual users, that will be referred to as bots. The bots followed *all* the thirty selected *pages*, most of them being media pages (see Appendix C), in addition to the political candidates and their parties.

Each bot kept open an Internet browser window (Firefox or Chrome) accessing the Facebook page. The bots were prepared to collect data on the posts to which they were exposed using the Facebook Tracking Exposed [1] browser extension. The extension auto-scrolls the Facebook page for three minutes at scheduled moments. Every auto-scroll produces a set of impressions which are stored at a local database. Each collection of posts is referred to as a *snapshot*.

1) One bot was scheduled to collect snapshots every ten minutes. We refer to this user as a *high* sampling user.
2) The *regular* bot collects posts every hour. There are four regular bots.

Each post appearing in a snapshot counts as a post *impression*. At each bot, Facebook Tracking Exposed collects all impressions and stores their corresponding publisher, publication time, impression time and impression order.

C. Experimental results accounting for Facebook News Feed filtering

Next, we report the results obtained using the methodology described in Section III-B.

Figure 2(a) shows the average VoA as a function of the timeline size for the two types of bots. To generate that curve from the collected measurements, for each point K in the abscissa, we truncated the snapshot size to K impressions, i.e., all the impressions that appeared in positions greater than K were not considered in the VoA calculation.

We also compared the model against an alternative approach to Facebook News Feed: a timeline following a reverse chronological order, i.e., a FIFO News Feed. To this aim, for each snapshot taken, we changed the order of the impressions such
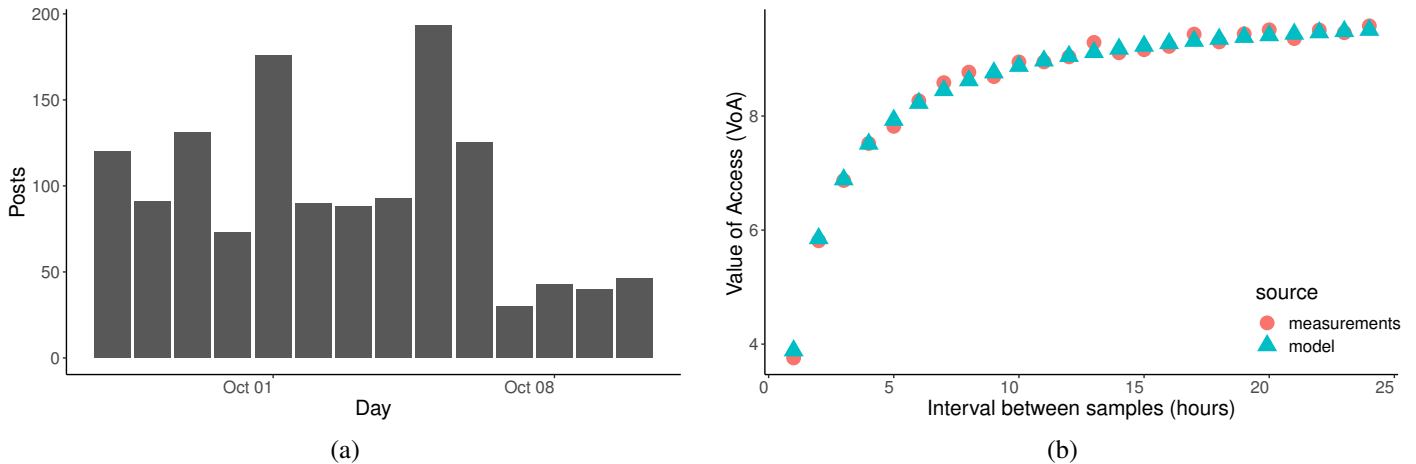
Fig. 1. (a) Total number of publications per day of selected pages and (b) parameterization and validation of the model with Facebook API data

that most recent posts belonging to that snapshot appear at the topmost positions of the timeline's snapshot.

In each subfigure of Figure 2, the VoA predicted by the model for each value K in the abscissa is compared against the measured VoA and against the FIFO timeline for the two types of bots. At the high sampling bot, both model and measurements have similar shapes, with model predicting higher VoA than observed in measurements. For small values of K, i.e. $K < 5$, the FIFO News Feed and model are very close, while for larger values of K, i.e. $K > 10$, the FIFO News Feed approaches the News Feed measurements.

At the regular sampling bots the model again typically predicts a VoA higher than the measurements of Facebook News Feed. After reordering the posts, the VoA of the FIFO News Feed is accurately captured by the model. Therefore, the difference in the VoA as predicted by the proposed model and the measurements is attributed to News Feed reordering. This finding reveals that Facebook Facebook creates a *position bias* [5] by showing repeated posts and is consistent with the conclusions of [17]. Note, for instance, that when $K = 10$ it is expected according to the proposed model that the timeline can be fulfilled with new posts, i.e. VoA= 10, but the Facebook News Feed recommends only 5 new posts on average. To reach a VoA of 10, a Facebook user must scroll 20 posts on average.

Due to reordering, Facebook News Feed algorithm shows repeated posts, in detriment of newer ones, at the top of the timeline. One possible reason for this behavior is that Facebook shows the same post multiple times to assure that users will more likely see a post. To overcome this effect, users must scroll their timelines beyond the repeated posts to see new contents and obtain all the value of the access. Another possible reason is that Facebook strategically reduces the VoA as a way to increase the users' access rate. It worth noting that the position bias has important consequences on the behaviour of users as there is a strong correlation between post positions and click rates [6], [36].

When $15 < K < 35$, the model underestimates the FIFO News Feed VoA. According to equation (2), if $A > K$ only the

K most recent posts enter the timeline, while the older ones are dropped by the social media. In reality, however, those posts can appear in future accesses. This effect becomes more noticeable in the region wherein VoA increases non-linearly with respect to K. In the linear region, i.e. $VoA = K$, it is expected that all snapshots are filled with K new posts on average. However, beyond this region posts that are dropped in an access may be necessary to fill timelines in future accesses, if $A < K$ in those latter accesses. In future work, we plan to improve the model by incorporating a buffer to handle bursts of posts through an $M/G/K$ with bulk services.

D. Post overlap

Next, we evaluate the amount of overlap between the posts viewed by the bots. This analysis objectives are two-fold: first the overlapping serves as an hypothesis check to confirm that by accessing the social network less often the bots indeed collected roughly a subset of the posts observed by the bots with larger access rates. Second, the overlapping can serve to verify if Facebook News Feed algorithm has influence over VoA.

Table II shows the overlap between the impressions observed by the high sampling bot and all four regular bots. The notation goes as follows:

- a post in XY was viewed by the two bots,
- a post in $X\bar{Y}$ was viewed only by X,
- a post in $\bar{X}Y$ was viewed only by Y,
- a post in $\bar{X}\bar{Y}$ was not seen by any of the two bots.

The high sampling bots viewed at least 80% of the regular sampling bots' impressions, indicating that regular bots impressions are a subset of the impressions of the high sampling bot.

Table III shows that the overlap between the four regular bots ranges between 74% to 89%. Figure 3 shows the ECDF of the number of posts that were viewed by x overlapping bots. It can be seen that almost half of the posts were viewed by all the 4 bots and one quarter were viewed by only on bot. It is worth noting that all bots follow the same set of pages and
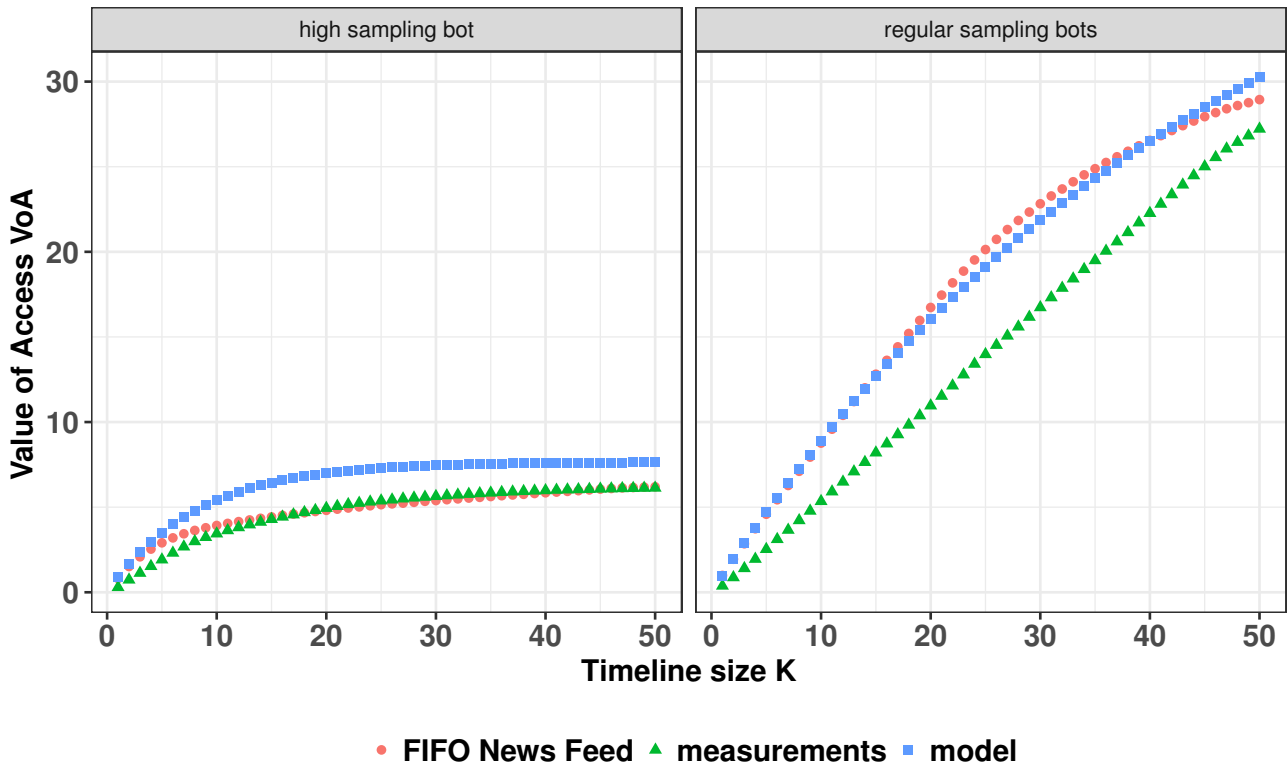
Fig. 2. Model (squares) compared against the measured VoA (triangles) and against a News Feed following a chronological order (circles).

are scheduled to access Facebook simultaneously, thus, these overlapping patterns help to detect the impact of the Facebook News Feed algorithm in the bots impressions. These results together with the findings reported in Section III-C indicate that Facebook News Feed personalization algorithm reduces the VoA when deciding which posts will be displayed to the users, particularly in the topmost positions.

Tables II and III also indicate that the low sampling bots collected posts that were not seen by any other bot. This indicates that studying the impact of the rate at which the network is accessed is non-trivial when we account for the filtering effect of the News Feed. A number of factors such as the time of the day, the sex of the bot and the dynamic evolution of number of likes to a given post are some of the various variables that affect the popularity of posts, which in turn impact the VoA. Although Tables II and III indicate that an analysis which neglects such second order effects is already valuable, they also point to the fact that there is significant space for improvement in the analysis of the VoA beyond the simple models considered in this paper.

IV. FINDING THE OPTIMAL SAMPLING RATE

In this section we pose the optimal access rate problem. We consider users who have a cost to access their social networks. This cost reflects, for instance, the users attention budget, as well as the energy spent from the batteries of mobile devices and possible monetary costs associated to using data plans for

TABLE II
POST OVERLAP BETWEEN THE HIGH SAMPLING BOT AND THE REGULAR SAMPLING ONES

	R_1	$\bar{R}_1$		R_2	$\bar{R}_2$
H	8205	2778	H	7242	3751
$\bar{H}$	651	2679	$\bar{H}$	1767	1563

(a) Regular I   (b) Regular II

	R_3	$\bar{R}_3$		R_4	$\bar{R}_4$
H	8381	2602	H	8111	2872
$\bar{H}$	1128	2202	$\bar{H}$	801	2529

(c) Regular III   (d) Regular IV

cellular access. Let c be the cost incurred by a user per access, measured in the same units as the VoA.

The *utility* per time unit experienced by a user is given by the expected number of new impressions per view ($\overline{V}$), times the rate of views per time unit (μ), minus the cost $c\mu$ required to recover these impressions. Therefore,

$$U(\overline{V}, \mu) = \mu\overline{V} - c\mu. \qquad (13)$$

Substituting (12) into (13), we express the utility as a function of λ and K:

$$U(\mu) = \lambda\left(1 - \left(\frac{\lambda}{\lambda + \mu}\right)^K\right) - c\mu. \qquad (14)$$

TABLE III
POST OVERLAP BETWEEN THE REGULAR SAMPLING BOTS

	R_2	$\bar{R}_2$			R_3	$\bar{R}_3$
R_1	6776	2080		R_1	6776	2080
$\bar{R}_1$	2223	3224		$\bar{R}_1$	1453	4004

(a) Regular 1,2   (b) Regular 1,3

	R_4	$\bar{R}_4$			R_3	$\bar{R}_3$
R_1	7850	1006		R_2	6855	2144
$\bar{R}_1$	1062	4395		$\bar{R}_2$	2654	2660

(c) Regular 1,4   (d) Regular 2, 3

	R_4	$\bar{R}_4$			R_4	$\bar{R}_4$
R_2	6641	2358		R_3	7734	1775
$\bar{R}_2$	2271	3043		$\bar{R}_3$	1178	3623

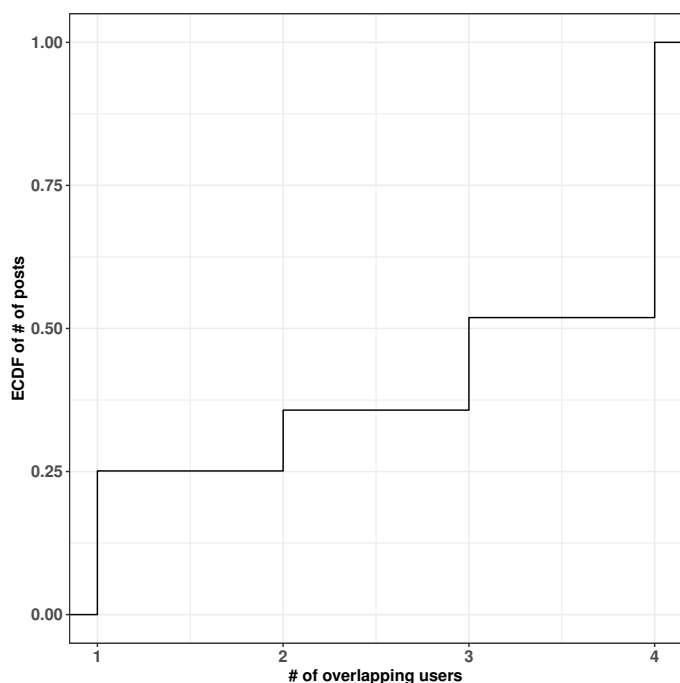(e) Regular 2,4   (f) Regular 3, 4

Fig. 3. ECDF for the number of posts viewed versus the number of overlapping users

Therefore, given λ and K, the problem of determining the optimal access rate to a social network consists of maximizing the utility by adjusting the access rate μ,

$$\max \quad U(\mu), \quad \mu \geq 0, \tag{15}$$

where $U(\mu)$ is given by (14). The utility is maximized when its partial derivative with respect to the control variable is equal to zero, i.e., when $\partial U/\partial \mu = 0$. The partial derivative $\partial U/\partial \mu$ is given by

$$\frac{\partial U}{\partial \mu} = K \left(\frac{\lambda}{\lambda + \mu} \right)^{K+1} - c. \tag{16}$$

Then, letting $\partial U/\partial \mu = 0$ we obtain the optimal access rate, $\mu^\star$,

$$\mu^\star = \lambda \left(\frac{\sqrt[K+1]{K}}{\sqrt[K+1]{c}} - 1 \right). \tag{17}$$

Next, we numerically evaluate the behavior of the optimal access rate, as a function of the different system parameters.

V. NUMERICAL EXAMPLES

In this section we present numerical examples obtained from the proposed model. Our objectives are (i) to perform sensitivity analysis of the metrics of interest given the parameters of the model (proposed in Section II) and (ii) to illustrate the sensitivity of the solutions of the optimal access rate problem (introduced in Section IV) to its parameters.

A. Impact of timeline size K and ratio $\rho = \lambda/\mu$

Next, we evaluate the marginal increment of VoA resulting from users scrolling down their screen to capture more access impressions, i.e., we evaluate the impact of an increase in K on VoA. Figure 4(a) shows the impact of K on VoA, for a fixed value of ρ. At the limit when $K \to \infty$, $\overline{V}(\rho, K) = \rho$. Note also that in all the scenarios considered the convergence for values close to the asymptotic value occurs for $K \leq 100$.

For example, when $\rho = 10$, that is, when the access rate is 10 times smaller than the publication rate, $\overline{V} \approx \rho$ for $K \geq 40$. Thus, even if a user accesses the network infrequently, at $\mu = \lambda/10$, he can still absorb almost every novelty provided he can recover and read $K = 40$ posts per access.

When users capture K impressions per access, the maximum reachable VoA is K. How does the VoA actually experienced by users compare to the maximum achievable VoA?

In the dotted line shown in Figure 4(a) we have the VoA corresponding to the situation in which all received publications are new. Figure 4(a) shows that even for small values of K, e.g., $K = 10$, the experienced VoA already significantly distances itself from K. This effect is caused both by fluctuations in the number of publications between accesses and by variations in the interval between accesses. *Note that if the accesses and the publications are deterministic and synchronized, the dotted and full curves will overlap. However, publications arise organically, due to exogenous factors, and dealing with unpredictability is an intrinsic aspect of the problem.*

Figure 4(b) shows the increase of VoA as a function of ρ. When $\rho \to \infty$, $\overline{V}(K, \rho) = K$. Note that the greater the value of K, the greater the value of ρ required to reach the asymptotic value $\overline{V} = K$. In fact, the greater the value of K, the more active the sources need to be in order to fill in the timelines of users with novel impressions.

B. Optimization problem

Figure 5(a) shows the utility experienced by a user per time unit as a function of the access rate μ, for five values of K. In this figure, λ has been parameterized with the data returned
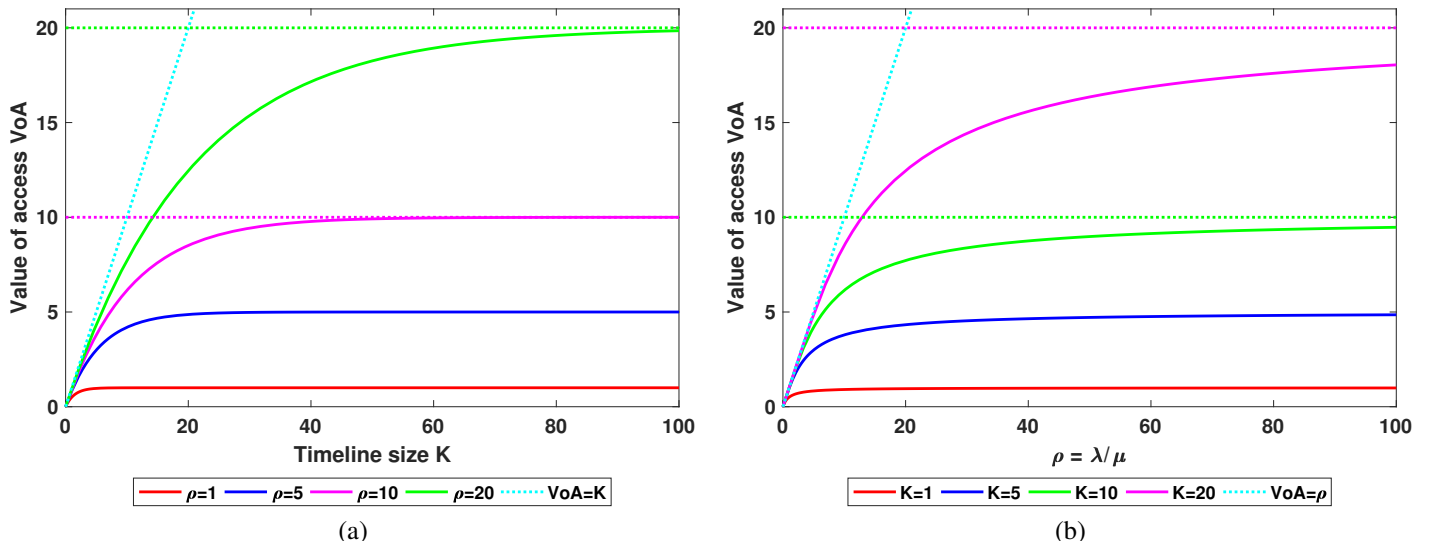
Fig. 4. Sensitivity analysis: impact of different parameters (K, ρ) on the value of an access (VoA).

by the Facebook API as described in Section III-A. The cost c is fixed equal to 1. This figure shows that, given the other parameters constant, the higher the value of K, the lower the optimal access rate. For example, when K increases from 2 to 20, $\mu^\star$ decreases from 1.1663 to 0.688 accesses/hour. For a given pair (λ, μ), the greater the value of K, the greater the utility. This is because we assume in our model that the cost of an access is independent of K. More complex cost functions will be subject of future work.

Figure 5(b) shows the utility experienced by a user per time unit as a function of the access rate μ for four values of λ (with $K = 10$). In this case, as λ grows the optimal access rate increases. For a given pair (K, μ), the utility increases with respect to λ. In fact, more active sources will favor an increase of VoA, remembering that such value is limited by K.

Figure 5(c) shows the optimal access rate $\mu^\star$ as a function of K. Note that for values of K close to one, $K \approx 1$, as K increases, $\mu^\star$ increases as the timeline can still be filled with additional novel posts. However, if K is further increased the system eventually transitions to a regime wherein as K grows the optimal sampling rate $\mu^\star$ decreases. This is because the probability of observing repeated impressions eventually becomes non-negligible.

Figure 5(d) shows the optimal access rate $\mu^\star$ as a function of λ, for $K \in \{2, 5, 10, 20\}$. This figure illustrates the linear behavior of $\mu^\star$ as a function of λ (see (17)). Next, consider Figure 5(d) for a given value of λ. Note that when K grows from $K = 2$ to $K = 5$, $\mu^\star$ increases as K grows, as the timelines are still not saturated by the new posts being created. As K increases in the range from $K = 5$ to $K = 20$, however, $\mu^\star$ decreases. This is because in this regime the timelines are already saturated, and the probability to receive repeated impressions increases. In this case, *if users can consume more publications per access, they can access the network less frequently (see also Fig. 5(c)).*

C. Practical implications on social network design

The numerical examples presented in this section illustrate that VoA can be limited either by ρ or K. The system then goes through two operating regimes:

- Consider the regime in which $\overline{V} \approx \rho$. We consider a fixed and given value of K, and note that in this regime the increase of K typically does not result in an increase of $\overline{V}$. If the user is able to consume K new impressions per access, he does not need to use social media filters;
- In the regime wherein $\overline{V} \approx K$ the user is potentially losing many new publications. In that case, he needs the help of the social network's recommendation system to filter out the most relevant publications.

Understanding these two regimes helps in the design of online social network timelines. Characterizing the regime in which recommendation algorithms such as the Facebook News Feed are needed to add value to users is key to determine when and if filters play a more prominent role and to assess their impact on the fairness and accountability of online social networks.

VI. RELATED WORK

The rate at which users access their online social networks has implications that range from Psychology and Economics to network infrastructure and performance. Next, we discuss prior work on each of those three domains. Then, we also briefly survey previous efforts on analytical models for online social network timelines. *We are unaware of previous work considering analytical models to estimate the optimal rate at which users should access their online social networks, and that has parameterized such models with real data.*

A. Psychology and addiction

The appropriate rate for users to access online social networks is a theme that has been widely discussed in the sphere of Psychology [3], [20], [21], [31], [32]. According to Sean Parker, founder of Napster and former president
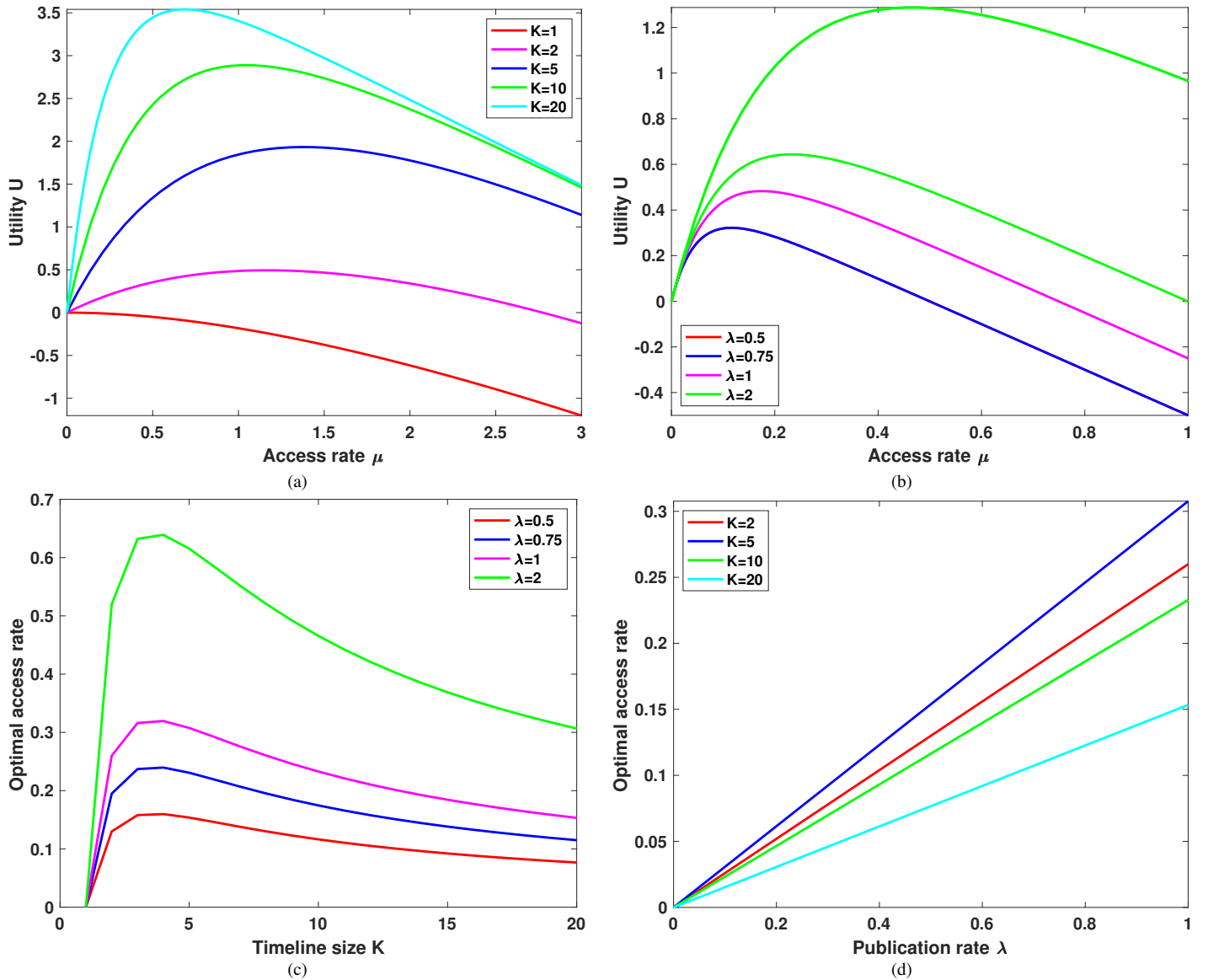
Fig. 5. Sensitivity analysis: impact of different parameters (K, λ, μ) on users' utility ((a) and (b)) and optimal access rate ((c) and (d)).

of Facebook, *Facebook is a system designed to exploit human vulnerability* [35]. Accordingly, Tristan Harris, a former Google employee, claimed that *technological artifacts can hijack people's minds* [19]. Although in this work our focus is exclusively related to technical questions regarding the amount of novelty users get each time they access a social network, the psychological implications of this work are evident. In particular, if users know in advance that they will receive little news when accessing the network at any given moment, they will have more incentives to further space their access.

B. Attention economy and ads

The rate at which users access online social networks has immediate economic implications, given that advertisements generate profits based on *click rate*. The click rate, in turn, is strongly influenced by the rate at which users access their online social networks. If advertisers have access to

information on the recommended rate for users to access their online social network throughout the day, they can better decide when to pay for certain advertisements to be displayed, e.g., depending on the target audience [22], [30].

The dynamics of collective attention and the way that popularity of topics evolves over time was studied in [27]. Combining these dynamics with user's behaviour expressed by the sampling rate and filtering algorithms [17], [28], one can derive insights on the dynamics of topic popularity and production of information [10], particularly when there is contention for visibility and attention.

C. Infrastructure and age of information (AoI)

The rate at which users access online social networks has a direct impact on the traffic passing through the network. Since social networks are currently one of the largest sources of traffic, recommendations on the rate at which users should

access their pages can have cascading effects on the network as a whole. The age of information (AoI) has gained considerable visibility in the scientific community, and we envision that the contributions of this work are relevant in this new domain [2], [4], [7], [9], [23], [25], [38], [39].

D. Modeling and evaluation of timelines' performance

Facebook's News Feed was modeled as a FIFO queue in [18]. In [17] it has been shown that FIFO timelines are a particular case of time-to-live (TTL) timelines [12]. TTL timelines are versatile and flexible, allowing the reordering of publications. In this paper, we build on top of [17] and [16] to account for the impact of the user's access rate, opening up new perspectives of analysis and enabling the formulation of the problem of optimal access rate of users.

VII. Discussion

The ideas presented in this work can be implemented in an intermediate layer between the user and the social network, e.g., through an application that consumes the Facebook News Feed and learns the parameters of the model proposed in this paper. Platforms like Gobo [29] already act as a *middleware* between Facebook and users, and can serve as a basis for implementing the ideas presented here.

The results contained in this paper do not consider all the impact that online social networking algorithms such as Facebook's News Feed can have on VoA. In future work, we intend to quantify the VoA by user profile and by social network, e.g., using platforms such Facebook and Twitter. From this characterization of the VoA, we envision to use VoA as a criterion for comparing online social networks personalization algorithms and their effects on the information diets [26], e.g., accounting for bots with multiple political orientations [17].

VIII. Conclusion

Social networks are increasingly present in the daily lives of millions of Internet users, and a significant portion of the world's population accesses social networks multiple times on a daily basis [24]. Each access to an online social network yields a certain *value* to its user, which is a function of numerous subjective aspects. For this reason, measuring the *value of access* (VoA) is a challenging endeavour.

In this work, we overcome the challenges associated to the assessment of VoA by proposing a simple but quantifiable approach to estimate the VoA experienced by a user of an social network. In particular, we assume that this value is equal to the amount of new impressions presented by the social network to the user. Using this approach, we proposed an analytical model to quantify the VoA. The model was validate using real-world data from Facebook experiments. Next, we parameterize the model using real data, and indicate how the model can be used to obtain the optimal access rate.

The results enable us to conclude that Facebook News Feed algorithm reduces the VoA by deciding that some repeated post are more relevant to the users than other fresher and unseen posts, specially in the topmost positions. Even bots following the same set of publishers, and accessing Facebook simultaneously, experienced a personalized information diet. We believe this work is yet another step in a research agenda that aims to develop social networking platforms respecting users' preferences. This agenda necessarily involves contemplating the psychological, economic and infrastructure aspects inherent to the problem of determining the optimal rate at which users should access online social networks, which we intend to embrace in future works.

Appendix

A. Terminology

Next, we summarize the basic terminology used throughout this paper.

- **Post** is a message created by a source
- **Access**, also referred to as *view*, is the act of refreshing a timeline
- **Timeline** is a sorted list of posts presented to a user. Every time a user accesses the network, a timeline is prepared. The timeline may contain posts that have already been presented to the user before. Nonetheless, there are no repeated posts in a given timeline
- **Impression** is a post appearing in a timeline of a user
- **Timeline size** is the maximum number of elements in a timeline
- **Access rate** is the rate at which a user accesses the network. At every access, the user can possibly receive a timeline containing novel impressions
- **Novel impression** is an impression corresponding to a post that has not yet been presented to that user
- **Repeated impression** is an impression corresponding to a post that has already been presented to that user
- **Value of access (VoA)** is the number of novel impressions in a timeline
- **Access cost** is the cost of accessing a social network, e.g., measured in monetary units, but reflecting the attention span dedicated to the access
- **Utility of an access** is the value of the access minus the associated costs.

B. Expression for $\overline{V}$ under the exponential model

From (7) and (8) we have

$$\overline{V} = K - \sum_{i=0}^{K} \frac{(K-i)}{i!} \lambda^i \mu \int_{\tau=0}^{\infty} e^{-(\lambda+\mu)\tau} \tau^i d\tau \qquad (18)$$

The solution of the integral above can be found in standard tables of integrals,

$$\overline{V} = K - \sum_{i=0}^{K} \frac{(K-i)}{i!} \lambda^i \mu (\lambda+\mu)^{-(i+1)} \Gamma(i+1) \qquad (19)$$

Then, applying the definition of the gamma function, $\Gamma(i+1)$,

$$\overline{V} = K - \sum_{i=0}^{K} \frac{(K-i)}{i!} \frac{\lambda^i \mu}{(\lambda + \mu)^{(i+1)}} i! \qquad (20)$$

Reorganizing the terms in (20),

$$\overline{V} = K - \frac{\mu}{\lambda + \mu} \left(K \sum_{i=0}^{K} \frac{\lambda^i}{(\lambda + \mu)^i} - \sum_{i=0}^{K} \frac{i\lambda^i}{(\lambda + \mu)^i} \right). \quad (21)$$

After algebraic manipulation, and noting that the two summations in (21) have closed expressions, we obtain a simple expression for $\overline{V}$ (see (12)),

$$\overline{V} = \frac{\lambda}{\mu} \left(1 - \left(\frac{\lambda}{\lambda + \mu} \right)^K \right). \qquad (22)$$

C. List of publishers and pages

Table IV reports the list of publishers and pages followed by the bots considered in our experimental setup.

REFERENCES

[1] Claudio Agosti. Facebook tracking exposed, 2018. https://facebook.tracking.exposed.

[2] Eitan Altman, Rachid El-Azouzi, Daniel Sadoc Menasche, and Yuedong Xu. Forever young: Aging control for smartphones in hybrid networks. In *Mobihoc*, 2019.

[3] Cecilie Schou Andreassen, Torbjørn Torsheim, Geir Scott Brunborg, and Ståle Pallesen. Development of a facebook addiction scale. *Psychological reports*, 110(2):501–517, 2012.

[4] Onur Ayan, Mikhail Vilgelm, Markus Klügel, Sandra Hirche, and Wolfgang Kellerer. Age-of-information vs. value-of-information scheduling for cellular networked control systems. *CoRR*, abs/1903.05356, 2019.

[5] Ricardo Baeza-Yates. Bias on the web. *Communications of the ACM*, 61(6):54–61, may 2018.

[6] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, 2015.

[7] Ahmed M. Bedewy, Yin Sun, Sastry Kompella, and Ness B. Shroff. Age-optimal sampling and transmission scheduling in multi-source systems. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Mobihoc '19, pages 121–130, New York, NY, USA, 2019. ACM.

[8] Pia Ceres. Screen time controls. *Wired*, 2018. http://tinyurl.com/ioswired.

[9] Augustin Chaintreau, Jean-Yves Le Boudec, and Nikodin Ristanovic. The age of gossip: Spatial mean field regime. *SIGMETRICS Perform. Eval. Rev.*, 37(1):109–120, June 2009.

[10] Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. The production of information in the attention economy. *Scientific Reports*, 5(1):9452, aug 2015.

[11] M.B. Crawford. *The World Beyond Your Head: On Becoming an Individual in an Age of Distraction*. Farrar, Straus and Giroux, 2015.

[12] Mostafa Dehghan, Laurent Massoulie, Don Towsley, Daniel Menasche, and Yong Chiang Tay. A utility optimization approach to network cache design. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9. IEEE, 2016.

[13] Richard Durrett. *Essentials of Stochastic Processes*, volume 7 of *Springer Texts in Statistics*. Springer New York, New York, NY, oct 2012.

[14] Facebook. About ad auctions, 2019. http://bit.ly/2Ou1PUE.

[15] Facebook. Welcome to News Feed. Retrieved August 02, 2019 from: https://newsfeed.fb.com/?lang=en, 2019.

[16] Anastasios Giovanidis, Bruno Baynat, and Antoine Vendeville. Performance analysis of online social platforms. In *2019 Proceedings of IEEE INFOCOM*, 2019.

[17] Eduardo Hargreaves, Claudio Agosti, Daniel Menasché, Giovanni Neglia, Alexandre Reiffers-Masson, and Eitan Altman. Fairness in online social network timelines: Measurements, models and mechanism design. *Performance Evaluation*, 129:15–39, feb 2019.

[18] Eduardo Hargreaves, Claudio Agosti, Daniel Menasché, Giovanni Neglia, Alexandre Reiffers-Masson, and Eitan Altman. Biases in the facebook news feed: a case study on the italian elections. In *Proceedings of the IEEE/ACM International Conference on Social Networks Analysis and Mining (ASONAM 2018)*, 2018.

[19] Tristan Harris. How technology hijacks people's minds: from a magician and Google's design ethicist, 2016. http://www.tristanharris.com/essays/.

[20] Fu-Yuan Hong, Der-Hsiang Huang, Hung-Yu Lin, and Su-Lin Chiu. Analysis of the psychological traits, facebook usage, and facebook addiction model of taiwanese university students. *Telematics and Informatics*, 31(4):597 – 606, 2014.

[21] Zaheer Hussain, Boban Simonovic, Edward JN Stupple, and Maggie Austin. Using eye tracking to explore facebook use and associations with facebook addiction, mental well-being, and personality. *Behavioral Sciences*, 9(2):19, 2019.

[22] Bo Jiang, Nidhi Hegde, Laurent Massoulié, and Don Towsley. How to optimally allocate your budget of attention in social networks. In *2013 Proceedings of IEEE INFOCOM*, 2013.

[23] Igor Kadota and Eytan Modiano. Minimizing the age of information in wireless networks with stochastic arrivals. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Mobihoc '19, pages 221–230, New York, NY, USA, 2019. ACM.

[24] Simon Kemp. Digital in 2018. *Wired*, 2018. http://bit.ly/2uChFUm.

[25] Antzela Kosta, Nikolaos Pappas, Vangelis Angelakis, et al. Age of information: A new concept, metric, and tool. *Foundations and Trends® in Networking*, 12(3):162–259, 2017.

[26] Juhi Kulshrestha, Muhammad Bilal Zafar, Lisette Espin Noboa, Krishna P Gummadi, and Saptarshi Ghosh. Characterizing information diets of social media users. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'15, pages 218–227, 2015.

[27] Philipp Lorenz-Spreen, Bjarke Mørch Mønsted, Philipp Hövel, and Sune Lehmann. Accelerating dynamics of collective attention. *Nature Communications*, 10(1):1759, dec 2019.

[28] Stefania Milan. When Algorithms Shape Collective Action: Social Media and the Dynamics of Cloud Protesting. *Social Media and Society*, 1(2), 2015.

[29] MIT. Gobo – your social media, your rules. https://gobo.social/, 2018.

[30] Steffen Rendle. Social network and click-through prediction with factorization machines. In *KDD-Cup Workshop*, page 113, 2012.

[31] Tracii Ryan, Andrea Chester, John Reece, and Sophia Xenos. The uses and abuses of facebook: A review of facebook addiction, 2014.

[32] Mohammad Salehan and Arash Negahban. Social networking on smartphones: When mobile phones become addictive. *Computers in Human Behavior*, 29(6):2632–2639, 2013.

[33] C. Shapiro and H. Varian. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press, 1998.

[34] Herbert A. Simon. Designing organizations for an information rich world. In Martin Greenberger, editor, *Computers, communications, and the public interest*. Baltimore, 1971.

[35] Olivia Solon. Ex-Facebook president Sean Parker: site made to exploit human 'vulnerability'. *The Guardian*, 2017. http://bit.ly/2HJTufh.

[36] Zeynep Tufekci. How facebook's algorithm suppresses content diversity (modestly) and how the newsfeed rules your clicks. https://tinyurl.com/facebooksuppress, 2015.

[37] Twitter. Bidding and auctions, 2019. http://bit.ly/2JXj6H8.

[38] Xianwen Wu, Jing Yang, and Jingxian Wu. Optimal status update for age of information minimization with an energy harvesting source. *IEEE Transactions on Green Communications and Networking*, 2(1):193–204, 2017.

[39] Roy D Yates and Sanjit K Kaul. The age of information: Real-time status updating by multiple sources. *IEEE Transactions on Information Theory*, 65(3):1807–1827, 2019.

TABLE IV
LIST OF PAGES FOLLOWED

Page	Address
Jair Bolsonaro	https://www.facebook.com/jairmessias.bolsonaro/
PSL	https://www.facebook.com/PartidoSocialLiberalBR/
Lula	https://www.facebook.com/Lula/
Fernando Haddad	https://www.facebook.com/fernandohaddad/
PT	https://www.facebook.com/pt.brasil/
Geraldo Alckmin	https://www.facebook.com/geraldoalckmin/
PSDB	https://www.facebook.com/Rede45/
Marina Silva	https://www.facebook.com/marinasilva.oficial/
Rede Sustentabilidade	https://www.facebook.com/RedeSustentabilidade18/
Ciro Gomes	https://www.facebook.com/cirogomesoficial/?ref=br_rs
PDT	https://www.facebook.com/pdt.org.br/
Henrique Meirelles	https://www.facebook.com/hmeirellesoficial/
MDB	https://www.facebook.com/MDBNacional15/
Álvaro Dias	https://www.facebook.com/ad.alvarodias/
Podemos	https://www.facebook.com/podemos19/
Guilherme Boulos	https://www.facebook.com/guilhermeboulos.oficial/
PSOL	https://www.facebook.com/psol50/
Cabo Daciolo	https://www.facebook.com/depudadocabodaciolo/
Patriotas	https://www.facebook.com/Patriota51Oficial/
João Amoes	https://www.facebook.com/JoaoAmoedoNOVO/
Partido Novo	https://www.facebook.com/NOVO30/
Quebrando o tabu	https://www.facebook.com/quebrandootabu/
Jornal O Globo	https://www.facebook.com/jornaloglobo/
BBC Brasil	https://www.facebook.com/bbcbrasil/
BuzzFeed Brasil	https://www.facebook.com/BuzzFeedBrasil
Carta Capital	https://www.facebook.com/CartaCapital/
Mídia Ninja	https://www.facebook.com/MidiaNINJA/
Huffington Post Brasil	https://www.facebook.com/HuffPostBrasil/
Jornalistas Livres	https://www.facebook.com/jornalistaslivres/
El Pais Brasil	https://www.facebook.com/elpaisbrasil/
Revista Forum	https://www.facebook.com/forumrevista/
Brasil de fato	https://www.facebook.com/brasildefato/
DW Brasil	https://www.facebook.com/dw.brasil/
Observatório da Imprensa	https://www.facebook.com/Site.ObservatoriodaImprensa/
Congresso em foco	https://www.facebook.com/congressoemfoco/?ref=br_rs
Le Monde Brasil	https://www.facebook.com/diplobrasil/?ref=br_rs
The Intercept Br	https://www.facebook.com/TheInterceptBr/?ref=br_rs
Agência Pública	https://www.facebook.com/agenciapublica/
Spotniks	https://www.facebook.com/spotniks/
Poder 360	https://www.facebook.com/Poder360
O Estado de SP	https://www.facebook.com/estadao/?ref=br_rs
Revista Isto é	https://www.facebook.com/revistaISTOE/?hc_ref=SEARCH
Folha de SP	https://www.facebook.com/folhadesp/
Revista Veja	https://www.facebook.com/Veja/
O Antagonista	https://www.facebook.com/oantagonista
Band News	https://www.facebook.com/BandNews
Correio Braziliense	https://www.facebook.com/correiobraziliense
Jornal GGN	https://www.facebook.com/JornalGGN/
MBL	https://www.facebook.com/mblivre/
Pragmatismo Político	https://www.facebook.com/PragmatismoPolitico/
Nexo	https://www.facebook.com/nexojornal/